

The Transition from Cases and BLAISE 2 to BLAISE III at the National Agricultural Statistics Service

Asa Manning, National Agricultural Statistics Service, USA

1. A History of the Conversion Effort

NASS was one of the first organization to use Blaise III for production data collection and editing. Blaise 2 had been used, first for interactive editing and then data collection on several smaller surveys beginning in 1992. A pilot project to study the feasibility of using Blaise 2 for our major surveys had been initiated in three states (CO, IN, & WY) in September 93. The goal was to produce both data collection and interactive editing instruments from the same source code. In addition, the interactive editing instrument would not only be used to edit data collected in Blaise, but data keyed from paper questionnaires.

This multiple mode concept was fairly successful in Blaise 2. However, the instruments for our largest applications exceeded the capacity of the software. In addition, NASS had identified a number of enhancements that would be needed in the CATI system before pursuing a larger scale implementation. So NASS was waiting anxiously as the initial versions of the Blaise III 1.0x series were released.

The first production use of Blaise III took place in March 1994 in the original three pilot states. The freedom from the memory limitations of Blaise 2 was an obvious advantage. A number of significant bugs and limitations were encountered, but with some real time adjustments in our procedures, the survey was completed on time. The promise of the software was very enticing, but the amount of work left to be done was daunting.

We communicated the problems that occurred during the survey to Statistics Netherlands. NASS requested that they make a visit to one of our three pilot states, prior to the next survey. To Statistics Netherlands credit, they did. Lon Hofman, Mark Pierzchala, and Asa Manning spent a week in the Indiana SSO during April 1994 repeating the March experience.

Those days were key to the survival of our pilot study in NASS. First, NASS management needed to see this type of commitment from Statistics

Netherlands. Secondly, Lon was able to identify and eliminate many bugs from the software. He also was able to offer us advice on certain issues related to our Novell Local Area Networks (LAN). The spirit of cooperation gained from that visit helped get us through what was to be a long road ahead. Little did we know at the time, but it would be over two years before the release of a version of Blaise III that would finally provide almost all of what NASS needed.

The pilot study continued with each quarterly Agricultural Survey through September 95. The first step to expand into additional states occurred in the July 1995. The CASIC Group spent a week in St. Louis, Missouri, training 13 states (12 new states and one of the original pilot states). Two representatives of each state were trained to manage a Blaise III survey in their office. The agenda included sessions on setting up a Blaise survey on the state LANs, working with the CATI management system, managing the flow of data from data collection to editing, using interactive editing, and managing the overall survey process. Each participant was then charged with returning to their respective office and transferring the training to the other staff.

The states attending that first workshop were states involved in NASS's Cattle on Feed program. This application was slated to be downsized in late 1995, with Blaise III as the centerpiece of data collection and editing. PC-SAS on the LAN was to be used for post-edit analysis and summary. This application was scheduled to change for October 1995, but was delayed until January 1996 due to internal changes to the survey program. Thus, the first survey conducted by states trained during the workshop, was the December Agricultural Survey. Fourteen states used Blaise III. Texas was delayed for internal reasons for several quarters. There were some problems encountered, primarily because this is NASS's largest and most complicated CATI application. Some of those challenges still remain and those will be mentioned later.

Another group of 12 states (including two original pilot states) was trained during November 1995. Three of those new states started using Blaise III in January as the Cattle Survey debuted in Blaise III. This brought our total number of states using Blaise III to 17. The experience on this application was improved as states began to build on their December experience, and benefited from the smaller, less complex nature of the Cattle survey. The other seven new states used the March Agricultural Survey Application, bringing the total number of states with Blaise III experience to 24.

We delayed further expansion until the summer of 1996. This allowed the newly released Blaise 1.1x series to mature. The upgraded software offered NASS many needed enhancements. CATI Management was upgraded with many additional features for the telephone supervisor. Manipula was enhanced with an integral check and other features which allowed NASS to dramatically speed up many processes. The new Manipulus software offered NASS the potential to develop a more user friendly interface for the statisticians using interactive editing. Basically the Blaise III 1.1x series included many of the features that were in the original Blaise 2 software.

