Adapting Blaise and SurveyTrak for the developing world: Case study in Ghana

Elana Cohen-Khani, Statistics Netherlands
Yu-Chieh Lin, Beth-Ellen Pennell, Gina-Qian Cheung, Survey Research Center, University of Michigan

Introduction

This paper describes survey and monitoring solutions developed to conduct in-person interviews for large scale community and household panel surveys in Ghana. These innovations include a highly flexible parallel block instrument design to allow the collection of household, individual, agricultural, and business level data in an integrated fashion, as well as a case management and data transmission system that allowed researchers rapid access to a rich array of paradata. Project managers were able to make use of timestamps, keystroke data, GPS data, and respondent contact records to track interviewer performance and movement, and map local infrastructure for analyses. Adaptations were also made to address the frequent lack of phone and internet connectivity that is common in rural areas of developing countries. This paper will discuss the challenges introduced by adapting old tools to a new environment and how these solutions enabled a more effective and efficient data collection process.

Background

The EGC-ISSER Socioeconomic Panel Survey (the Panel) is a collaboration between the Economic Growth Center (EGC) at Yale University and the Institute for Statistical, Social, and Economic Research (ISSER) at the University of Ghana. The first wave of the Panel took place between 2009 and 2010 and included 5,009 randomly selected households, containing roughly 18,000 people, spread amongst 334 enumeration areas in all of Ghana's ten regions.

The current and second wave of the Panel began in late 2013 and will be finished in mid-2015. The Wave 2 sample included the original 5,009 households as well as new households formed by household members from the original sample. The Panel aims to revisit the same households periodically over the next two decades to monitor their progress across a wide range of outcomes.

Design

Survey - Blaise

The Panel contains separate questionnaires administered at the community and household levels with two separate instruments. The community level data is primarily collected before the household level data so that some of it can be preloaded into the household instrument. The community level survey has modules for schools, health facilities, financial institutions, infrastructure, land use, livestock, crop prices, crop unit conversions, extension services, shocks, employment, daily wages, district government, social and political groups, and community infrastructure projects.

The household instrument collects data at four levels: household level, individual level, plot level, and business level. The information collected at each level is:

Household	Individual	Plot	Business
Non-resident spouses	Background	Agricultue	Basic information
Non-resident relatives	Employment history	Land tenure	Assets
Durable goods	Education	Tractor and plough use	All employees
Financial assets	Migration	Seeds	Important employees
Social networking	General health	Labour	
Information seeking	Womens' health	Harvest	
Consumption	Children's module	Chemical inputs	
Housing	Psychological and social	Sales and storage	

Table 1. *Household Survey sections by level of analysis*

Case Management – SurveyTrak

For case management, data transmission, and data processing, we used SurveyTrak, which is a sample management system created by Survey Research Center at the University of Michigan. It has been used to manage hundreds of CATI/CAPI survey, and its strength is managing longitudinal panel surveys.

Challenges

The Questionnaire

The questionnaire takes about 12 hours to complete for households where most sections are relevant, and requires speaking to multiple respondents in the house privately. Enumerators were not permitted to spend more than two hours per day with each household. Thus, the instrument needed to be designed in such a way that sections and subsections could be accessed easily and out of sequence, and enumerators needed to have a clear view of their progress and be able to know, at a glance, what they needed to do to complete a case.

To handle this, with the assistance of staff at the SRC, we built survey layout that allowed enumerators plenty of flexibility while also restricting them to specific routing where it was necessary. We used a set of color-coded hyperlinks to give enumerators immediate feedback about their progress and also to provide quick access to each section. The main parallel tab is a dashboard from which the different sections of the survey can be accessed, and enumerators can return to this tab from any point in the survey.

