Integrating the use of Data Centers into the NASS CAI Structure

Roger Schou, National Agricultural Statistics Service, USA

Over the last several years, the National Agricultural Statistics Service (NASS) has begun to have difficulty hiring CATI interviewers in several of our field offices. The wages that we can offer to prospective interviewers in cities like Denver, Colorado; Minneapolis, Minnesota; Lincoln, Nebraska; and Nashville, Tennessee are not competitive. As a result, the use of CATI has dwindled to virtually nothing in those and other field offices. On the other hand, in offices located in smaller cities with lower costs of living, hiring interviewers is not a problem. Therefore, NASS has opened a Data Center in Cheyenne, Wyoming with 30 computers dedicated to CATI data collection. They also have an additional 22 computers that are used mainly for training, but 11 of them could be used for data collection, as there are 11 phones in the training room. NASS will open another official Data Center in Helena, Montana in early 2002 with 31 computers dedicated to CATI data collection, and 12 computers for training, which could also be used for data collection. Later in 2002, another Data Center will open in Louisville, Kentucky.

The Wyoming Data Center currently services four other states on a regular basis. The office in Montana services one state, performing Data Center calling activities utilizing their office computers.

In a typical state office, the CATI interviewers use the regular office staff's computers. There may be a few older computers dedicated to daytime interviewers. During the evening hours, interviewers are scattered throughout an office, sharing the desk space and computers of the professional staff. The interviewers are part-time employees and only work when the state is conducting larger surveys.

In contrast to the typical state office, our official Data Centers have and will have computers dedicated to CATI data collection. They are all in one general area, making supervision much easier. Although the interviewers are still employed on a part-time basis, they have steadier work because they are not dependent on one state's survey load. They are conducting surveys for multiple states. Since the workload is steadier, the level of expertise of the interviewers increases yielding higher quality data.

The majority of the Manipula setups used in the Data Center effort are generic from survey to survey. Therefore, coordination between Headquarters and the state offices is minimal, once they have received the applications. Most of the coordination is between each Data Center and their User States.

The basic functionality is relatively simple. The office with little or no CATI capacity (the User State) transfers cases selected for CATI to the Data Center. The Data Center collects the data for these cases and returns them to the User States. Staff in the User States then interactively edit the cases using Blaise. By returning the case to the User States for editing, we assure that all data in the state is edited in the same manner. A more detailed description of some of these processes will follow.

Each User State identifies the cases within their sample to be collected in CATI by the Data Center. Since the samples are usually available before the final version of the Blaise instrument, the ASCII files used to initialize the Blaise data set are zipped into a zip file. These ASCII files originating from the User States will only contain the cases to be collected by the Data Center. The zipped file is placed on the File Transfer Protocol (FTP) server and an e-mail message is sent to the Data Center's official mailbox. Each User State then initializes their entire sample into a Blaise data set. The cases corresponding to the ones sent to the Data Center are marked as non-CATI cases, by using a Manipula setup to compute a field in the management block. This field is then used to exclude these cases from the day batch.

Upon receiving the e-mail message from a User State, the Data Center moves the zipped file over the Wide Area Network (WAN). After unzipping the file, they concatenate the input files from the User State to the input files for their state. After they have concatenated each of the User State's files with their own, they initialize their Blaise data set, which will contain their own sample and all of the User States' CATI samples. Since all of the cases to be completed by the Data Center for a given survey are in one data set, managing the survey requires little additional effort on their part.

The zipping and unzipping processes and the e-mail message are all done within a Visual Basic menu script, so they are as easy as clicking a button. Currently, the only manual processes are the actual moving of the forms over the WAN, which is simply a "drag and drop" operation, and the concatenation of the User States' files with their own.

As the Data Center completes the User States' cases, they are sent back to the appropriate state on a daily basis. The Data Center clicks on the "Send CATI Completes Home" button on their CASIC interface, a Visual Basic interface. This process invokes a Manipula setup that creates a Blaise data set containing a copy of all completed cases that have not been previously sent. After it creates a copy of a case, it marks the process switch field in the management block in the Data Center's data to indicate that the case has been sent. The process continues by zipping the newly created data set and placing the zipped file on the FTP server. An e-mail message is then sent to the respective User States, notifying them that a new data set exists on the FTP server.

