United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Development Division
Washington DC 20250

RDD Research Report
Number RDD-10-00

July 2010

Technology Advancing Data Collection: Thin Client Computer Assisted Personal Interviewing in the National Agricultural Statistics Service's 2010 Field Data Collection Program

Michael W. Gerling

James M. Harris

EXECUTIVE SUMMARY

The National Agricultural Statistics Service's (NASS) primary purpose is to provide timely, accurate and useful statistics on United States and Puerto Rico agriculture. NASS conducts over 400 surveys annually to estimate crop production, livestock, production practices, farm economics, etc. NASS has 45 field offices across the United States and one in Puerto Rico that are responsible for collecting agricultural data. These field offices employ various data collection methods: personal interview using a paper questionnaire, mail, telephone, and also Computer Assisted Telephone Interview and self-administered web.

Computer Assisted Personal Interviewing (CAPI) involves an interviewer going to the respondent with the questionnaire residing on a computer (generally a laptop) and recording all responses onto the computer. This data collection method is successfully being utilized by a number of organizations (e.g., the Natural Resources Conservation Service, the Census Bureau, Research Triangle Institute International, the Federal Emergency Management Agency and the National Opinion Research Center). Improved timeliness of the data, reduced data entry costs and improved data quality are several benefits experienced by those implementing CAPI. However, all of these organization's CAPI solutions involve storing data on the device and then uploading it later in the day. The primary flaw with this solution is that it lends itself to the possibility (although remote) of data being compromised if the device is lost or stolen.

In the spring of 2009, Michael Gerling, Mark Harris and Beth Edwards of NASS designed an innovative, thin client CAPI solution by leveraging wireless broadband technology and a recently completed web based data collection system. Hence, the name thin client CAPI was coined. In general, a field interviewer accesses the Internet using a low cost netbook with a wireless broadband aircard. The interviewer then accesses NASS' data collection website and brings up the questionnaires that need to be completed by that particular agricultural operator. The interview is conducted real time over the Internet with no data ever residing on the netbook. To date, this innovative approach to CAPI, while evolving, continues to be successful.

Technology Advancing Data Collection: Thin Client Computer Assisted Personal Interviewing in the National Agricultural Statistics Service's 2010 Field Data Collection Program

Michael W. Gerling and James M. Harris[1]

Abstract

The National Agricultural Statistics Service (NASS) surveys farmers and ranchers across the United States and Puerto Rico in order to estimate crop production and number of livestock, to assess production practices, and to identify economic trends.

Computer Assisted Personal Interviewing (CAPI) involves an interviewer going to the respondent with the questionnaire being displayed on a computer (generally a laptop) and recording all responses onto the computer. Although, CAPI has been around since the 1980's, all forms of CAPI have centered around both the data and the data collection instrument being stored on the device. Also, over the past three years, there have been several accounts of agencies' and businesses' losing or having laptops stolen containing clients' personal information. Although the possibility of someone bypassing the security mechanisms of the device and decrypting the data is small, it still exists. This possibility of data being compromised combined with the sensationalism of the news media can cause the general public to question participating in surveys, adversely affecting the entire survey industry.

NASS' CAPI solution, coined thin client CAPI, is an innovative thin client approach which leverages wireless broadband technology and a recently completed, web-based data collection system called Electronic Data Reporting (EDR).

In general, a field interviewer using a low-cost netbook accesses the Internet using a wireless aircard. The field interviewer then accesses NASS' data collection website and brings up the questionnaires that need to be completed by that particular agricultural operation. The interview is completed real-time over the Internet with no data ever residing on the netbook.

Key Words: Agriculture, CAPI, Data Collection, Wireless Broadband, Aircards

---

[1/] James M. Harris, Director of Research and Development Division and Michael W. Gerling, Mathematical Statistician for the National Agricultural Statistics Service - Research & Development Division, located at 3251 Old Lee Highway, Fairfax, VA 22030. Special thanks to Beth Edwards for solving the technical issues to keep thin client CAPI both simple to implement and cost effective.