In Figure 1, under the "Rosters" subheading, are links to other tabs where parallels at the person (individual), plot, and non-farm enterprise (business) level are accessible. These links only become active

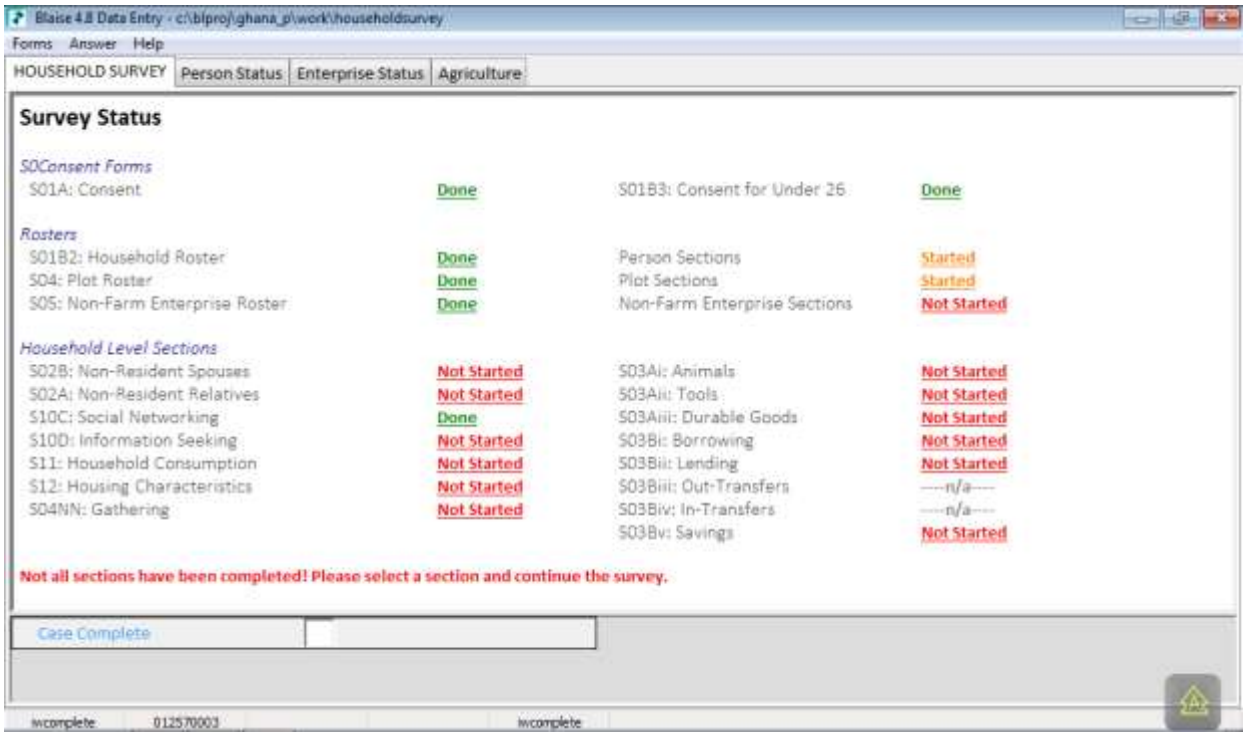

Figure 1. *Main survey parallel in for Household Survey*

when the matching roster on the left hand side of the page has a "Done" status.

As an example of restricted routings built into the modular system, most of the sections rely upon the household roster being complete, so most of the sections are inaccessible until the household roster is done. The Out-Transfer and In-Transfer sections rely on information collected in the Non-Resident Spouses and Non-Resident Relatives sections, so they will stay inactive until those other sections have been completed

Figure 2 shows the "Person" level of parallels, with names filled in from the household roster displayed on the left and a set of color-coded hyperlinks for each section relevant to that person visible on the right. Both of the individuals displayed below are men, so the Women's Health sections are inaccessible.

