Moving from 43 Generated Unique Instruments to One National Instrument

Roger Schou, National Agricultural Statistics Service, USA

1. Background

At the National Agricultural Statistics Service (NASS), we have been generating instruments for our Crops/Stocks Survey for over ten years. It is the only national survey that NASS conducts, where the instruments are unique for each state. The Crops/Stocks Survey instrument collects crop acreage and production data, stocks data, and general livestock data of farmers across the country. It is conducted quarterly. It has been a very successful concept.

The generation of instruments results in 43 unique instruments generated in our field offices. It allows each state to ask crops pertinent to their state in their preferred order. For example, some states prefer to ask soybeans before sorghum, while others prefer sorghum first. Also, within a given crop, some states prefer to ask the irrigated questions before the non-irrigated questions, while others prefer non-irrigated first.

The ability to generate their own instruments from pre-tested headquarters-written code gives the states the luxury of having it their way. Since each state field office collects their own state's data, this is the preferred method, at least from the field office perspective. There are some maintenance issues from the headquarters' point of view.

1.1 Generator Maintenance Issues

The generation process is driven by a specifications database, which is a Blaise data set. It contains one record for each quarter for each state. It keeps track of things like which crops are asked, in what order they are asked, any special file extensions, whether irrigated questions precede non-irrigated questions, and much more information needed to assemble and generate an instrument from a library of pre-tested code.

There are many files that are specific to a given quarter. The developer must remember to change the extensions of a number of files to the "live" version for that quarter, and distribute them to the states.

Many of the similar crops share a rules section. Making a change for one crop may adversely affect many other crops if the change is not implemented correctly. Sometimes there is extra routing that pertains to one or two crops. This additional routing can be confusing when reviewing the rules for a crop that does not use that particular routing. Also, the error checks are often more complex because all the crops sharing the rules file must be taken into account.

Since there are 43 unique versions running in the field offices, testing the instruments becomes a tedious process. Testing may be tedious with one national instrument as well, but getting to the point of testing is greatly reduced. Each version must be generated in headquarters in order to test the instruments and to support the questions posed by the field office staff. The process of generating all 43 of the unique instruments has been automated, but it still takes 1.5 hours to complete when all of the code is error free. In order to duplicate a reported problem during the

survey, the instrument must be generated (or copied from an earlier generation process) and sometimes a test sample must be initialized.

2. Changes in NASS

As survey management and technology continue to evolve, changes are needed to adapt to the evolution. Two changes in NASS that have a direct impact on the way we conduct the Crops/Stocks Survey are the introduction of Data Collection Centers, and the Question Repository. Both of these changes have been driving forces in the decision to move from the 43 generated unique instruments to one national instrument for the Crops/Stocks Survey.

2.1 Introduction of Data Collection Centers

A few of the NASS field offices have begun to function as Data Collection Centers, because hiring and maintaining a steady crew of telephone interviewers is much easier in states that have less competition in the work force. In other words, field offices in larger metropolitan areas have more difficulty hiring interviewers than offices in smaller cities.

As the Data Collection Centers increase the number of states for whom they are collecting data, the survey management of the Crops/Stocks Survey is becoming overwhelming. Since the Crops/Stocks Survey has a unique instrument for each state, the Data Collection Centers must setup each state's survey as a separate entity. They have quickly maxed out with the number of states that they can manage for a Crops/Stocks Survey. Since the scheduler looks at one particular survey, and each state is setup as a separate survey, call management is quite convoluted.

2.2 Question Repository

NASS is currently building a Question Repository. This will be used to create both paper and WEB versions of questionnaires for NASS surveys. One of the requirements of such a system is standardization. After reviewing the generation of unique Blaise instruments and the number of years spent developing this generator, it was determined that the Question Repository would create unique paper and WEB instruments for each state, but the crop order would need to be chosen from one national standard order. In addition, within a given crop, we would need to ask the cropping practices (irrigated and non-irrigated questions) in the same order.

By not allowing the states to uniquely customize their paper forms (as far as question order is concerned) an estimated two to four years of development time of the Question Repository will be saved. This estimate is based on the amount of time invested developing and testing the generator in Blaise.

3. A Decision Was Made

An executive decision was made to create one national Crops/Stocks Survey instrument that would only ask crops, stocks, and livestock data that are relevant to each state, with the crops and cropping practices asked in one national standard order. It was important that the states were made aware of what they would have to give up as well as what they would gain.

There were open lines of communication between headquarters and the states, and we found that the states were not as attached to maintaining their crop orders as they once were. From survey administrators, to developers, to sample designers, to analysts, to upper management, the idea was passed around and discussed, which led to total buy-in and finally the executive decision. The projected date for this standard order to be in production in all states (in Blaise and on paper) is March 2005. There will be smaller sample tests (exclusively on paper forms in September 2004 and December 2004) to measure any impact of changing the existing order of crops.

4. From Building Specifications to Driving Specifications

The specifications database is still a big part of the Crops/Stocks Survey, but its role has changed from containing "data needed in order to build an instrument" to "data needed to run the instrument." Instead of Manipula reading the specifications database in order to generate code, the data model reads it as an external file to determine which crops need to be asked in a particular state and whether cropping practices data need to be asked.

The specifications database still contains one record for each quarter for each state. Many of the fields that are required for building an instrument no longer have any value, because the instrument code already takes that into account. Similarly, new fields have been added to drive the instrument's routing of questions.

5. One National Instrument

In order to minimize the maintenance impact on a small headquarters development staff, we decided to create one national instrument that would collect data for any given state on any given quarter. The cost of the original development of this complex instrument should quickly pay off with the lower maintenance projected for the future of the instrument.