1.0 BACKGROUND

The National Agricultural Statistics Service's (NASS) primary purpose is to provide timely, accurate and useful statistics on United States and Puerto Rico agriculture. NASS conducts over 400 surveys annually to estimate crop production, livestock, production practices, farm economics, etc.

NASS has 45 field offices across the United States and one in Puerto Rico that are responsible for collecting agricultural data. These field offices employ various data collection methods: personal interview using a paper questionnaire, mail, telephone and also Computer Assisted Telephone Interviewing (CATI) and self-administered web.

In the 1980's, Computer Assisted Personal Interviewing (CAPI) was tested and implemented in Indiana and Pennsylvania (Eklund, 1991) and (Pordugal, 1995). However, due to the high initial start-up costs of the hardware (laptops), short battery life, and difficulty in updating the survey instruments, this mode of data collection was discontinued. Meanwhile, other government agencies, institutions, and private companies continued to pursue CAPI and have incorporated CAPI into their arsenal of data collection tools. The Natural Resources Conservation Service, the Census Bureau, the National Opinion Research Center, and the Research Triangle Institute International are all successfully using laptops, tablet PCs and personal digital assistants (PDAs) to collect data. Overall, they are experiencing improved timeliness of the data, improved data quality, and reduced data entry costs (Gerling, 2004).

In 2003, NASS' North Carolina Field Office pioneered NASS' use of PDAs in data collection. They used PDAs to collect people's opinions on North Carolina's State Fair, in addition to Cotton Objective Yield Survey data (Neas, Molina, Hardegree, Gerling, 2006). Also, in 2003, NASS developed a Question Repository System (QRS) that builds paper questionnaires, and then feeds a web based self-administered form creation system called Electronic Data Reporting (EDR). This system creates questionnaires that are rendered on the Internet for agricultural operators to complete on-line.

Since NASS's first exploration into CAPI, hardware costs (laptops, PDAs and tablet PCs) have decreased while performance and functionality have dramatically improved. Laptops that once cost over $1,500 can now be purchased for around $300. In 2007, the introduction of low-cost netbooks also forced the prices of these devices even lower. This made the timing right for NASS to re-examine CAPI in 2009.

2.0 CURRENT DATA COLLECTION PROBLEMS

Currently, NASS' field data collection costs continue to rise. Postage costs for both the regular and express mailing of paper questionnaires between the field office and field interviewers are increasing. Personal delivery of completed questionnaires to the field offices, although timely, is not optimally cost effective, due to the time involved and mileage costs.

Another issue NASS faces is that field interviews which occur on the last couple of days of enumeration result in completed questionnaires that are sent or hand-delivered to the field office, sometimes too late to be edited correctly and keyed properly into NASS' data collection system.

To combat the in-swell of data entry needed at the end of the data collection period, temporary data entry staff are hired. This is a costly solution that negatively impacts field offices' budgets.

Also, the number of "cell phone only" households is increasing, and this is causing the survey industry trouble since now people may answer the phone, but are in an unsafe environment (i.e. driving) while trying to answer questions. And secondly, a majority of cell phone plans have a finite number of minutes to use, and those with cell phones may not want to respond to a survey that is going to use up their monthly minute's allotment. This is pressuring the survey industry to conduct more face-to-face interviews just to obtain data.

Finally, over the past five years, there has been a dramatic rise of private companies and government agencies reporting laptops, containing customer's data, being stolen or lost. This in turn creates a firestorm of bad publicity and investigations into what occurred and why. In some cases, staff was fired and companies had to pay for identity theft services for their customers whose data may have been compromised. Overall, this gives organizations for which data security is so essential a black eye and may also adversely affect the trust that clients have placed on the survey industry to protect and safeguard their data.

Therefore, NASS wanted a solution that protects the agricultural operators' data, which is inexpensive, for which it is easy to train field interviewers, and that leverages current NASS systems while solving the problems noted above.

