

USING PERSONAL DIGITAL ASSISTANTS TO COLLECT SURVEY DATA

S. M. Nusser

Department of Statistics, 220 Snedecor Hall, Iowa State University, Ames, IA 50011-1210

D. M. Thompson

USDA Natural Resources Inventory and Analysis Institute
202 Snedecor Hall, Iowa State University, Ames, IA 50011-1210

G. S. DeLozier

Aristar, Inc., Akron, OH 44333

Keywords: CASIC, computer-assisted, data collection, pen computing, handwriting recognition, electronic data recorders, natural resource surveys, environmental monitoring

Introduction

New technologies for computer-assisted survey information collection (CASIC) are being developed as graphical user interfaces, mouse/pen-based computing, handwriting recognition, hand-held computers, and use of the Internet become more widespread (Connett, Mockovak and Uglow, 1994; Gilluly, 1994; Nichols and Apple, 1994; Shanks, 1994; Forster and Snow, 1995; Hansen, 1996). The utility of these new technologies depends on the survey setting, including the type of survey data to be collected, the size of the survey, and conditions for collecting survey data.

We consider CASIC technologies for conducting longitudinal natural resource surveys that involve collecting survey data from multiple sources, including field observations, office records, photo-interpretation and geographic information systems (GIS). Examples of such surveys include natural resource and land use surveys, soil quality surveys, soil mapping updates, forest inventories, and environmental monitoring surveys.

We focus on the National Resources Inventory (NRI), conducted by the U.S. Department of Agriculture's Natural Resources Conservation Service (NRCS) in collaboration with the Iowa State University Statistical Laboratory (ISUSL). The NRI is a national longitudinal survey in which data on hundreds of variables are collected periodically from approximately 300,000 area segments (PSUs) and 800,000 points. The national survey was conducted in 1982, 1987 and 1992, and interim special studies have been conducted in 1995 and 1996 using the national sample as a frame. The survey instrument for the national survey includes sections on land use, soil characteristics, agricultural

practices, wetlands, rangelands, vegetative cover, and wildlife habitat. Historical data on sample units are used to collect information and validate sample locations. NRI surveys are conducted using a mobile work force spread throughout the U.S. When specialized skills are required to obtain survey data, multiple data gatherers may collect information on the same sample unit.

In recent years, the NRCS has used internally-created computer-assisted survey instruments (CASIs). In preparation for the 1992 NRI, NRCS developed a CASI on UNIX and DOS computers for state offices. The survey software was designed to manage the survey data base and guide data gatherers through the survey instrument with logic for navigation and branching, and rules to validate and cross-check observations. A separate program on the state office computer was used to check the data for compatibility and consistency. Once data were complete and checked, the survey information was transmitted to the ISUSL on diskettes and tapes. State data were subsequently checked by ISUSL with ancillary data bases and more extensive edit routines. Questions and problems were resolved by personal calls and exchanges of faxes.

The 1992 NRI instrument was designed for direct entry of data as they were collected by photo-interpreters operating in centralized remote sensing facilities. However, paper worksheets were used more frequently than expected because of limited computer resources, data gatherer preferences and abilities, and the need to gather information at locations without convenient access to computers. As a result, the CASI functioned more as a data entry system. The software was discarded after the 1992 NRI was completed.

CASIC Objectives

Recent developments in hand-held computers and telecommunications have provided the basis to develop

a new approach to CASIC that addresses deficiencies in previous systems and that supports the increasing need to collect complex survey data on an annual basis. In designing a new system for NRCS surveys, several design goals were identified:

1. Store, manage, and exchange data via networks with one central data base system.
2. Use small, inexpensive, and robust data collection devices to support data gathering by a mobile work force.
3. Send/receive sample units, messages, and survey management information to and from the hand-held computer and central server.
4. Rapidly develop CASIs with commercial software for large and small surveys.
5. Create CASIs based on a simple, intuitive user interface allowing complex questions to be displayed.
6. Execute data quality checking rules and procedures inside the CASI as data are collected.
7. Make on-line supporting materials accessible while operating the CASI.

These goals were used to develop a CASIC system through the collaborative efforts of NRCS, Aristar, Inc., the ISU Computation Center, and the ISUSL. In the remainder of the paper, we describe the survey system developed for NRCS surveys, and discuss this system as it applies to other surveys and the field of CASIC.

System Overview

The CASIC system has three main components: the central data base and server, the electronic data gathering device, and the software used to collect data and communicate with the data base. A DEC 1000A UNIX server supports an ORACLE data base that contains data for each sample unit in the survey. In addition to new variables to be collected during the survey, the data base typically contains historical data on each sample unit that is used to support data gathering and verification of the sample unit's location. The data base also includes tracking information on the history of data collection for the sample unit. System robustness and security are achieved by using RAID for disk storage, having a duplicate server attached to the same disk array designed to take over database operations in the event of a hardware failure, and by having data gatherers login to parallel front end machines rather than the data base server. The modem bank is also connected to parallel terminal servers.

