THE NASS QUESTION REPOSITORY SYSTEM

Daniel G. Beckler, USDA National Agricultural Statistics Service
3251 Old Lee Highway, Room 305, Fairfax, VA 22030, dan_beckler@nass.usda.gov

1. INTRODUCTION

The United States Department of Agriculture's National Agricultural Statistics Service (NASS) conducts hundreds of surveys annually on the nation's farmers and agribusinesses. Multiple questionnaire versions are needed for most surveys to address differences in agriculture between states. Since the majority of NASS' surveys are multi-modal, questionnaires also need to be developed for multiple data collection modes: mail, face-to-face, telephone/CATI, and most recently the World Wide Web. To efficiently create the numerous questionnaires needed for all of the survey, state and mode combinations, NASS developed a client-server based Question Repository System (QRS) (House, 2000). The QRS includes a user interface to build properly formatted questions for the various modes which are stored in a central database. The stored questions may be retrieved and used to build questionnaires, which can be saved, printed or ported to a Web server. This paper describes the capabilities and some technical details of the QRS.

2. SYSTEM REQUIREMENTS

The basic requirement of the QRS was to efficiently create publication quality paper and Web questionnaires. However, in order to fit into NASS' environment, the QRS needed to accommodate additional requirements. These included:

- The paper and Web questionnaires needed to be created by staff without Hypertext Markup Language (HTML), or any other type of programming experience. This requirement was necessary since, due to staff resource limitations, the primary QRS users would be existing questionnaire designers who have extensive experience with word processors, but are not programmers.

- The QRS needed to store individual questionnaire components – questions, headers, footers, instructions, etc. – as individual entries in a central database repository. This requirement would allow reusing the same questionnaire components that are common among multiple state versions in a given survey, as well as those that are used in multiple surveys. This would provide greater standardization within and across surveys as well as increase efficiency in creating questionnaires.

- Users needed to be able to identify subsets of the individual questionnaire components in the central database to be used to generate paper and Web questionnaires simultaneously.

- The process to generate the paper and Web survey instruments needed to be as efficient as possible. This requirement was crucial since NASS conducts over 400 surveys annually, most of which have several questionnaire versions to accommodate uniqueness across states. The survey instruments for these surveys range in size from a single page with only a few questions, to over 30 pages with several hundred questions. Furthermore, due to resource limitations, all versions of both the paper and Web questionnaires needed to be created by NASS' existing questionnaire design staff.

- The system needed to be robust enough to accommodate virtually any type of questionnaire design, yet provide tools to ensure standardization across questionnaires. Although most surveys NASS conducts are relatively static and repeated at set intervals (weekly, monthly, quarterly, or annually), there is enough variation in survey programs to necessitate a flexible system. Also, NASS routinely implements new surveys for other organizations that often collect data not encountered previously.

- The Web survey instruments needed to comply with the Government Paperwork Elimination Act (GPEA) and with Section 508 of the Americans with Disabilities Act.

- The system had to use existing Agency metadata that would benefit the questionnaire development process, as well as contribute to NASS' metadata warehouse. This would benefit other existing and future systems.

- The QRS needed to be developed and implemented in roughly a three year period with a limited budget. A limited pilot QRS was developed during the first year and a production version containing essential features was developed during the second year. Third year development consisted of enhancing and refining the production system.

3. QRS DESCRIPTION

The Question Repository System that was developed accomplished the requirements listed above. The NASS QRS is a client-server application consisting of three main components: (1) a user interface client written in Visual Basic.NET (VB.NET), (2) Microsoft Word with an added application written in Visual Basic for Applications (VBA), and (3) a Sybase database.

The VB.NET client serves as the "control panel" for the QRS. The client has three main functions: (1) it allows users to search and preview the individual questionnaire components (questions, headers, footers, etc.) as well as completed survey questionnaires, (2) it allows users to build paper and Web survey questionnaires by specifying a subset (known as a *recipe*) of the individual questionnaire components as well as supplying necessary metadata (such as the survey's title, the states the survey is valid in, the date of the survey, etc.), and (3) the client instantiates Microsoft Word (with the added VBA application layer). Microsoft Word is used to produce all individual questionnaire components.

Perhaps the most unique component of the QRS is the Visual Basic for Applications application that was written for Microsoft Word. This VBA application adds considerable functionality to the off-the-shelf version of Microsoft Word.