3.0 PROPOSED SOLUTION (Thin Client CAPI)

After much research involving discussions with numerous vendors on netbooks, laptops, aircards and wireless broadband technology, Michael Gerling, Beth Edwards and Mark Harris of NASS' Research and Development Division developed an innovative new CAPI approach. Their solution was to use low-cost netbooks equipped with aircards to access and leverage NASS' EDR data collection website to collect the data in real time. Thus, no data would ever be stored on the device, mitigating any concerns of the data being stolen or compromised.

This solution was coined thin client CAPI (CAPI^{TC}), and it addresses and mitigates all of the problems discussed in Section 2.0. However, the plan for its large-scale implementation is not yet feasible since wireless broadband coverage is not entirely available across the United States and some surveys are not available on EDR's website.

Hence, M. Gerling proposed a staggered implementation to occur over several years across a handful of states each year. The initial states selected would be based primarily on the availability of wireless broadband coverage. The surveys chosen would be from those already available in EDR.

These selected states would distribute the netbooks to field interviewers based on their anticipated workloads and computer skills.

Approximately ten field interviewers per selected field office would be provided with a netbook having wireless broadband access. The number of devices allocated per field office was capped at ten, due to budgetary constraints. As the implementation budget

increases, more field interviewers would be provided the necessary hardware and training to conduct interviews via CAPI^{TC}.

Training would be a mixture of (classroom, field, and video) on how to use the netbooks and navigate EDR surveys. Afterwards, field interviewers would be ready for CAPI^{TC} enumeration.

3.1 CAPI^{TC} Interview

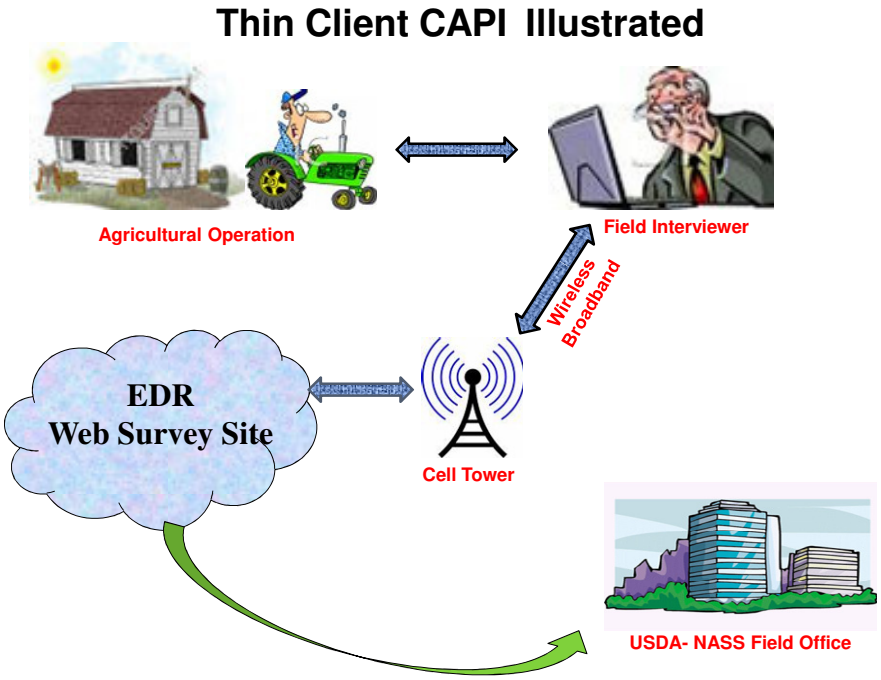
A typical CAPI^{TC} interview would start by the field interviewer arriving at the agricultural operation of interest and obtaining the operator's permission for the interview. The field interviewer would then log into the netbook, connect to the EDR website, and obtain the questionnaires for the operation. Next, the field interviewer would conduct the interview over EDR via wireless broadband access.