The data collection device is a personal digital assistant (PDA). We currently use the Newton

MessagePad 130 (NOS 2.0) accompanied by a modem and a 4MB storage card (Figure 1). The MessagePad is a pen-based PDA that contains capabilities beyond that of many electronic data recorders, including communications and data sharing between applications. The operating system is based on object-oriented paradigms. A LCD, optionally back-lit, graphical user interface is used in the display. The MessagePad supports handwriting recognition and also provides pop-up keyboards. The cost of the units and necessary peripherals is about \$1000.

The survey software is commercially developed by Aristar, Inc., and currently requires Newton OS 2.0 for execution (i.e. is not limited to the MessagePad). The software contains a full-featured survey instrument with on-line access to supporting materials. Communications capabilities with the central database are accomplished via modem. Currently, communications are used for data exchange (checking sample units in and out of the central data base) and survey management.



Figure 1. The Newton MessagePad 130, the hand-held computer selected for data gathering.



Computer-assisted Survey Instrument

We describe data entry features using an instrument developed for the 1996 NRI Land-Use Study, conducted by NRCS for the 1996 growing season to investigate the effect of the Federal Agricultural Improvement and Reform Act of 1996 and prices on land use and conservation practices. The 1996 NRI is based on a subsample of approximately 4,000 PSUs selected from the 1992 NRI sample. The basic design is a two-stage area sample, in which a typical primary sampling unit (PSU) is a quarter section of land (160 acre square), and the second stage sample unit is a point within the PSU. One to three points were selected for each PSU.





In what follows, a "section" refers to a set of questions or a display that is designed to be visible within the display area of the PDA. A "form" is a scrollable sequence of sections.

Overview. The CASI consists of several components. First, a screen is displayed for selecting

1996 NRI Data Gathering	
MLRA/FIPS/PSU/Status	
106/20013/040101R/NS	-----
106/20013/040401R/NS	-----
72/20023/010101B/NS	-----
72/20023/010201P/NS	-----
72/20023/010402G/NS	-----
72/20023/020102R/NS	-----
72/20023/020501P/NS	-----
72/20023/030301G/NS	-----
72/20023/040202B/NS	-----
72/20023/040401R/NS	-----
72/20023/040502B/NS	-----
7/53001/010103O/RT	-----
7/53001/010204O/RT	-----
7/53001/010504O/RT	-----
7/53001/020201R/RT	-----

Find	Open	Connect	Setup
------	------	---------	-------












 Names
 Dates
 Extras

 Undo
 Find
 Assist

Figure 2. Screen for selecting PSUs for data gathering, and for connecting to the central data base system.

and transferring a group of PSUs from or to the server, and for selecting and opening a PSU for data entry (Figure 2). After opening a PSU's form, a section appears that displays PSU information and buttons to skip to other sections for each point in the PSU (Figure 3). Below the initial screen are a series of scrollable sections to record PSU data. By tapping on a point's button in the first PSU section, the user moves to the data entry sections for that point's form.

Navigation within a form. After opening a PSU form, the user can press an overview button at any time to select a section associated with PSU variables. The overview button is the dot between the up/down scrolling arrows at the bottom of the screen (see Figure 3). Alternatively, forward (down arrow) and backward (up arrow) scrolling within the PSU can be used to navigate through PSU sections. The same features are available in the point forms. As with most CASIs, skip control is executed conditional on answers to questions. For example, in the 1996 NRI point forms, data collection sections to gather conservation

2005/020307P:PSU Specific Data	
PSU	
State	Kansas Note
County	Atchison
FIPS	20005
PSU	020307P
Point #1	Point #2 Point #3
General Information	
MLRA	106 Note
HUC	10270104
USLE R (rainfall factor)	200
WEQ C (climatic factor)	0
Gathered by? (96)	








 Names
 Dates
 Extras

 Undo
 Find
 Assist

Figure 3. First PSU screen, with direct navigation to sample point forms, and with intra-PSU form navigation through MessagePad's scrolling arrows and overview button at the bottom of the screen.

20005/020307P:PSU Specific Data

PSU Cropland Inventory

ID	Polygon T	Ac.	Pt.	Pract.	Notes
1	C-C	12	1	1	Small gro
2	C-N	20	2	0	
3	N-N	20	3	0	
4	C-N	15	0	0	Idled
5	C-C	80	0	1	
6	N-N	10	0	0	Road

Total Acres: 157.0

Names Dates Extras Undo Find Assist

Figure 4. Example of a grid whose dimension requires scrolling. The "Notes" column (shown above as an angled column) can be viewed by scrolling to the right.

practice information are displayed only when cropland has been recorded for a point.