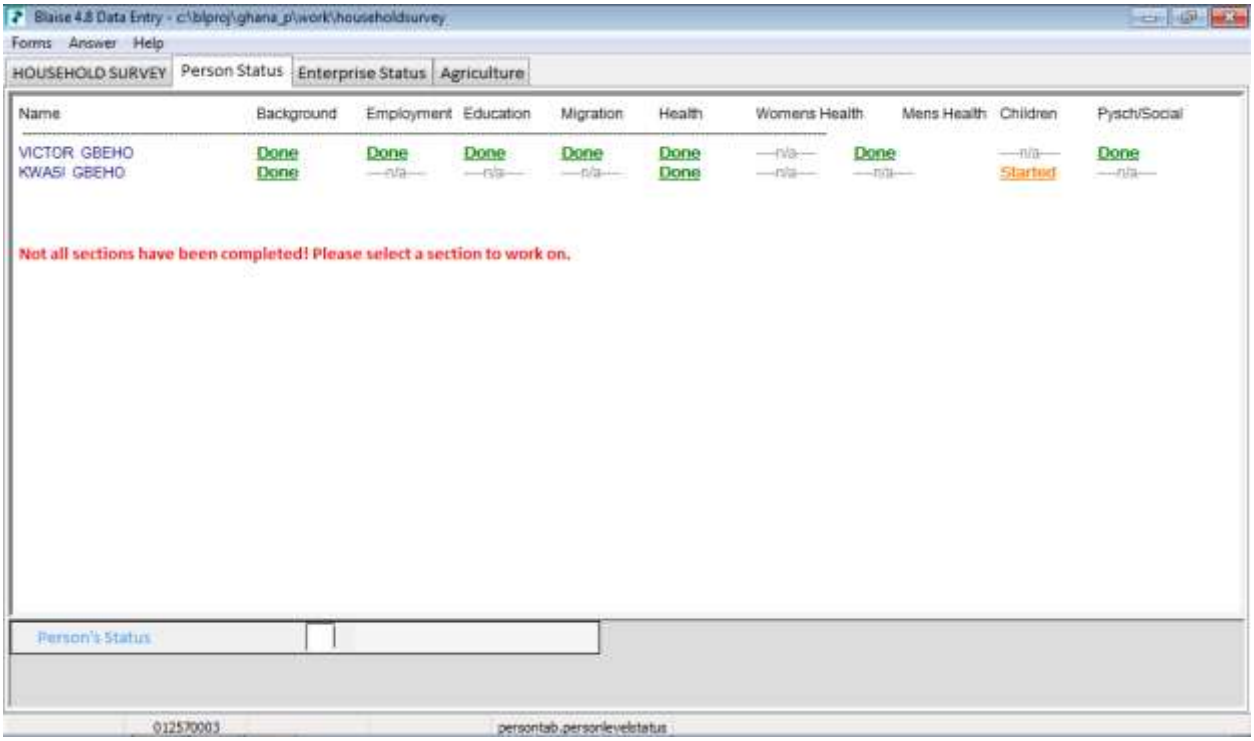

Figure 2. *Person level tab showing color-coded hyperlinks for each person's relevant parallel blocks*

Figure 3 shows the "Plot" (Agriculture) tab.