To understand the complexity of the one instrument concept, one needs to first understand the crop data being collected. There are over 50 program crops collected over the year. Some of these crops are further broken down into varieties. Not all crops are collected in all states, because all states don't grow all crops. On certain quarters, some states collect crop data by cropping practice, while other states do not. In some quarters only planted and/or harvested acreages are collected. In other quarters planted and/or harvested acreages and production numbers (either total production or yield) are collected. When a crop is broken down into varieties, not all states collect the same combinations of the varieties, and some states don't even break them down at all. Some states collect crop data for the varieties by cropping practice.

Between the generator system and the national instrument concept, there are differences in the individual crop files. The following two sections highlight these differences.

5.1 Individual Crop Files in the Generator System

In the generator system, every crop has one file for every structure it may be involved in over the year. Some examples of crop structures follow: 1) planted acreage only, 2) planted and harvested acreage, 3) planted and harvested acreage, production and yield, 4) planted and harvested acreage, production and yield for grain and silage, and so on. The structures for a particular crop change as the quarters change.

All of these crop files are created by the generator, and consist mainly of an "assignment block" and the fields section for the crop and its structure. The assignment block computes some locals and auxfields for things such as error limits, crop names, production and yield units, and conversion factors. The assignment block then is used as a type for one of the fields in the crop block. Sometime the unit of production differs from state to state on certain crops. In this case, the assignment block is different and the crop file is saved with a different extension.

The rules for the crop block are more generic and are shared between crops that have the same structures. Many times there is extra code in the rules file that pertains to one or two crops that are sharing it. This can cause some confusion when looking at the block from another crop's perspective. Also, when making a change to the rules for a particular crop, it may affect a crop that the developer is not aware of at the time.

5.2 Individual Crop Files in the National Instrument

The new national instrument has a maximum of two files for every crop. The first file is for the crop itself and the second file is for any cropping practices in which the crop may be involved. The individual crop block accounts for all possible structures of the crop. As in the generator system, the crop file consists of the assignment block and the fields section for the crop.

There is one corresponding rules file for every crop file, making a maximum of two rules files per crop. Since the rules are specific for each crop, there is no extraneous information or routing for other crops. A crop may now be updated without the risk of the change affecting other crops. However, if the update is required to a specific structure, then the change needs to be made in multiple crop files. This is bit more labor intensive than the generated process, but in the long run, the single national instrument is easier to maintain.

It is very important to the editors in the field offices to be able to use the jumping capabilities of the field tags in Blaise. Therefore, even though many of the crops' field sections are the same, none of them are shared. The majority of the fields have a field tag that is equivalent to an item code which corresponds to the paper form. This allows the editor to jump to any portion of the questionnaire without paging through numerous screens to get to their data.

5.3 Prepare Directives

Obviously, the number of extra fields needs to be kept to a minimum. Instead of defining all possible fields for the entire year and only routing the needed ones in any given quarter, prepare directives are used to only include the fields needed for the quarter. There are still a number of extra fields that some states do not need, but they must be present in case the Data Collection Center is calling for multiple states. However, the total number of extra fields is much less than it would be without the use of the prepare directives. This instrument works for any state in any quarter, as long as it is newly prepared each quarter.

The use of the prepare directives has greatly reduced the size of the instrument. For example, in quarters where only planted acreage is collected, the fields for harvested acreage and production do not exist in the instrument. This is not a big issue for CATI, but is a factor in Interactive Edit, as there will be fewer "empty" screens through which to page.

Prepare directives also eliminate the need for the developer to change extensions on files that are different from quarter to quarter. The prepare directives allow the appropriate file to be included in the appropriate quarter. Many of the tedious steps from the generated process have been eliminated.

Great care was taken when assembling the code to insure that 1) it would prepare for each quarter, and 2) the correct crops and cropping practices were routed for the appropriate states. It is very easy to have the correct code in place, and yet have it ignored because of the prepare directives. However, once the code is in place and tested, maintenance is projected to be much lower in the national instrument than it is in the generated instrument.

The developers have adopted an indentation standard with respect to the prepare directives, in order to more easily detect where they start and stop. This becomes even more helpful when the prepare directives are nested.

6. Conclusion

Converting from the generated instrument schema to the one national instrument schema is a large undertaking, but the end result will be a more easily maintained instrument that meets the requirements of NASS. The generated instrument scenario worked very well for over ten years as a proven method for creating customized instruments for the states. It no longer meets the needs of the environment to which NASS is moving.

The one national instrument allows the Data Collection Centers to collect data for the Crops/Stocks Survey for multiple states, keeping the data in the same dataset. Each state's survey will no longer need to be treated as a separate entity in the Data Collection Centers. The call scheduler will handle the call backs for all of the states for which they are responsible. Data Collection Centers may add additional states in a given quarter much more easily than with the generated instrument. Maintenance and survey preparation time by the limited headquarters staff will be greatly reduced.

The Question Repository has motivated NASS to move to a standard crop and cropping practice order for the Crops/Stocks Survey. This will require fewer resources in developing the Question Repository. The standardized question order opened the door to move to one national instrument for the Crops/Stocks Survey in Blaise, since the original driving force behind the generator was the ability to customize the questionnaire in the states.

With the idea of only maintaining one instrument for all states in all quarters, we wanted to minimize the size of the data model as much as possible. This is preferable by our data editors in the field offices, because they don't have so many "empty" pages to page through to get to their data.
The prepare directives in Blaise provide for a much more streamlined instrument, while continuing to use the same instrument for all quarters.