Specifically, the VBA application provides a host of formatting tools that ensure standardization. Such things as font sizes, line spacing, margins, tab sets, table row/column sizes and answer box sizes are controlled by a series of custom tool bars added by the VBA application. In addition to ensuring a standard appearance for the final questionnaires, these features provide tremendous time savings for the questionnaire designers.

The VBA application also provides a formatting wizard for HTML input fields (i.e., text boxes, radio buttons, check boxes, and drop-down boxes). This feature allows a questionnaire designer with literally no HTML programming experience to produce virtually any type of Web questionnaire.

Additionally, the VBA application provides a "fill variable" input wizard that allows questionnaire designers to insert variable fields that will be replaced with current information when the questionnaire components are actually used in a survey. For example, if the current year is desired to appear in a question (e.g., how many acres of corn did you plant in 2004?) the questionnaire designer would not actually enter "2004," but rather insert a fill variable (e.g., how many acres of corn did you plant in CURRENT_YEAR). This field (CURRENT_YEAR) would be replaced with the appropriate year when the question is actually used in a survey instrument. This feature allows for greater reuse of questionnaire components, and consequently, more efficient questionnaire creation.

The same concept of "fill variables" is used for question numbering. The VBA MS Word application allows questionnaire designers to label questions with dynamic question numbering. When questionnaires are built, the QRS sequentially numbers all questions correctly. This feature allows questions to be able to be used in different locations in multiple questionnaires.

The VBA MS Word application also allows question designers to build skip patterns within questionnaires. For the designer this process involves identifying the screening question and the questions that the respondent will be routed to when certain conditions are met. This process is accomplished with a dialog box. This feature allows questions to be built with skip logic by QRS users with no programming background.

Finally, the VBA application provides a tie to NASS' data warehouse – specifically, the master variable names and associated metadata. This data warehouse contains a repository of all official master variable names for all of NASS' questionnaire items as well as all previously reported survey data. The master variable names are used by a variety of post-data collection editing and summarization programs. The MS Word VBA application allows questionnaire designers to embed (as bookmarks) master variable names in questionnaire answer cells that allow Web data to be captured to appropriate variables.

All questionnaire components as well as all completed questionnaires are stored in a Sybase database as binary large objects (BLOBs). The individual questionnaire components are stored in both their native Microsoft Word format files (.doc files) and in MS Word generated XML format files. The MS Word files of the individual components are used to assemble completed paper questionnaires, which also are stored as MS Word format files. The XML files of the individual components are used by a Web server to dynamically create Web questionnaires. In addition to the BLOBs, the database also stores many metadata fields related to the components and completed questionnaires.

Storing the individual questionnaire components in their native MS Word format, and using those Word files to assemble paper questionnaires provides the benefit of retaining all formatting (font, alignment, spaces, etc.) details of each component. This ensures the constructed paper questionnaires will be of publication quality.

3.1 OTHER QRS FEATURES

In addition to the many formatting related features, the QRS provides several features to aid in the development and management of questionnaires.

One of the most powerful QRS features is the notion of a *recipe*. A recipe is simply a subset of questionnaire components (i.e., a header, all of the questions, a footer, instructions, etc.) that build a questionnaire for a particular survey. These recipes are saved in the QRS database and may be retrieved at a future time to generate questionnaires for the same survey for another reference date.

Another useful feature is the ability to build questionnaire components for specific modes (currently only paper and Web). While being built, components can be identified to be used for the Web, for paper, or for both. If separate paper and Web versions exist for a specific component, these components may be linked together to maintain their "equivalence" relationship. After questionnaires are built using these components, the appropriate one will be used for each mode (i.e., the paper version will be used to generate the paper questionnaire and the Web version will be used to generate the Web questionnaire). Furthermore, the system allows individual portions of a component to be masked for a specific mode. This allows a single component to be built for both paper and the Web, but still allow flexibility for each mode. For example, a questionnaire instruction may request respondents to "Return the completed questionnaire in the enclosed postage paid envelope." Clearly, this line is not appropriate for the Web mode, consequently, this line could be flagged to appear on the paper questionnaire only and be masked on the Web.

The QRS also contains a feature to track the history of questionnaires. Although somewhat uncommon, changes to questions from one implementation of a survey to another do occur. The QRS provides users with a history of all changes to questions as well as an assurance that the most current version of questions are used.