Question and response formats. NRCS surveys frequently require complex response structures. For example, grids are used to accommodate the longitudinal nature of the survey. Each question in a related set of questions is associated with a row in the grid, and each column designates a survey year. The grids often contain historical data collected during previous surveys. Some grids are large enough to require scrolling capabilities. An example of such a grid is contained in the PSU Cropland Inventory section in the 1996 NRI instrument (Figure 4). The grid contains scrolling arrows in the lower right corner.

Paged sections are useful for collecting the same data on each of several subunits. For example, in the 1996 NRI instrument, data must be recorded for each aerial photograph used to support data collection. The film format, type of imagery, date and scale of each photograph are recorded using pick lists on a single page (Figure 5). The user "flips" the pages, each containing the same layout, to enter data for distinct photographs. Pages are turned by tapping on the left or right arrows at the upper right corner of the section. The page number is presented between the arrows.

Several formats are available for entering specific data values, including hierarchical pick lists, keyboard

20005/020307P:PSU Specific Data

Information Resources

Source of aerial photography **USDA**

Film format: **9" x 9"**

Kind: **Color**

Date of photography: **10/9/96**

Scale: **1:7,920**

Index:

Figure 5. Example of a paged section.

entry, numeric keypad entry, and direct handwriting. Hierarchical pick lists can be used to collect data on a few interrelated variables within one question. Editable pick lists can also be created.

Edit checks and calculations. Hundreds of edit rules are required to check data collected in NRI surveys. These rules are used as range checks and to evaluate the consistency of recorded data with other variables and with historical data. To minimize the computational burden, checking routines are associated with each section of a form. After the user enters data, s/he taps on the check box at the top left of the section (see, for example, Figure 5). Checks are executed only if all variables have been entered for a particular edit rule. If all variables have been entered for a section and no warnings or errors are detected, the check mark persists. If errors or warnings are detected, short messages explaining the error are provided and stored for subsequent recall or deletion. First, a list of errors, i.e. the first few words of each message, is presented (Figure 6). The user may tap on a message to see the full text, and then scroll through the messages using the left/right arrows in the error box. Messages may be discarded after being viewed.

In addition to logical edit rules, on-line calculations are executed. This is particularly useful to check calculated values that are derived from a function of several observed variables and that are included in the final data base.

Completion codes. After the user has completed work on a specific PSU and its points, a global checking routine is initiated in the PSU Completion Check section (Figure 6). The routine consists of

20101/010403P:PSU Specific Data		
PSU Completion Check		
Section	Checked	Notes
PSU	NA	
General	There is an inconsistent change in Contouring/Cross slope - Missing Contour stripcropping - Missing or Terraces - Missing or invalid data Shape - Missing or invalid data: 0 Acres - Missing or invalid data: 0	
Info. Re		
Croplan		
Pract D		
<input type="button" value="Discard"/> <input type="button" value="←"/> <input type="button" value="→"/> <input type="button" value="✕"/>		
Return incomplete PSU		
Point 1 Status	IP	
Point 2 Status	CO	
Point 3 Status	NS	
Completion Status	IP	
<input type="button" value="⌚"/> <input type="button" value="?"/> <input type="button" value="⌨"/> <input type="button" value="✕"/>		

Figure 6. PSU Completion Check form with error message summary.

running the section checks and determining a completion status for each section, which is stored in the central data base when the PSU is returned to ISUSL. A global complete code is determined for each point form as well. The point form codes are: not started (NS), in progress (IP), unable to complete (UC), and complete (CO) (Figure 6).

On-line support. Other commercial software is used to create a hypertext version of a document that provides detailed descriptions of the rules used to check for errors and inconsistencies in the data. When errors or suspicious observations are detected, a full description of the problem can be examined on-line while the data gatherer is using the CASI. As with PSU and point forms, the overview button or the scrolling arrows can be used to navigate to sections of the document. In addition, the user can electronically mark a page in the document for future reference.

Notes. In any section, the user can record special notes to be stored in the data base. This is accomplished by tapping a "Notes" button available in every section (see, for example, Figure 5). Notes may be reviewed as a group and edited before being stored in the data base.

Communicating with the data base server. The Aristar, Inc. software contains protocols to communicate with the central data base server for several functions, including querying the data base for available sample units and checking out or returning sample units. By tapping the "Connect" menu on the screen listing PSUs available (Figure 1), the user can call into a terminal server via modem using a simple user interface. Queries or data transfer can be executed from this same menu.

