The 1997 Census of Agriculture Experience at NASS

Roger Schou and Asa Manning, National Agricultural Statistics Service (USA)

Introduction

The Census of Agriculture is conducted every five years for the years ending in two or seven. So in the decade of the 90's, the Census years would be for 1992 and 1997. Through the1992 Census, the Bureau of Census of the Department of Commerce was responsible for every survey. However, as the budget for Commerce was reduced during the mid-90's, funding for the Census of Agriculture was eliminated. This left the 1997 Census in jeopardy, but opened the door for the United States Department of Agriculture and specifically, the National Agricultural Statistics Service (NASS) to step in. Since NASS is responsible for every other federal government survey related to agriculture, the agency has always hoped someday to be able to "complete" its mission and assume responsibility for the Census of Agriculture. NASS approached congress about the Census being transferred to NASS. This was officially done prior to the 1997 Census of Agriculture.

Although NASS had a goal in its mission statement to assume responsibility for the Census of Agriculture in the future, the timing of the transfer presented a daunting task. Not only did the transfer happen quickly, but the 1997 Census of Agriculture was right around the corner. Planning for taking over the Census began during 1996. At that time, the mailing of the Census questionnaires was less than 18 months away. Not only was the timing tight, but the scope of the job ahead dwarfed any survey NASS had ever done. The Census of Agriculture attempts to collect data on every farm in the United States. A sample of 50,000 would be large by typical NASS standards, so attempting to gain a response from slightly less than two million farms in the U.S. was certainly a challenging task.

The Bureau of Census is essentially a centralized agency. Except for a large mailing facility in Jeffersonville, Indiana and CATI calling centers in Arizona, Maryland, and Indiana, all the staff assigned to the Agricultural Division, work in Suitland, Maryland, a suburb of Washington, D.C. In the past, almost all the staff assigned to work on the Census was located in either Suitland or Jeffersonville. When the responsibility for the Census was transferred to NASS, all staff in that division were given the option to transfer to NASS as well. A majority of the staff did transfer. That staff played a major role in educating the staff in NASS about conducting a Census of Agriculture survey, and served as the coordinators for the Census effort.

NASS is much more distributed with over half of our staff located in 45 field offices located around the country. As NASS planned for the Census, it wanted to take advantage of the local agricultural expertise of our field offices, but also had to tap the talents and experience of the staff from the Bureau, who had been involved in the prior Censuses.

Large Samples and Limited Time Frames

Another limiting factor was the tight time frame. The planning for the 1997 Census of Agriculture had been under way for about three years before NASS got involved. With such a short time frame available before the start of the Census, NASS was limited in what could be changed.

The Bureau of Census maintains a large facility in Jeffersonville, Indiana for handling very large volumes of questionnaires. NASS has nothing to compare to that infrastructure and it was decided that NASS would contract the big "paper shuffling" tasks back to the Bureau of Census. This would include questionnaire mailing, receiving and checking in the returned questionnaires, and capturing the data from those questionnaires.

The following random facts will help illustrate just how "much paper" was handled by the Jeffersonville facility for the 1997 Census of Agriculture.

First mailing	3.2 million questionnaires
Second mailing	1.3 million questionnaires
Third mailing	0.8 million questionnaires
Returns	2.7 million questionnaires (86.0 %)
Peak number of Jeffersonville staff	292
Peak number of data entry staff	126
Avg. questionnaires keyed/week	88,000
Total pieces of printed material	15.2 million

NASS decided to move certain data collection functions to the field offices. All follow-up for operations that did not respond by mail would be handled there. Some contacts were handled in person. However, a significant amount of data was collected by Blaise CATI. Whereas CATI was handled in the Bureau calling centers in 1992, when we looked at our options for 1997, the potential offered by our field offices in combination with the Blaise software were seen as definite strengths.

Once the decision was made to conduct the CATI work for the Census in the field offices, it was imperative that the local area network hardware be upgraded. NASS purchased two new Dell file servers for each field office. Additional Pentium workstations were purchased in order to assure that the low end workstation for CATI was at least a 486-33. In reality, for evening calling most interviewers would have Pentium workstations available.

The processes that followed editing, from analysis to publication were largely handled in the field offices, where the local expertise was most valuable. After all, these offices have a knowledge of the agriculture in their state that could never be equaled by the staff in HQ. Even here, the Bureau's computer systems in place for analysis and publication were used by NASS. Developing new systems was out of the question, due to time constraints. Essentially, the NASS staff was trained to work with the centralized Bureau processing systems. The NASS wide area

network was used to provide critical communication links between the Bureau computers and the field offices.