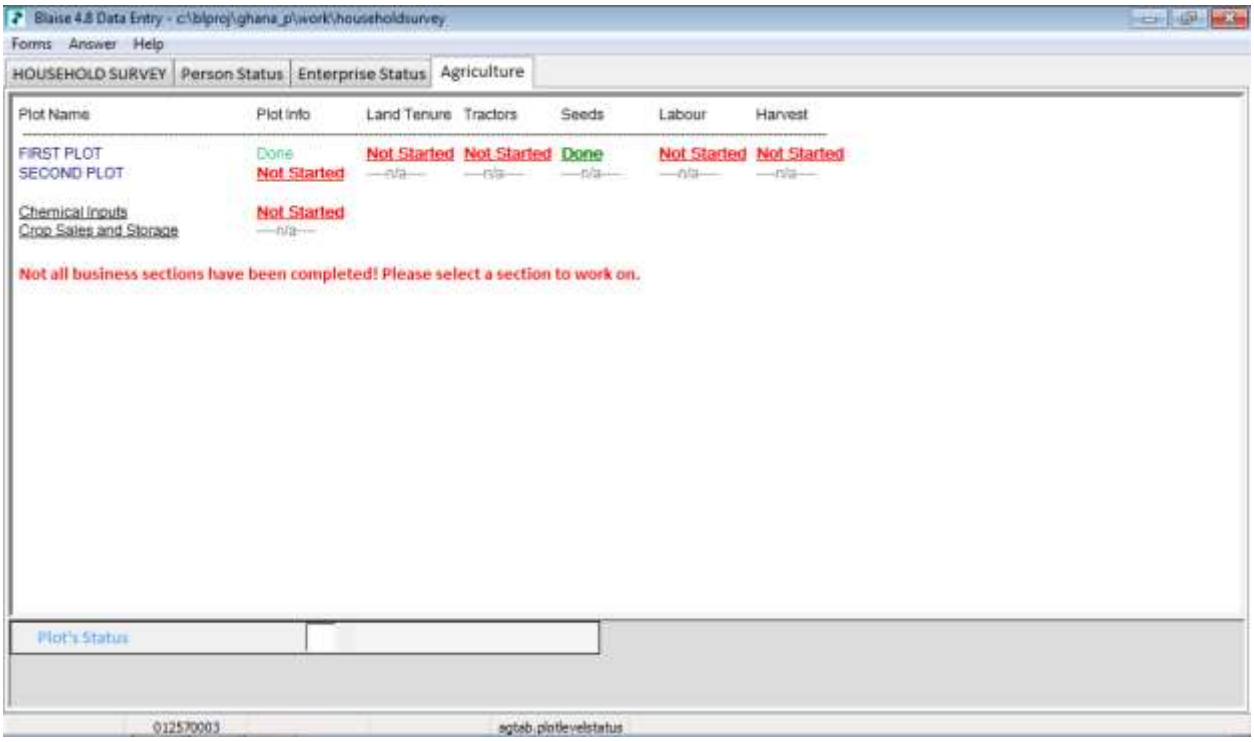

Figure 3. *Agriculture (plot) level tab showing color-coded hyperlinks for each plots's relevant parallel blocks*

Sample Management

Beyond the challenges of designing the survey instrument itself, SurveyTrak software had to be extended to handle situations where internet access was either poor or non-existent. SurveyTrak was already equipped to handle visit records, sending and receiving samples from the central server via the internet, and scheduling future calls and contacts.

Figure 4 shows the main screen from which enumerators can select households from the list assigned to them and start an interview. It displays the last result code, the status, the last date the record was touched, keeps a count of contact attempts, and gives the status of the contact.