Discussion

The use of portable CASIC systems in natural resource surveys is not new. The U.S. Forest Service and commercial foresters have used portable electronic data recorders for several years in their forest surveys. Scott (1990), Gilluly (1994) and Hansen (1996) describe hand-held, typically DOS-based systems that include keyboard entry, range checking for entered values, logic and calculation capabilities, and the use of peripheral ports for attaching other equipment such as global positioning system (GPS) units. Hand-held electronic data recorders have also been successfully used for simple in-person health surveys in developing countries (Forster et al, 1991; Forster and Snow, 1995).

Electronic data entry coupled with on-line edit routines on desktop, laptop and hand-held devices have historically provided significant gains in the efficiency of conducting human and natural resource surveys. However, Shanks (1994) notes that CASIC systems have not fully taken advantage of recent technologies related to graphical user interfaces, object oriented programming, and the internet. The PDA CASIC system uses these features to provide expanded capabilities relative to hand-held electronic data recorders. The PDA system adds survey management tools, network-based data transfer, simple navigation, complex question and response formats, and hypertext on-line help to the suite of features available on hand-held devices.

Negroponete (1995) suggests that the perfect metaphor for a computer interface is a piece of paper and a pencil, noting that people are already "trained" to use these devices. In developing computer-assisted information systems, the challenge is to imbue virtual paper and pencil with logic and utility to capture information in digital form, and in a way that makes the interface "disappear." Part of the design of the PDA software was to make the data gathering features emulate the utility and form of the paper worksheet of past NRIs. While this system may fall short of Negroponete's goal, it is our hope that the system's

capabilities will be sufficiently simple to be accepted and fully utilized by data gatherers. Early tests with a Soil Quality Pilot Study for NRCS have shown that the data entry features are approaching this goal. Users of the software were asked to complete a survey documenting their experiences and suggestions. Many of their comments were used to redesign and improve the utility of the current system.

New developments in the system are planned. Research is under way to incorporate e-mail and to electronically update survey instruments. Development software is being created by Aristar, Inc. that would allow survey researchers to design their own survey instruments using a simple user interface. In addition, Aristar, Inc. has begun work on a product to view the data on a desktop computer in a form that mimics what is seen on the PDA. Both products are under development for Macintosh and Windows operating systems.

Finally, although the PDA CASIC system described in this paper is being developed for a complex longitudinal survey that requires a mobile work force, the system is easily adapted to simpler survey environments. This is especially true for surveys involving data collection through observation rather than human interviews. However, Forster et al. (1991) and Forster and Snow (1995) have shown that hand-held devices can be used in simple in-person surveys. It would be interesting to apply this technology to a more complicated questionnaire for human subjects.

Acknowledgments

We thank Bob Dayton, Masoud Kazemi and Dick Dorsch for their extensive contributions to this project. We are also indebted to the Iowa State University Computation Center staff, who designed the data base server and terminal systems. Tom O'Connor, Kathie Reinertson, Melissa Swanson, and Herb Wilson provided valuable support in testing the CASIC system. Finally, sincere thanks to Matthew Gessner, Michael Gessner and Matthew Orley for their tireless work in creating the communications and rules software.

References

- Connett, W. E., Mockovak, W., and Uglow, D. (1994) CAI systems--the users' perspective. *Bureau of the Census 1994 Annual Research Conference Proceedings*, U.S. Department of Commerce, Washington, D.C., 665-680.
- Forster, D., Behrens, R. H., Campbell, H., and Bypam, P. (1991) Evaluation of a field data collection system for health surveys. *Bulletin of the World Health Organization*, 69(1):107-111.
- Forster, D. and Snow, R. W. (1995) An assessment of the use of hand-held computers during demographic surveys in developing countries. *Survey Methodology*, 21(2):179-184.
- Gilluly, D. (1994) Integrating computer technology for field foresters. *Journal of Forestry*, 92(5):24-27.
- Hansen, M. H. (1996) Portable technologies. *Journal of Forestry*, 94(6):29-30.
- Negroponte, N. (1995) *Being Digital*. Knoff, New York, 243 pp.
- Nichols, W. L. II and Appel, M. V. (1994) New CASIC technologies at the U.S. Bureau of the Census. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Alexandria, VA, 757-762.
- Scott, C. T. (1990) Computers in the field. *Forest Farmer*, 49(3): 8-10.
- Shanks, M. J. (1994) Evolution vs. revolution in computer-assisted surveys: trends and issues concerning the next generation of CASIC technology. *Bureau of the Census 1994 Annual Research Conference Proceedings*, U.S. Department of Commerce, Washington, D.C., 681-696.