After completing the interview and submitting the reported data, the field interviewer would thank the respondent for his/her time, log out of EDR and the netbook, shut down the netbook, and return it to its case.

The collected data would then reside on the EDR server, from which the data could be processed by the field office for review, editing and analysis. No data would ever be stored on the netbook.

Finally, the field interviewer would return to his/her vehicle and plug the netbook into the provided car charger, so that it would be ready for the next interview. See Figure 1.

Figure 1:



The primary limitations of CAPI^{TC} and the proposed remedial solutions are as follows:

Limitation -- Connectivity is established, but subsequently lost.

Solution -- All data collected on previous web pages would be saved. Data for the current web page and subsequent pages would then be collected via paper. Later on, when connectivity is re-established, the field interviewer would enter the data from the paper questionnaire into the EDR system.

Limitation -- Connectivity is unavailable at the time of the interview, or the environment/weather makes conducting a CAPI^{TC} interview impractical.

Solution -- Data would be collected on the paper questionnaire and later entered into the EDR system.

3.2 Expected Benefits

Computer Assisted Personal Interviewing has the following benefits:

- ✓ it minimizes paper questionnaires (including storage and eventual shredding costs),
- ✓ it reduces the amount of data entry from the office,
- ✓ it provides the ability for supervisors to review work right after the interview is conducted,
- ✓ it improves data quality by using edit checks, and
- ✓ it reduces mailing costs in shipping materials between the office and the field interviewers.

Thin client CAPI has all of these benefits with the added perks of doing this more cost effectively and eliminating the risk of a respondent's data being compromised.

4.0 IMPLEMENTATION

Even good change has resistance and CAPI^{TC} was no exception. The paradigm of having the questionnaire and data on the device and at night uploading and downloading information had been around for years and it was "How everyone else was doing it."

However, through additional research and testing, the walls of the old paradigm were brought down and CAPI^{TC} is now viewed as a vital part of moving NASS forward to meet the ever-changing agricultural industry.

To implement CAPI^{TC}, a team composed primarily of staff from NASS' Census and Survey Division, responsible for the operational workings of CAPI^{TC}, was formed.

4.1 Equipment

The netbook selected was the Lenovo S10e, priced at $300. It was purchased with an extended battery which provides approximately six hours of use. The screen is visible indoors and outside. This model also comes with an express slot for the aircard. The express aircard format was selected over the typical USB aircard since the express type lessens the likelihood of the field interviewer losing the aircard, while minimizing any damage to the aircard if a netbook was dropped or bumped.

Aircards were from three wireless broadband carriers (AT&T, Sprint and Verizon). The aircards themselves were free but a monthly service fee of approximately $40 was incurred per card.

Finally, an inexpensive netbook case ($20) and a car charger for netbooks ($9) were purchased.

4.2 Where to Implement?

After reviewing aircard coverage maps, and conducting testing of the aircards, the team decided to implement CAPI^{TC} in New Jersey and Indiana.

One aircard was distributed per field interviewer based on the overlap of their assigned data collection territories with the broadband coverage maps from each of the carriers. There were some instances, of field interviewers receiving aircards from more than one carrier in order to determine which carrier would provide the greatest connectivity. Although having multiple carriers' aircards would be ideal, NASS found that providing each field interviewer with only a single service provider's aircard was more cost effective.

5.0 INITIAL FINDINGS

In general, NASS' field interviewers consist primarily of retirees with varying levels of computer skills. For the initial implementation, those field interviewers having the best computer skills were selected. This would not only help in ensuring a good start but also enable these first CAPI^{TC} interviewers to be future trainers for those not initially receiving the training.

Instruction consisted of a half day in the field office educating office staff on CAPI^{TC} and addressing any concerns. The second day focused on instructing the field interviewers how to use their netbooks and aircards and how to access and navigate the EDR system. The instruction was classroom style with several opportunities for hands-on practice.