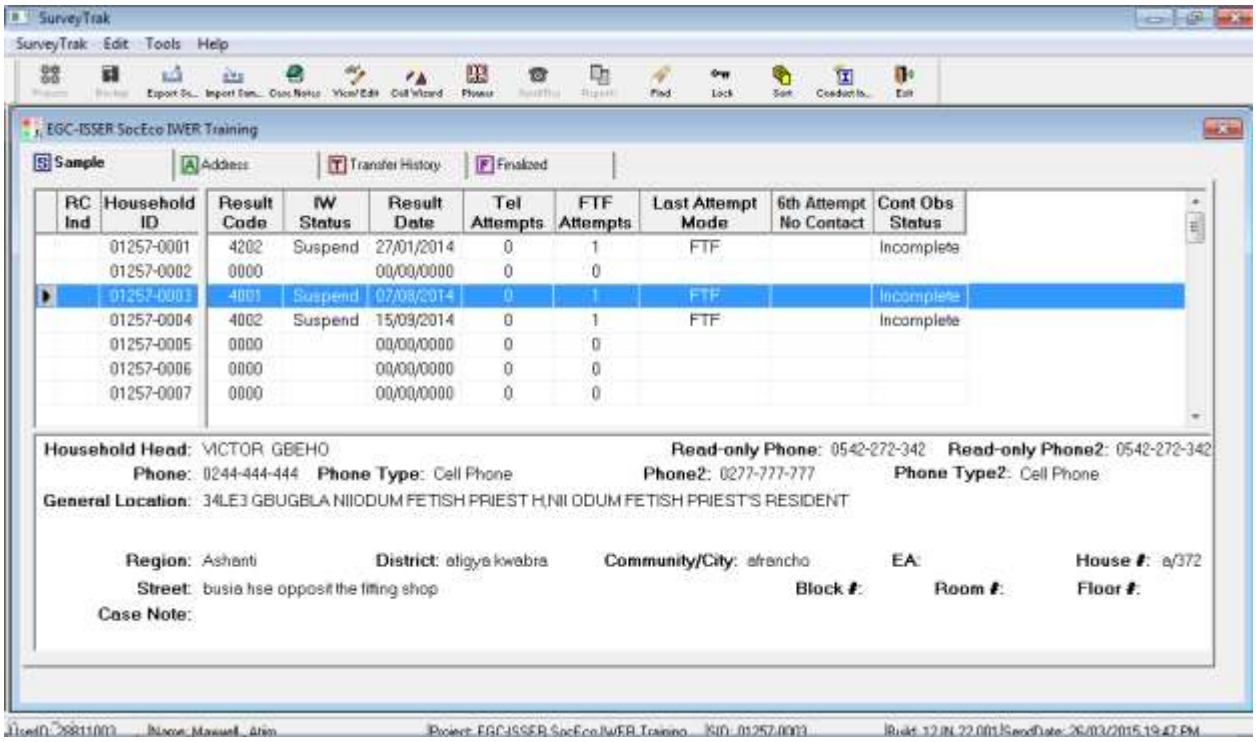

Figure 4. *Sample Selection screen*

At the bottom of the page is the contact information for the head of the household and an address or location. It is still uncommon in most Ghanaian communities for people to discuss locations using street addresses and numbers. Rather, people usually give directions in terms of relationships to landmarks.

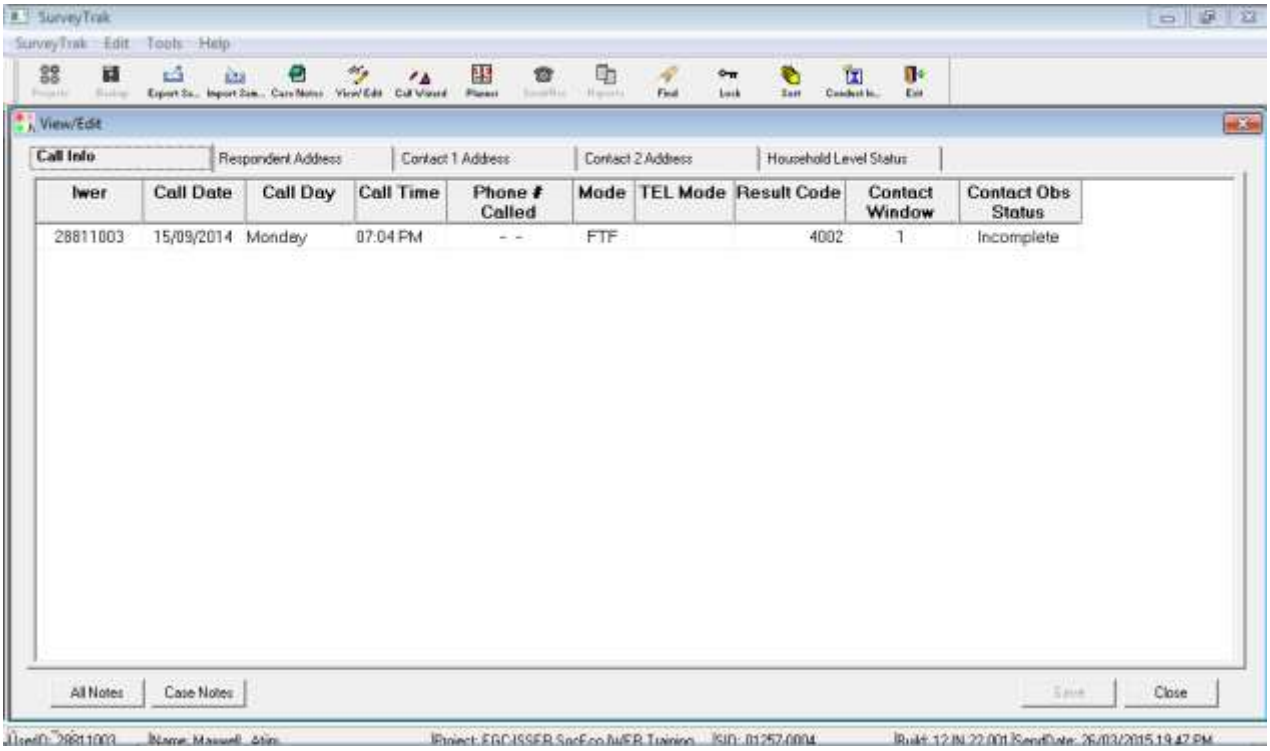

Figure 5. *Call Info*

Selecting a line in the window seen in Figure 4 and then pressing the View/Edit button pulls up the window seen in Figures 5 and 6.