The following table provides some details on the Blaise CATI applications, planned and/or developed, that were part of the NASS effort on the 1997 Census of Agriculture.

Blaise CATI Applications
Developed by NASS for 1997 Census of Agriculture

Name	Expected time frame as of 6/98 frame	Actual time	Specifications Set	Instrument Used	Completed CATI Interviews
Screener		10/97-12/97	10/97	screener	238,000
Advanced Follow-Up		2/98-5/98	2/98	main census	56,000
Large Farm Follow-Up	3/98-6/98	3/98-6/98	2/98	main census	20,000
Not on Mail List (NML)	3/98-4/98	3/98-4/98	2/98	main census	1,000
Low Response County	4/98-6/98	canceled			
Last Call Follow-Up		4/98-5/98	4/98	main census	51,000
Classification Error Survey	4/98-9/98	4/98-9/98	2/98	CES	20,000
Non-Response Survey	4/98-10/98	4/98-7/98	2/98	NRS	20,000

In the column labeled "Expected time frame," there are no dates listed for the Screener, Advanced Follow-Up, or Last Call Follow-Up. This reflects that these applications were "birthed" as the process took shape. Lead times here were minimal. In some cases, an application would be needed in the field offices within a few weeks of learning that there would be such an application. NASS also was determined to conduct this Census in a shorter time frame than had ever been done before, and it was common for the schedule for applications to be accelerated. Fortunately, by structuring the applications in a modular fashion and enforcing certain standards across all Census CATI applications, we were able to react quickly to the challenges posed by the fluid nature of the specifications.

The ever accelerating schedule also created an environment where minimal training was possible. In fact, the only opportunity that we had to present our plans to field office staff in person was a one hour session in September 1997. If you glance at the above table, you can clearly see that specifications were far from firm at that point and two applications had not even been thought of yet. All we could provide at that time was a rough overview of how we thought things would unfold. We supplemented that training later on with a two-hour teleconference. Again, that training was very limited. The only other way we had to pass information to the field offices was a written document that accompanied each application when it was delivered to the field offices. Yet, despite this limited training that the field offices received, most of them coped very well with the Census CATI applications. We give credit to the intuitive nature of the user interfaces of the Blaise system itself, the rigid adherence to NASS standards where possible, and the sophisticated utilities which allowed us to build "user friendly" interfaces for interviewers and statisticians.

Using Coding Results to Influence Routing

The challenges presented to the Blaise developers were also technical. Four of the applications used the instrument for the main Census questionnaire, which was 11 pages long. The interviews could take up to an hour. The most challenging aspect of the instrument was that it had to be able to handle any of almost 300 crops. Trigram coding was used to build a list of the crops grown on the operation. Since there is a shorter list of possible livestock commodities, we used a more traditional method to build that list. The interviewer would then review the profile of the operation and if needed, refine the profile. Once the profile was set, the detailed acreage, production, and sales questions were asked for each crop, and inventory and sales questions for each livestock commodity.

Our usage of Blaise coding programs had been very limited. Using them during the CATI interview would clearly be a totally new experience for our interviewers. Our interviewers embraced this new technique more enthusiastically than we had ever anticipated. We expect that this approach will find its way into other NASS applications in the future.

Conducting the Agricultural Census in the United States required some clever programming to insure robustness within the Blaise instrument as well as the system that supported the application. From the standpoint of maintenance and support, having one instrument for the main Census questionnaire for all states was the only feasible alternative. At first glance, this may not seem to be much of a challenge. But looking more closely, one soon realizes the diverseness of the data to be collected across such a large population. As mentioned earlier, one scope of the Agricultural Census was to collect data on all crops grown in the United States. Each state had a specific list of crops that would typically be grown there, but very few states had the same list. Each state's list would be a subset of 294 unique crops which were identified prior to the survey. To avoid asking embarrassing questions like pineapple crops in Alaska, and to keep the interview length as short as possible, the instrument was designed to first poll the respondent of the crops that they had raised. Once this list was prepared, more detailed acreage and production questions for each of these identified crops were routed.

The trigram method of lookup was used on an external file that contained 296 records: the 294 unique crops, one "unknown" category, and one "done" indicator. In addition to the crop's name, abbreviation, code, and description, the external file included such data as the item code, crop category, whether the acreage was collected in whole numbers or tenths, the conversion to bushels factor, the unit of production, and the point value based on the value of production of the commodity. Therefore, once the list was built for the crops grown by a respondent, the instrument had everything it needed for the identified crops. The crop description and the crop abbreviation were defined as the trigram key.