During the June Agricultural Survey, NASS took a long awaited step, when Blaise III 1.1i was used in two states (MT & KS), while Blaise III Version 1.05 was used in the other 22 states. The enhancements to version 1.1i were well received, but some signs of minor bugs existed. Our user interface for interactive editing was much preferred to the standard system interface, but slow form retrieval was identified as a problem. The positives far outweighed the negatives, so NASS went forward with plans for further expansion.

A third workshop was held for twelve new states during July 1996. This brought the total number of trained states up to 37. Version 1.1j was used at the workshop and shortly thereafter distributed to all states using Blaise III.

All those states used Blaise III on the September Agricultural Survey (including Texas). The improvements over version 1.05 were very popular with the end users. The states finally had a system with almost all of the features they had been looking for.

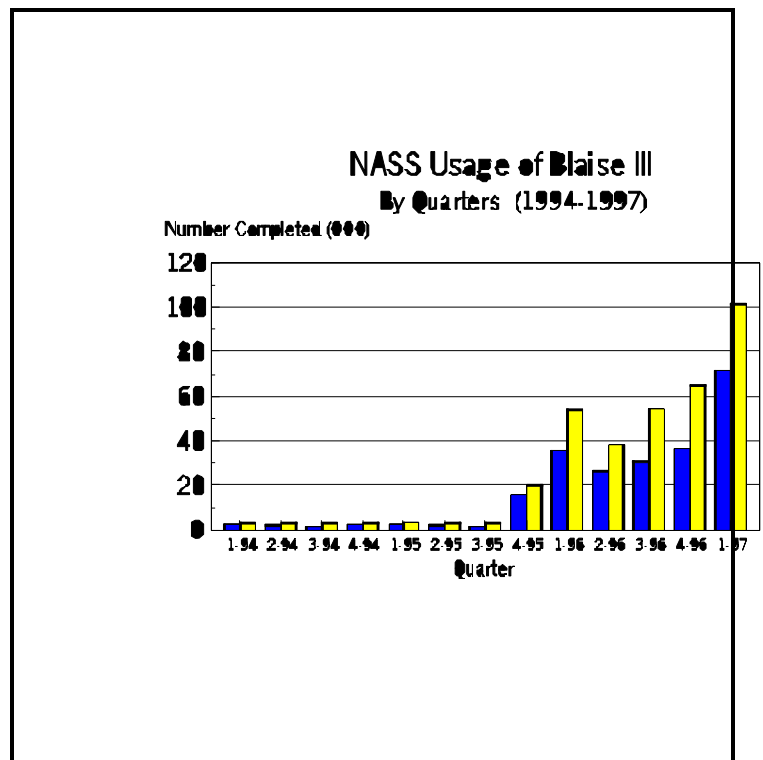
A small workshop was held in November 1996 to train five more states. This left only Nevada untrained. Version 1.12 was used at this workshop. It was distributed to those five states soon thereafter, so they could prepare for the January Cattle and Sheep surveys, when they would use Blaise III operationally for the first time.

The December Survey 1996 was conducted with Blaise in the same 37 states as September. Most states were very pleased with the results. In states with the largest instruments and largest samples, we continued to observe some performance type problems. Texas could not support as many interviewers as needed, without adversely affecting question to question performance. Other states may have seen signs of this as well. The slowness of form retrieval in Maniplus contributed to the staff in Texas being unable to complete all editing in Blaise. The need to "touch" each warning error to suppress it, proved to be too time consuming when samples were large.

During January 1997, the Cattle and Sheep applications (running Blaise III version 1.12) were used in 42 states. The smaller instruments reduced the impact of the performance related issues and resulted in the most successful use of Blaise III thus far.