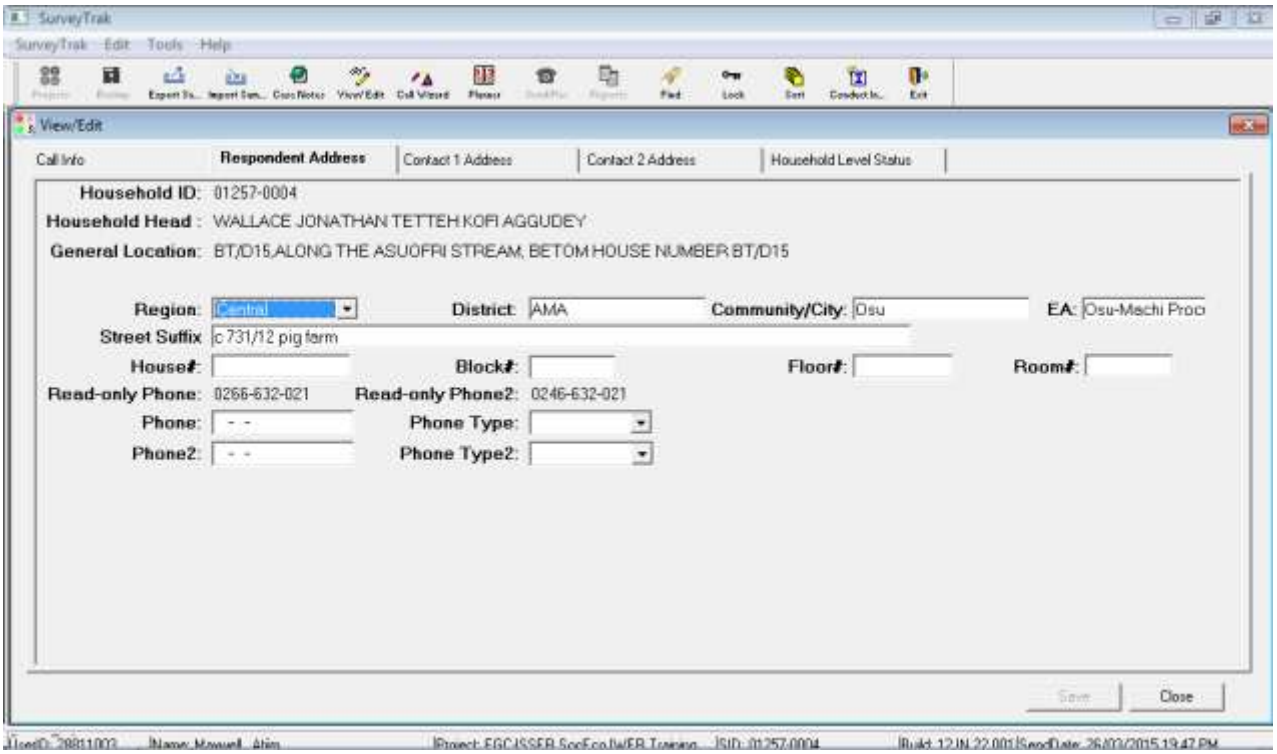

Figure 6. *Address Tab*

The Call Info tab displays a list of all contact attempts for that household, the three Address tabs allow the contact information for the household head, and two contacts from outside the household, to be recorded and updated. sd

The Household Level Status tab displays the same information as the dashboard in the Household Survey instrument, which allows enumerators to see the status of each section of the survey without having to open the survey itself.