The instrument was designed to hold up to 159 crops in nine categories. Each of these crop categories was defined as a table, and as the respondent identified the crops they produced, parts of the rows of the corresponding table were being computed from the data in the external file. Also, as a particular crop category was identified, LOCALS were being computed that would be used to determine whether each table would need to be routed, and other LOCALS were being incremented, keeping track of how many crops in each category had been reported.

Upon collecting the list of crops from the respondent, the appropriate tables were routed based on the LOCAL, and the acreage and production data for those crops were collected. The tables were processed from one up to the number of crops within that category. This helped keep the respondent on track, as data for crops within the same category were collected together, even though they may not have been reported by category.

Dynamically Creating Blaise Records with Maniplus and Manipula

Another technical challenge was presented by NASS's first experience with toll-free telephone numbers. Every Census questionnaire, including the Screener, gave the respondent a toll-free number that they could call with any questions. Staff in the field offices were trained to handle the inquiries posed by the incoming callers. On the screener application, since the interview was very short, it was decided to attempt to complete the interview once we had them on the phone. We had never dealt with "incoming" calls before and it took some effort to work this out.

Each state office conducted their own CATI portion of the Screener, whose main objective was to determine whether the selected unit was a potential farm, a non-farm, a duplicate, or an undeliverable address. The records with a Screener Outcome of potential farm, would later receive an Agricultural Census questionnaire. The Blaise instrument, developed to collect this data, used a series of screening questions to determine the outcome. In addition to the CATI samples, approximately 240,000 screener postcards were mailed, which included the toll-free telephone number. At the start of the survey, all of the toll-free calls came to our North Carolina field office. The mailed portion of the survey was treated as a separate application, although the same Blaise instrument was used to collect the data. The postcards that were returned by the respondent were scanned in at our North Carolina office, if there were no changes to the name and address.

The survey design had to allow for the cases of the mailed portion to be collected over the phone via the toll-free number, as well as entering the name and address changes. However, initializing all 240,000 cases into a Blaise data set would have been tremendous overkill, as only the forms with name and address changes and the call-ins would actually need a Blaise record. The case would also need to be retrieved in a timely fashion, so the respondent would not be lost. Using Maniplus, Manipula, and a Pascal tool, a system was designed that would dynamically build a Blaise record. The following describes the steps of the Maniplus setup.

First, the user was prompted for the identification number of the form to be retrieved or created. Second, the SEARCH method was used on the Blaise data set to see if the form already existed. If the form was not found, then the appropriate state ASCII file was searched using GREP, which is a Pascal tool that quickly searches the contents of a file for a particular string. The names and addresses of the 240,000 cases were split up by state, into multiple ASCII files with the corresponding state's number as the extension. This enabled the GREP search to find the record very quickly. The two-lined result of the GREP search was stored in a file. The first line was the name of the file that was searched, and the second line was the entire line that contained the id. Third, a Manipula setup was called that would update the Blaise data set with a new record containing the id, name, and address information from the GREP-created ASCII file. In order to process only the second line of this file, the computations and WRITE statement were enclosed in an IF-ENDIF construct using the RECORDNUMBER method.

The remaining steps of the Maniplus setup were the same, whether the Blaise form was newly created, or already existed. The GET method was used to retrieve the form from the data set. A dialog box (Figure 1) was then opened displaying the id and name fields and a check box used to obtain the Screener Outcome without invoking the Data Entry Program (DEP) and answering the series of screening questions. Four buttons were available in this dialog including one to accept the outcome entered, one to retrieve the form in the DEP, one to invoke a name and address update dialog (Figure 2), and one to cancel.