NASS plans to train Nevada in the next few months. That will bring all the states on-line. During this time frame the last CASES applications are also being replaced.

Figure 1 gives a graphical representation of the quantity of data being collected and edited with Blaise III in NASS.



It has been a much longer road than we expected when we first started this effort over three years ago. There have been occasions when NASS wished it had not been on the "bleeding edge" of the Blaise III software evolution, but we feel the software would not be where it is today without our input. To be honest, if NASS had started using Blaise III later than it did, there would still have been bugs for us to contend with. NASS puts pressure on many pieces of the system and uses the software in some unique ways. Thus other organizations would not have found all the bugs that NASS did. This is proved by the bugs that other organizations found even after NASS had used the software for several surveys. The shakedown cruise for any software is a painful process, and the smaller the number of users the more pain each one will probably feel. It is nice to be looking back on most of this process, at least for the DOS version.

NASS has converted a large number of applications from CASES or Blaise 2 to Blaise III. Data collection and interactive editing instruments are generated from one copy of source code for all these applications, except June Area. The June Area Survey is a large area frame survey where CAPI was tried, but has been discontinued. Each editing instrument must be capable of handling both data collected with Blaise and data entered from paper questionnaires. The following table provides information on each application.

Blaise III Applications In Use At NASS

The above table illustrates the large number of applications (16) that NASS has converted to Blaise III. With a full implementation of Blaise III nearly at hand, the volume of data collected with Blaise could approach 400,000 in 1997. Interactive editing volume could reach about 600,000. Another factor that bears mentioning here is that each of these applications is running on LAN's in the various state offices. This adds significantly to the support demands because there are subtle differences in LANs from location to location and in the staff expertise as well. Because there is hardly a month that goes by without a survey and in some months there may be up to four concurrent surveys, the support demand is virtually constant.

2. An Update on Computer Assisted Personal Interviewing (CAPI)

At past conferences, NASS's Computer Assisted Personal Interviewing research was discussed. During the last couple of years, the research was limited to work with the June Area Survey and major multiple frame surveys (Agricultural Survey, Cattle, Sheep, & Hogs) in Indiana. On the multiple frame surveys the laptops were used for personal interviewing, as well as telephone interviewing from the field enumerators home. NASS referred to this as CATHI (Computer Assisted Telephone Home Interviewing), to distinguish it from CAPI. This effort proved that Blaise was a capable tool for collecting data on a laptop computer in either mode. It was also proven that the technology could support the electronic transmission of data into the state office.

In late 1996 upper management in NASS decided to terminate further CAPI efforts. There were four major reasons.

1) NASS's survey program does not provide the type of surveys that lend themselves to CAPI. Most surveys are of very short duration (less than two weeks) and as documented above are frequent in occurrence. The overhead of supporting CAPI increases significantly per-survey unit in this environment. There ends up being an almost constant support

demand on the state office staff. More surveys with a longer period of data collection would make it easier to justify CAPI/CATHI at NASS.

2) The cost of laptop computers has increased from about \$2000 early in the study to over \$3000 currently. The demand for low end machines was not sufficient to make it economically feasible for vendors to produce them. With potentially between 1000 and 1500 field enumerators to equip with laptop computers, the start-up costs for CAPI are simply too high at this time.

3) NASS has given priority to upgrading its workstations in the state offices for Blaise and other applications. Servers and LAN cabling are out of date and also need improvement. Money is also being invested in a Wide Area Network to connect the field offices and HQ. This will provide the communication backbone for future technologies. These competing initiatives made it even more difficult to find funding for CAPI.

4) Support for CAPI would be more than most field offices could handle. Indiana had a very skilled staff that was determined to make CAPI work. Yet, even in Indiana, there were times when things did not go smoothly and the stress on the staff took its toll. Essentially, if Indiana was finding it a challenge, others would likely be overwhelmed. It might be most efficient to support CAPI from certain regional sites, but this would require some structural changes that simply were not going to be dealt with now.