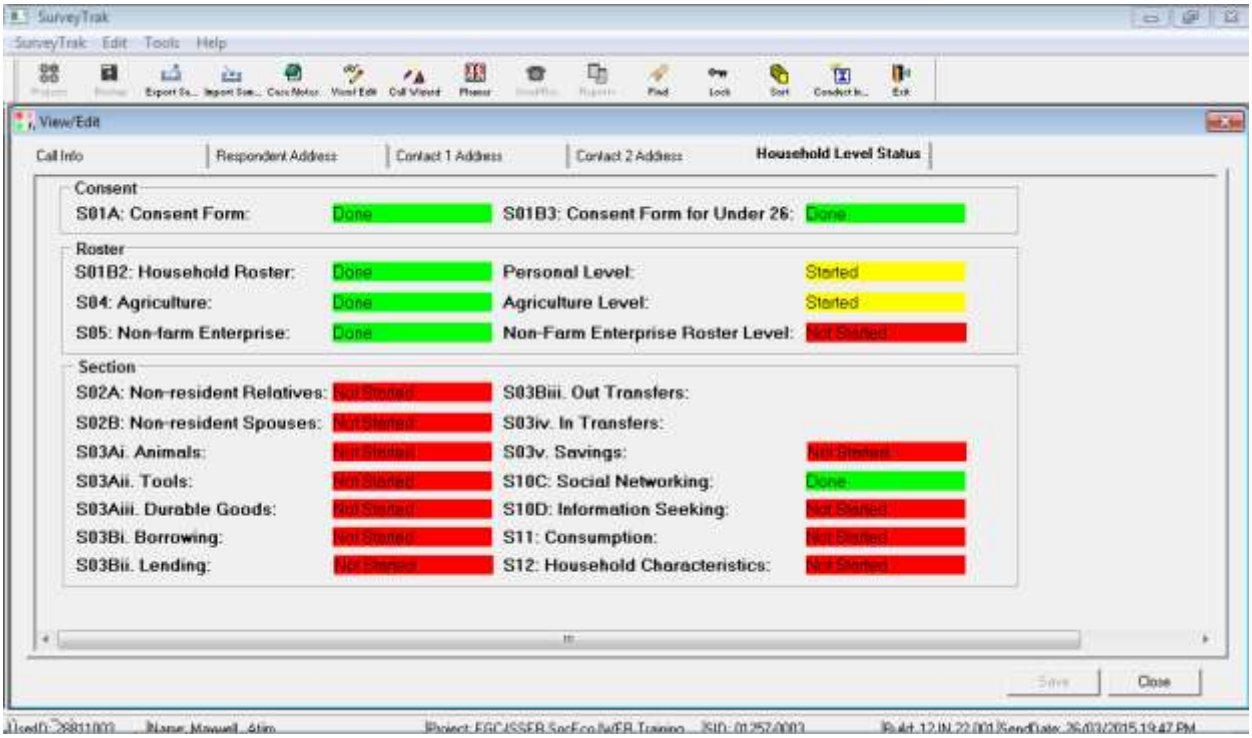

Figure 7. *Household Level Status Tab*

Data Transmission

Many of the enumeration areas were located in rural areas without reliable phone or internet connections, which meant that the usual Send/Receive process would not work in these situations when cases needed to be transferred between users or uploaded to the server when they were complete.

To address this critical issue, the SRC team devised a method to allow for offline transfers of cases between different computers using USB drives or SD cards. This made it possible for team leaders and field managers to collect completed cases from enumerators at their field assignments and send them back to the server when they reached an area with better connectivity. This also allowed team leaders and field managers to distribute new cases to enumerators as needed, even when they were far away from a good connection, all within the SurveyTrak system.

These accommodations made it possible for researchers to receive and review data and paradata with minimal delays, which is, arguably, one of the main benefits of computer-assisted surveying.

Conclusion

While there is a long history of Blaise and SurveyTrak, in addition to similar systems, being used for data collection in developed countries with reliable communications infrastructures, it has been less frequently used in developing world contexts, especially as part of an end-to-end solution. Using a combination of Blaise and SurveyTrak, we were able to deploy a comprehensive, efficient data collection system for a complicated survey instrument that worked effectively in an environment where that combination had not been tried before.