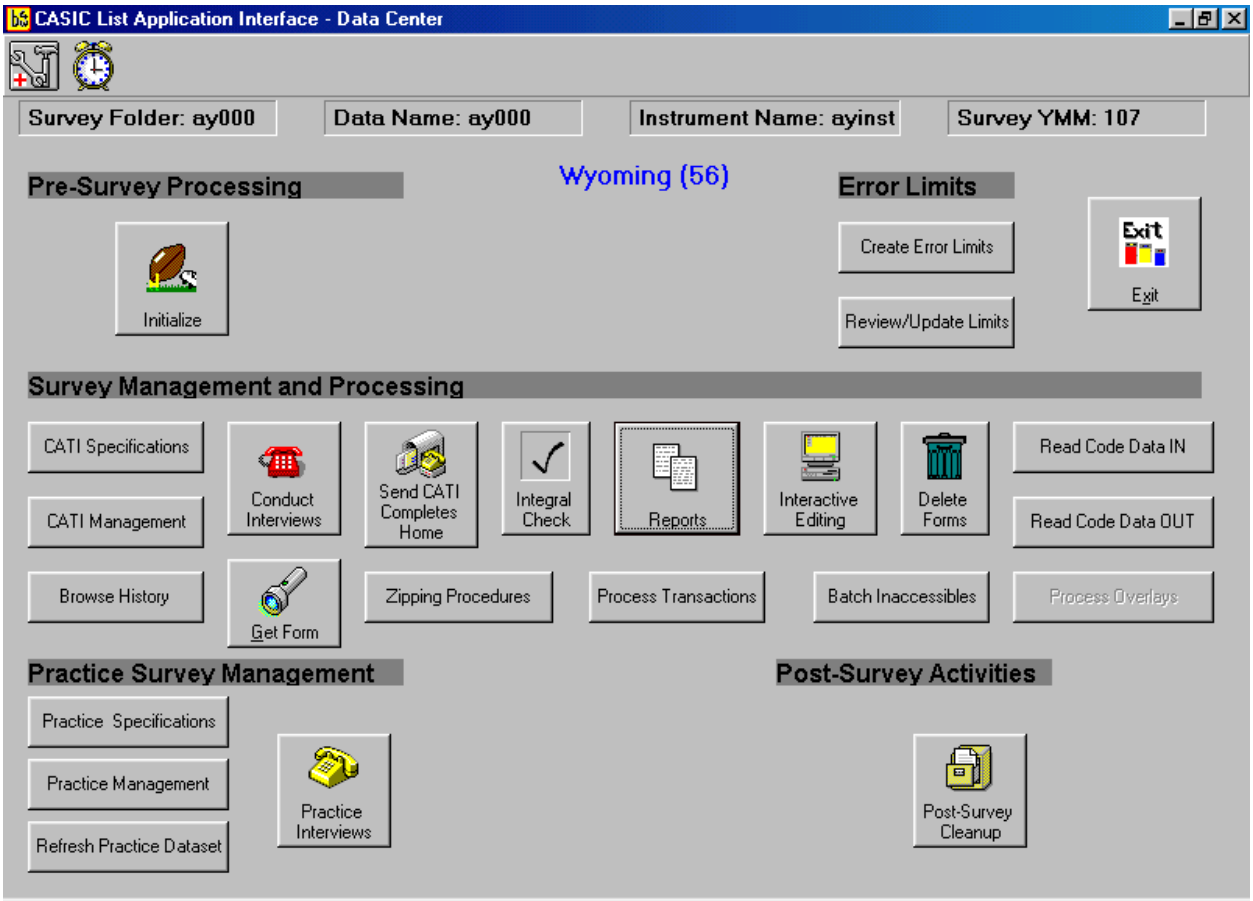

Figure 1 Data Center Interface (Visual Basic)

When a User State receives the message, they manually move the file over the WAN to a pre-determined folder. Then they click the "Merge Data from Data Center" button on their CASIC interface. This process unzips the data set into a temporary folder. Next a Manipula setup runs that locates the case in their data set and checks to see if it still has no data. If it has no data, the case is deleted from their data set, and then copied from the data set received from the Data Center. If for some reason, the User State had entered data for a case that was coming in from the Data Center, the Data Center's case is written to an overlays data set.