Overall, the field interviewers found using the netbooks and aircards rather straightforward. Most of the questions centered on the EDR system. Field interviewers had several questions on how survey questions were rendered on a web page and on the logistics of navigation between pages. Although the time discussing the on-line questionnaires greatly extended the training to a full 8 hours, the comments obtained will significantly improve EDR's on-line instruments not only for CAPI^{TC} but also for the agricultural operators who access the on-line questionnaires through self-administration.

As the field interviewers moved from the classroom to the real world, they found that some netbooks took over 15 minutes to boot up. The first solution was to have field interviewers always logged on to the netbooks. This circumvented the boot-up time, but with the netbooks always on, heat generated from the netbooks would literally fry the express type aircard, causing the ability to connect or hold a steady wireless broadband connection uncertain. The second solution was to remove all of the unneeded software that came on the netbooks. This latter solution was effective in greatly reducing the boot-up time problem without introducing new ones.

Overhead power lines and buildings with thick (especially metal) walls would also negatively affect the ability to connect or hold a broadband signal. This problem, however, appears to be mitigated by using Sprint's Overdrive or Verizon's Mi-Fi. These standalone devices are smaller than a deck of cards and contain a built in aircard and battery. Once the device connects, the device creates a Wi-Fi hot-spot which seems

better at penetrating walls. The netbook then obtains the Wi-Fi signal to access the Internet and EDR. Additional testing still needs to be conducted but preliminary testing supports the effectiveness of this approach.

6.0 LESSONS LEARNED

As with implementing any new technology, there are growing pains. However, the simplicity of CAPI^{TC} was the saving grace of the initiative. For those looking to adopt CAPI^{TC} into their business, the following sections of this report document some of the lessons learned when adopting this data collection method.

6.1 Equipment

Price isn't everything. Depending on the location and type of data being collected, netbooks are not the best solution. Paying a little more for a low cost convertible or tablet PC may shorten the interview and save more money over the life of the device. When a low cost convertible tablet PC was purchased and tested with field interviewers, all of the interviewers preferred using the tablet PC's stylus over the netbook's touch pad.

All devices (laptops, netbooks, tablet PC's, aircards, etc…) need to be configured correctly ahead of time. Attempting to do this at the training site wastes valuable time, and it's extremely difficult to make sure all equipment is configured correctly. All unneeded software needs to be removed during the configuration process. This will significantly shorten the boot up time.

Keep it simple. The fewer mouse clicks / key strokes for field interviewers to press or select on the device, the better. One shortcut that can be implemented is having the aircard automatically connect to the Internet after logging in.

6.2 Training

Recommend a combination of classroom, hands-on and outdoor training. Make sure that the training room has wireless broadband capability.

7.0 CAPI^{TC} or CAPI?

Currently, NASS' implementation of CAPI^{TC} is still in its fledgling state. As compared to other agencies' CAPI solutions, NASS' CAPI^{TC} approach appears to be more cost effective and requires less start-up costs.

CAPI^{TC} can use low cost netbooks while regular CAPI requires a more expensive laptop due to the added security measures needed for storing and transferring data.

CAPI^{TC} requires less technical support since there is less software required to maintain and all of the data transfer is done real time over the Internet. In contrast, CAPI requires additional data transfer and security software solutions which can conflict with one another. Typically, the technical staff required for standard CAPI tech support would be more than twice the size of the one needed for CAPI^{TC} technical support. (See Gerling 2004.)

CAPI^{TC} only requires a one full day of training, whereas standard CAPI training is typically conducted over several days. This additional training is due to the time it takes to instruct on how to upload and download data, as well as how to log in through all of the security protocols.

Standard CAPI, however, does shine in its ability to conduct an interview regardless of wireless broadband coverage. Although wireless broadband coverage is expanding, the ability to conduct an interview via a computer at any location is a major plus for CAPI.