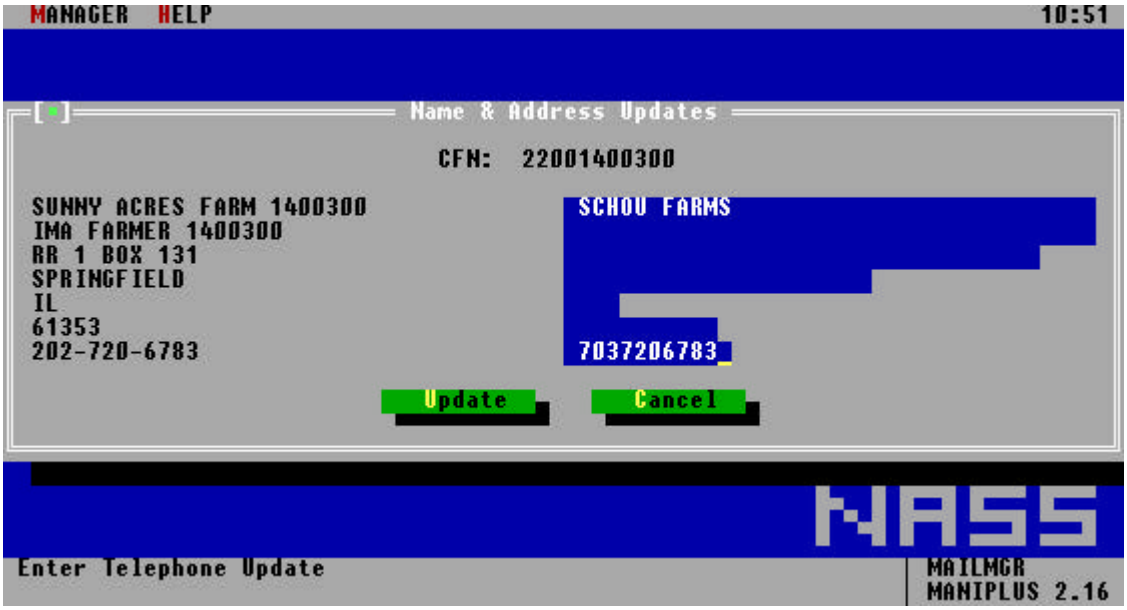


Figure 1 Name and address update dialog box

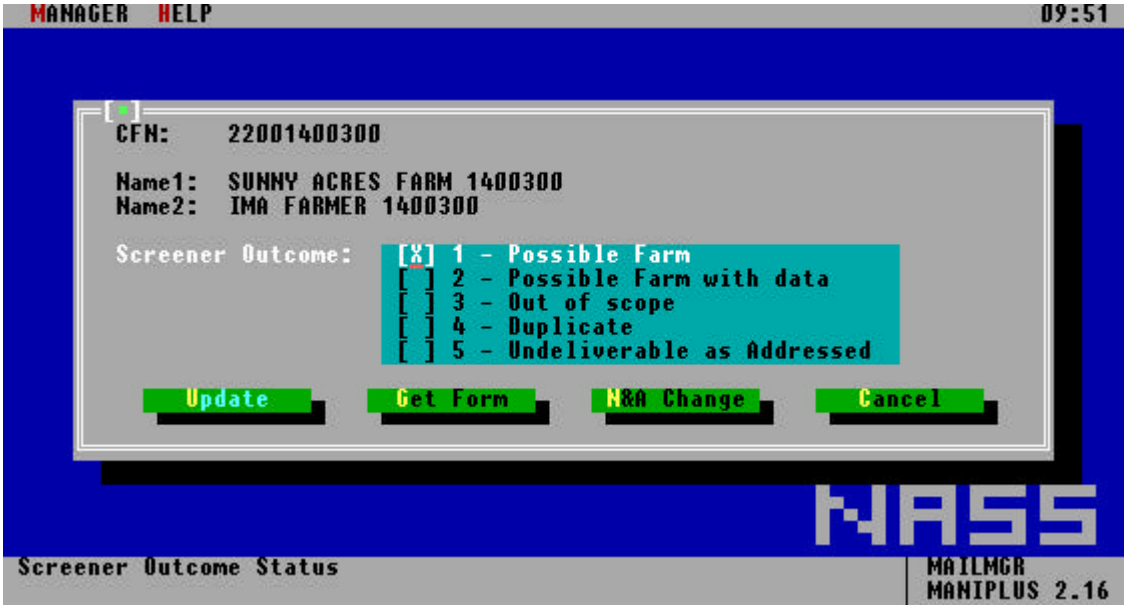


Figure 2 Screener outcome dialog box

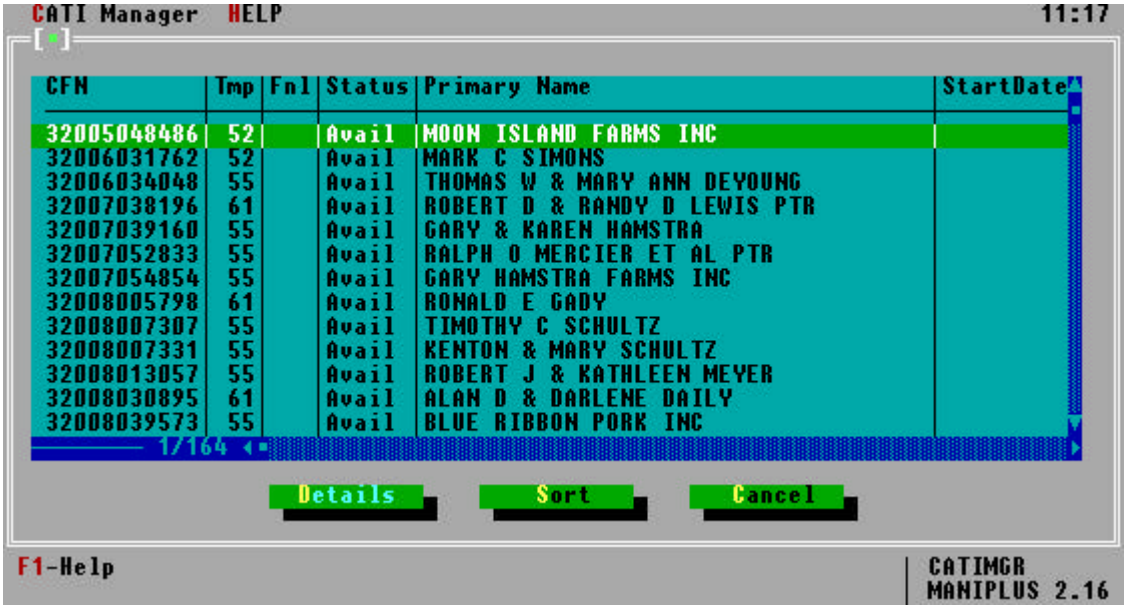
Using Maniplus as a Management Tool

All the Blaise applications used a couple of "outcome codes" to track cases through the system. A Final Outcome Code (FOC) was assigned when a case was resolved and ready for output. The Temporary Outcome Code (TOC) would be assigned if the case needed some type of special handling. This usually meant that the case could not be resolved by the normal CATI process. The "main Census" instrument was programmed to set these codes based on the results of a CATI call. A more challenging issue was building a user interface to offer the end users an efficient way to manage those cases with a TOC. We took advantage of one of the more powerful abilities of Maniplus, which was the ability to update certain key fields in the Blaise data set without starting the interview instrument. We had never done anything similar to this in NASS. The options that it opened up for us were invaluable.

Maniplus played a major role in many parts of the Agricultural Census. The Blaise call scheduler was used for the first telephone contact. Two Maniplus programs were developed to manage the special handling of forms with a TOC. One was for a second CATI contact and the other was for secondary non-CATI type of processing.

One of the specifications of the survey was to contact all refusals a second time, using a more experienced group of interviewers. Other examples of a second CATI contact included language barriers, claims by the respondent that the form was completed and returned but no form was received after a two-week period, and discovery of new correct phone numbers. Maniplus was used to develop a point and shoot interface (Figure 3) which could be sorted by the TOC. This grouped the similar types of contacts and allowed the experienced interviewers to know some history of the form. Upon selecting an id from the list, a dialog box was displayed with other pertinent information. From there, the interviewer could click on the Retrieve Form button, which would invoke the DEP, and the interview could be conducted.

Figure 3 Point and shoot interface



The other Maniplus program was geared for the statistician's use. It provided a similar point and shoot interface, but included a different subset of forms. This list could also be sorted by TOC. Examples of forms included in this list were second refusals, disconnected or incorrect phone numbers, six attempts to contact the respondent had been made, inaccessibles, and reported duplication. As with the secondary CATI interface, once an id was selected, a dialog box (Figure 4) was displayed with pertinent information. However, this dialog box used CONTROLS that could be edited. An FOC could be assigned, thus completing the case, without actually invoking the DEP.