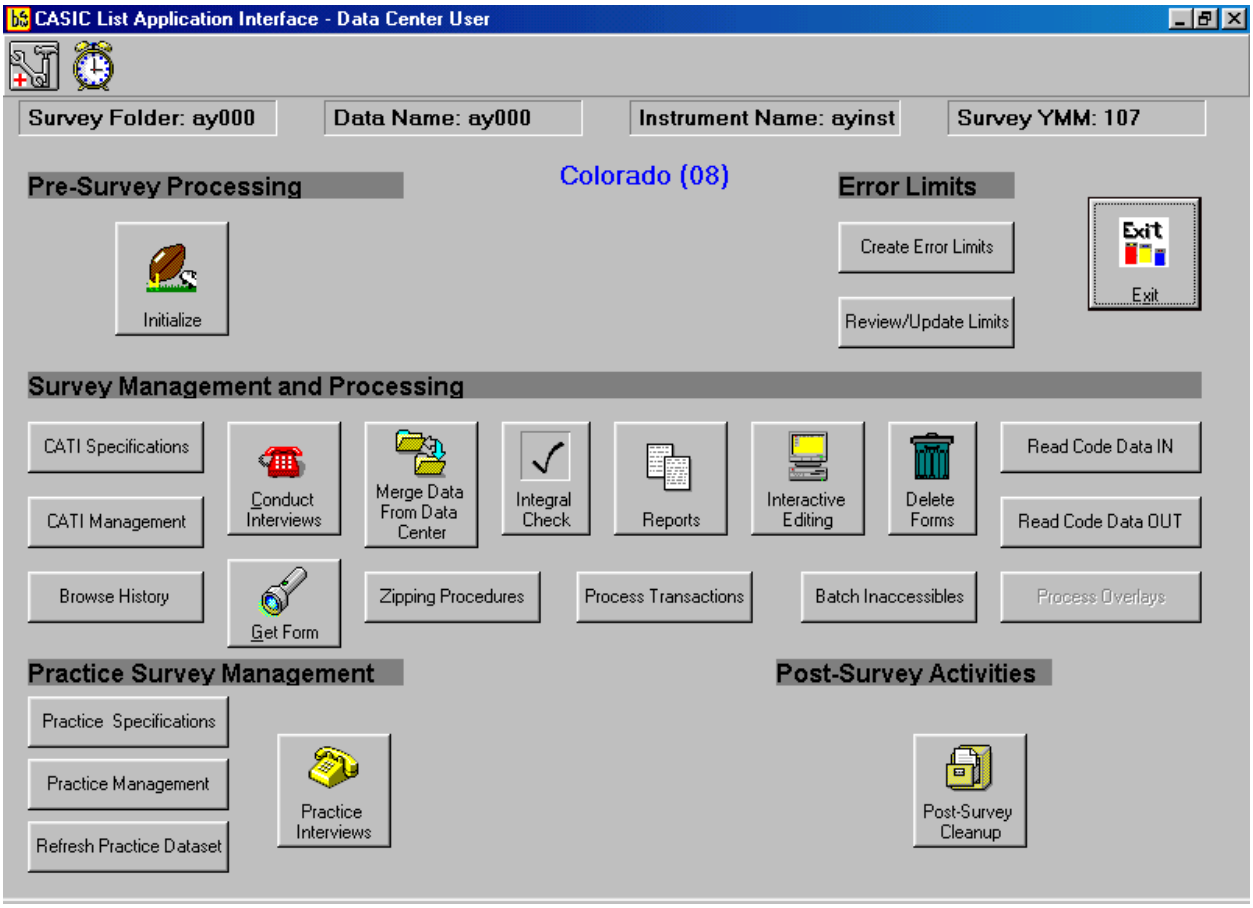

Figure 2 User State Interface (Visual Basic)

If an overlays data set exists, a ManiPlus setup can be invoked from the "Process Overlays" button on the CASIC interface. This setup will display a table of key fields for each case in the overlays data set. Upon selecting a case, a pre-determined set of fields from will be displays showing data from the case in the overlays data set and the survey data set. There are also buttons that will allow a user to start the data entry program for either case. Once a decision has been made as to which case to keep, the appropriate case is written to the survey data set, and the other one is deleted. Once a case has been resolved, it is removed from the table listing the key fields. This process can be repeated until the table is empty.