NASS may revisit CAPI in the future. Changes will have to be made in several areas before this occurs.

3. How Well Is Blaise III Doing ?

As chronicled above in the recent history of Blaise III usage in NASS, there has been much progress in the evolution of the software. We are convinced it is the most flexible, versatile survey processing software in the world. However, there are still areas that need to be addressed by NASS, Statistics Netherlands or both.

It has been NASS's goal since it first starting using Blaise to produce instruments for data collection and editing from one source code. This offers the potential advantage of cutting down total maintenance time. If data can be cleaned in Blaise IE, then other post-edit processing can move to the LAN as well. NASS has been very successful in completely downsizing all of its list frame survey applications. However, we have not been as successful on the more complex multiple frame surveys. The administrative coding is much more complicated on these applications, as is the CATI routing. The extra level of complexity when built into the instrument appears to be leading to some flakiness in the IE instruments. On rare occasions, a situation which should trigger an error in the editing instrument does not. This can not be traced to a logic problem, as the same exact error was caught on many other occasions. Even though this is happening on a small number of records, the impact on subsequent processes can be very severe.

NASS plans to explore the possibility of separating the data collection and interactive editing instruments. This would reduce the size of the data collection instrument about 30 percent and the interactive editing instrument more than 50 percent. By simplifying the logic in many error checks we hope to create a more stable interactive editing instrument. Since the checking mechanism is a major player in performance, we may also see an improvement there as a side benefit. As has been mentioned earlier, slow performance continues to be a problem with the integrated CATI/IE instruments in certain situations.

NASS uses Manipula a great deal in its Blaise applications. Many of the past support calls from the states dealt with Manipula failures. The number of failures increased noticeably when Dell pentium workstations running at 166 mhz began to be used. As it turns out, a bug in the system was creating occasional run time errors on pentium workstations. Our testing on the faster Dell workstations with version 1.15 looks positive. If these workstations prove to be reliable, it will reduce the time wasted when failures occur. Since many of the Manipula processes are potential bottlenecks, running them on the fastest workstation available has obvious advantages. Hopefully, we will be able to give a good report on the March results, when this paper is presented in Paris.

4. Agricultural Census - A Challenging New Application For NASS

An Agricultural Census is conducted in the United States every five years. This survey attempts to contact all farm operations and collect information about acreage, production, and value of sales for crops; production and sales for livestock; and other data relating to agriculture. Survey results are released at national, state, and county levels. Since this survey is close to a complete enumeration of the agriculture sector, NASS uses this data to "true up" its own estimates.

The Census Bureau, an agency of the Department of Commerce, conducted this survey through 1992. In 1995, the Department of Commerce came under scrutiny for elimination or downsizing. Although the Commerce Department survived, funding for future Agricultural Censuses was not included in their appropriations. NASS stepped forward and asked to have the appropriations added to its budget. Congress passed legislation in late 1996 completing the transfer of responsibility for the Agricultural Census to NASS. This brings virtually all surveys of the agricultural sector under NASS's control.

This survey is much larger than any survey that NASS has done before. Initial mailings in December of 1997 will total about 3.5 million questionnaires. NASS does not have time to put the infrastructure in place to handle the 1997 survey. NASS will contract many tasks on the 1997 survey back to the Census Bureau. The Census Bureau will be heavily involved in handling the processing of paper questionnaires, including mailings, checking in, and keying. NASS will handle all the CATI follow-up and editing. All CATI instruments will be developed in Blaise III. The existing interactive editing system developed by the Census Bureau, will continue to be used in 1997.

There will be three Blaise instruments developed for 1997. The main Agricultural Census Instrument will be used to contact large farms that do not return a paper questionnaire and to follow-up in counties where the response rate is inadequate. The Coverage Evaluation Survey instrument will be used to follow-up on operations that did respond and will gather information to identify duplication or operations that might have been misclassified. The Non-response survey instrument will be used to quickly phone non-respondents and determine whether the operation is in or out of scope.