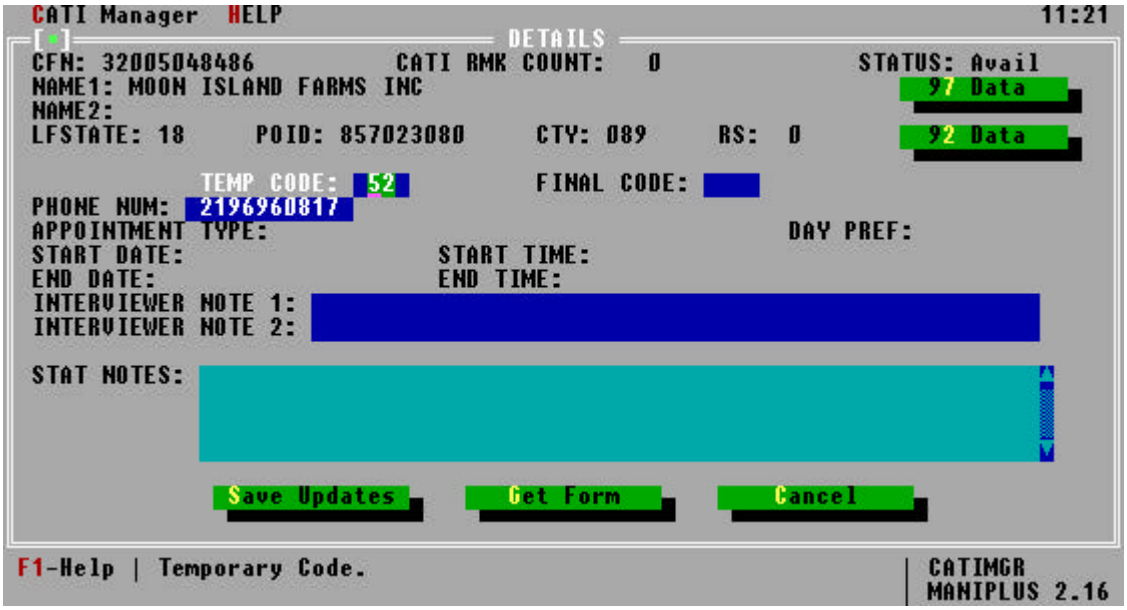


Figure 4 Dialog box with controls

Using Maniplus when Populating the Sample

On our more typical surveys, the name and address sample file is populated into a Blaise data set in its entirety. The Ag Census name and address files, however, were not as complete compared to what we are normally accustomed. Many records were missing phone numbers that would need to be looked up before populating them into the data set. Also, some of the applications had large samples of which only subsets would need to be populated at one time. For these reasons, it was evident that our procedures for populating a data set would need to be changed.

A Maniplus setup was written to handle the populate process. It did not use menus. It simply started with a dialog box prompting the user for an answer to the following question: "Are you adding records to an existing data set, or are you initializing from scratch?" If the user chose to start from scratch, two warning dialogs were invoked if the data set already existed. As a safety net, a batch file was run from Maniplus to make a backup copy of the existing data set. The user was then prompted as to whether all records in the provided ASCII input file should be populated, or just the records with a telephone number.

Before adding these forms to the Blaise data set, a procedure was invoked that assigned the appropriate time zone to each of the forms. To determine the time zone, this procedure first looked at the state field. Some states required that the county field also be considered to determine the time zone. In a handful of states, the county field was not sufficient, and so the zip code field was also examined. Parts of a couple of states do not take advantage of daylight-saving time, so the current date had to also be considered when time zones were assigned.

If the user chose the initialize process (versus the appending process), up to 200 forms were initialized into a practice data set. The names and the phone numbers were changed before the

data set was written. Interviewers can be trained prior to, as well as, during a survey using the live instrument with a practice data set. This newly created data set was also zipped for potential use, in the case of the 200 practice forms becoming exhausted.

Conclusion

As we look back on the uses of Blaise CATI on the 1997 Census of Agriculture, the one word that comes to mind is change. In the summer of 1997, we were planning for five CATI applications. Now as we look back on the completed effort, we can count seven. Four of the original five did occur and three other applications were born as the data collection effort evolved. The startup dates for some of the applications were also accelerated. Looking back on the CATI applications for the Census, it is amazing that we were able to consistently meet our target dates. Despite the number of new applications and ever accelerating schedules, the applications were up and running on time. Maybe even more remarkable is that almost no updates were made to the applications once they went into use. They worked from day one. Two factors can be credited here. First, the staff building and supporting the applications were extremely talented, experienced and dedicated. Many long hours, more than just a few off the clock, went into making this a success. The second factor is the flexibility and power of the Blaise system. We were able to build sound instruments that met almost every technical challenge. Utilities such as Maniplus allowed us to build user interfaces that could accomplish many tasks without even involving the instrument. The combination of a strong software in the hands of a gifted staff spawned success out of chaos.