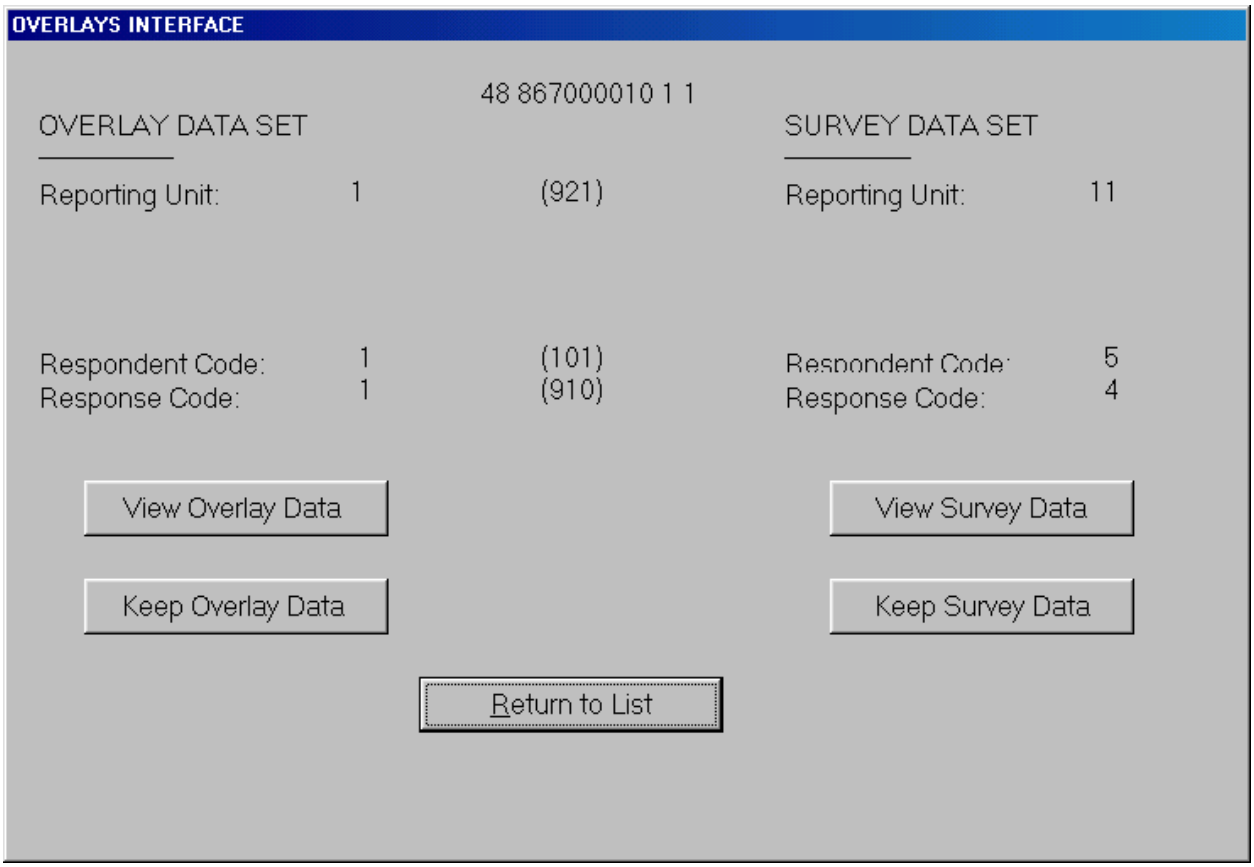

Figure 3 Overlays Interface (ManiPlus)

Except for the addition of the merge and zipping buttons, the User States' interface looks the same as it would for any other survey. All of the other processes run as they normally do in the User States, sharing the same set of Manipula setups. The interfaces for a typical survey, a User State, and a Data Center are very static, as seen in figures one and two.

The Data Center's interface also has just a few extra buttons (send completes, zipping, and a special report), however, many of the Manipula and ManiPlus setups running the other survey processes have additional code in them. This is due to the different handling requirements between a case that "belongs" to the Data Center's state, and a case the "belongs" to a User State. For example, the Data Center does not interactively edit a case that "belongs" to a User State. However, the data collection and interactive editing processes are all conducted in the same data set. The same data set that contains the Data Center's cases and the User States' cases. Thus, the ManiPlus setup, that runs the interactive editing process in the Data Center, filters out all cases except the ones that "belong" to the Data Center state. The same is true for the integral check and read out processes.

Many of the Manipula and ManiPlus reports the run in the Data Center were altered to either ignore all states but their own, or include extra reports for the User States' cases. For example, the status report, a Manipula setup that gives a report of all cases in the data set, normally provides distribution counts of cases to be completed and statuses of completed cases. If the survey being conducted has been designated as a Data Center activity, the status report gives the same information, but it is by state. It also has a breakdown of the data set as a whole for all states.

Another example is the disconnected phone report. The Manipula and ManiPlus setups that run to complete this process create a listing of cases where the phone number of the respondent has been disconnected. This listing only contains cases from the Data Center's state. Other files are created, one for each User State that has disconnected phone numbers in the data set, as well as a file of parameters for the e-mail package used by NASS. The Visual Basic menu script then sends an e-mail message with the attached file of that particular state's disconnected phones. The current NASS thinking is that if a Data Center encounters a disconnected phone number, the case becomes the responsibility of the User State to which it belongs.

NASS has been using this Data Center / User State relationship since September 1999 with much success. Some tweaking of the system was done during the early stages of implementation, as expected. However, the system matured rather quickly, and has been quite successful.