The main Agricultural Census Instrument will present NASS with several challenges. Each is detailed below :

1) The instrument will have up to collect data on up to 900 items and CATI interviews can last over an hour. However, most operations will probably have less than 100 items. To identify the commodities produced by a given operation, NASS plans to use a coding technique to first build a list of commodities. Once the list is complete, detail acreage, production, and value of sales questions will be asked for each commodity in the list. Our goal is to expedite the process of collecting the crop commodity data, while not missing any. NASS has never used the Blaise coding modules, so this will be a learning experience.

2) The instrument will be one of the biggest that NASS has put together. Question to question performance is a concern. It will help that the instrument is interview only. Thus the logic for checks will not be cluttered with coding for interactive editing or paper forms. To help meet this challenge, the instruments will be designed to maximize instrument speed. NASS is also upgrading to state of the art file servers (currently 486) in the field offices during 1997. States still using pre-10BaseT LAN cabling technology will be upgraded to 10BaseT. Additional pentium workstations will also be purchased, so that the bottom end CATI workstation will be a 486 running at 33 mhz.

3) The low response county follow-up will require that NASS complete enough CATI interviews to bring the response rate in those counties to a pre-determined level. This will allow NASS to use the quota feature of Blaise CATI Management for the first time.

4) The typical NASS CATI survey lasts no more than two weeks. The data collection period for this survey will last about four months. Field offices will be able to offer interviewers a longer period of active employment than in the past. NASS must also be aware for the first time that "interviewer burnout" may become a problem. We are also concerned whether the field offices can find a sufficient number of interviewers to meet our needs over such an extended period. The field offices will have to be innovative in their hiring, training, and management of the interviewers.

The CASIC Group in NASS will be kept very busy as we prepare for these applications. We are excited about the unique challenges that they present us.

5. Where To Now For Blaise III ?

As NASS wraps up its conversion from CASES and Blaise 2 to the DOS version of Blaise III, we are basically satisfied. However, there are issues that remain and can hopefully be met as Blaise III moves to Windows. Three major items follow :

1) One of the complaints that we are hearing from the field offices is that it takes too long to edit forms. The primary step that is consistently criticized is the need to touch every soft error to suppress it. NASS would like the option of displaying a listing of all soft checks in a pop-up window. The editor could scroll through the list and click on any message they wanted to review in detail. Once all messages were reviewed, the editor would be able to suppress all warnings with one key stroke from the pop-up window. NASS also needs to continue to minimize the number of micro edits, but that is a cultural change that takes time.

2) NASS would like to see the Blaise software be more open. The proprietary nature of the Blaise format presents barriers to users that need access to the data. Unless they are capable of writing Manipula programs, the data must be converted to another format before they can use it. This conversion itself, should be unnecessary. Blaise III should be able to read and write to other formats. If we could access XBase and SAS files directly, many cumbersome ASCII transactions could be eliminated. We suspect most users of Blaise III could benefit from this ability, and suspect that it is a top priority for many organizations using it.

3) NASS would also like to see the data collection and/or editing software run in a client server mode. The instrument could be located on a centralized platform (server), and users in the field offices would be able to access it. The centralized location of the instrument appeals to NASS, because it reduces the difficulty of supporting applications running in a distributed manner. NASS has taken the first step in this direction by connecting the field offices and headquarters LAN in an Wide Area Network (WAN). Client server is viewed as a longer term goal than the other two.

We have presented a chronology of NASS's experience with Blaise III. From the original pilot study to the planning for the 1997 Agricultural Census, Blaise III has become an integral part of NASS's survey processing methodology. It has not met all of our needs, but it is doubtful that any software could, given NASS's diverse survey program. Hopefully, as Blaise III moves to Windows, the remaining challenges can be met.