Focusing on data security and the public's concern with identity theft, standard CAPI is more vulnerable to data being compromised. This risk is due to collected data being stored on the device at various times during the data collection period. Eventually, a CAPI netbook, tablet PC, or laptop will be lost or stolen. Although the device is password protected and the data are encrypted, there is still a chance (although remote) that the data could be compromised.

If NASS was to implement CAPI and a device (containing agricultural operations' data) was stolen or lost, the problem would be how this would affect the agricultural industry's and the general public's perception of NASS. Another potential problem is the impact of any media coverage on a lost or stolen NASS CAPI device. Depending on what else is happening in the news, this could be a public relations nightmare for NASS and the survey industry. The Census Bureau, Internal Revenue Service, General Electric Corporation and Wells Fargo have all had laptops, containing clients' data, lost or stolen which resulted in negative media coverage, and employees either released or demoted. Also in the case of Wells Fargo, the company had to pay for a year's worth of identity theft services for those clients whose data may have been compromised.

In comparison, CAPI^{TC} data are never stored on the device, so a loss would strictly be a financial one not a data security one. Also, NASS could take a loss and turn it around to promote its data collection efforts and explain why the agency chose the CAPI^{TC} solution so as to protect the farmers' data.

Overall, as wireless broadband continues to expand, the CAPI^{TC} solution will become more and more cost effective.

8.0 FUTURE

Technology is continuing to evolve, and wireless broadband continues to cover more rural areas. Urban areas are seeing increases in data transmission speed with the introduction of 4G wireless broadband. Also, technology that combines both aircard and satellite technologies is making its way to market.

In April of 2010, Apple Corporation made a major move into the tablet industry with the introduction of the iPAD at an aggressive price point of $500. Also, Apple's Corporation's pricing of its wireless broadband service starting at $15 is placing pressure on all wireless broadband carriers to lower their prices as well. In early April, NASS was able to procure an iPAD and soon realized that this device has much promise. Currently, NASS is seriously considering making the iPAD the device of choice for future implementation.

CAPI^{TC} has also opened future opportunities. Over the next 3-5 years, this data collection method has the potential to provide directly/indirectly the following capabilities:

a. Enable field staff to complete timesheets and expense reports on-line.
b. Provide access to interviewer manuals on-line.

c. Improve data quality (via edit checks).

d. Provide real time routing and navigation for field interviewers driving to agricultural operations.

Over the next ten years, NASS' CAPI^{TC} could morph into CAI^{TC} (CAI = Computer Assisted Interviewing). In CAI^{TC}, the field interviewers would conduct both personal and telephone interviews using the EDR questionnaires as computer assisted telephone interview instruments. This will cause additional paradigms of staffing needs and data collection to be re-thought. *Note: At the time of writing, CfMC (a company based in San Francisco, CA) is now offering their research services for broadband CAPI and WebCATI solutions. Hence, this paradigm shift has already begun.*

Thin client CAPI's simple model provides flexibility to adapt to NASS' ever evolving business needs, to meet current and future demands of the agricultural industry and those of the federal government. The future holds vast opportunities and through research selecting the right opportunities to implement, routine processes will be easier and capabilities greatly expanded.

9.0 REFERENCES

Gerling, M. (2004) *"A New Look Into Portable Electronic Devices for Field Data Collection in the National Agricultural Statistics Service"*, (White Paper) United States Department of Agriculture, National Agricultural Statistics Service.

Neas, K., Molina, J., Hardegree, J., Gerling, M. (2006) "Using Personal Digital Assistants for the 2004 Cotton Objective Yield Survey", Research and Development Report RDD-06-01, United States Department of Agriculture, National Agricultural Statistics Service.

Eklund, B. (1991) *"Mobile Computerized Data Entry And Edit For The Livestock Prices Received Survey"*, SRB-91-10, United States Department of Agriculture, National Agricultural Statistics Service.

Pordugal, T. (1995) *"The Effects Of Editing And A Review Of Warning Errors From A CAPI/IE Perspective"*, SRB-95-05, United States Department of Agriculture, National Agricultural Statistics Service.