Another QRS feature allows questionnaire designers to automatically e-mail paper versions of questionnaires (i.e., the MS Word files) to NASS field offices or other NASS staff. This feature is used to distribute draft versions of questionnaires for review, or to send finalized versions to NASS' field offices for data collection.

While being primarily built to produce survey questionnaires, the QRS may also be used to produce forms used internally at NASS, which may then be delivered to NASS employees via NASS' intranet. These forms may be for administrative data collection efforts (such as leave permission slips and supply orders) or for internal organizations climate surveys.

3.2 IMPORTANCE OF METADATA

NASS is becoming increasingly aware of the benefits of metadata in virtually all aspects of the survey process. Consequently, metadata are playing a key role in the development and operation of the Question Repository System. Metadata generated by the QRS and by other systems are used extensively by the QRS to ensure accuracy in the generated questionnaires, as well as to allow the QRS and other NASS systems to interact more efficiently. Notable examples of the uses of metadata by the QRS include:

- All survey components are identified as to which states they apply. This information is used to ensure that the appropriate questions are delivered to the appropriate states. For example, NASS conducts a series of Quarterly Crops/Stocks Surveys each year that collect acreage and production information for a series of crops. The specific list of crops the surveys target vary by state. By attaching state metadata to each question, QRS users are able to create individual state questionnaires that contain the correct list of crops by building only a single recipe for the survey.

- The master variables obtained from NASS' data warehouse contain metadata for such things as a variable description, the type of data the variable would store (text, numeric, dollars/cents, etc.) and a measure of precision. All of this information is used to render the Web input field correctly, as well as to provide HTML alt tags for Section 508 compliancy.

- The questions the QRS produces serve as metadata for other systems. Hence, a user retrieving previously reported survey data from NASS' data warehouse can not only retrieve the survey answers, but also the question text which was used to collect the data.

4. NASS EDR ARCHITECTURE

The Question Repository System, along with its associated database, is the heart of NASS' overall questionnaire production environment. However, it is only one part of NASS' complete Electronic Data Reporting (EDR) system. Figure 1 illustrates the entire NASS EDR system for generating and delivering Web questionnaires.

The EDR Web server runs Apache software and relies on PERL common gateway interface (CGI) scripts to query the necessary databases to dynamically construct and deliver Web questionnaires. The CGI scripts perform three

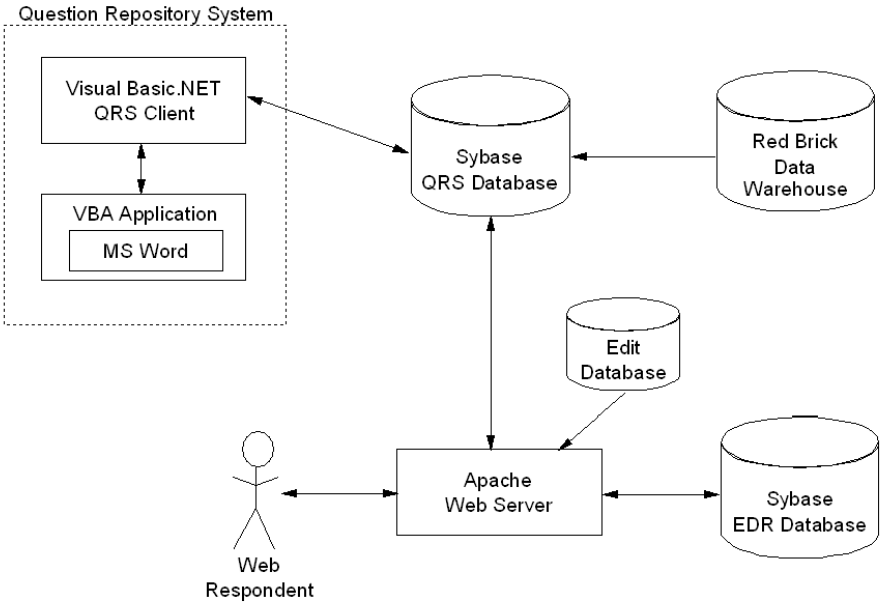

Figure 1: NASS Electronic Data Reporting System (simplified)

key functions: (1) confirm that a respondent-provided identification number exists in the EDR database, (2) identify what surveys a particular respondent is to receive, and retrieve the appropriate survey questions from the QRS database to dynamically create the Web questionnaire, and (3) store respondent-provided survey data to the EDR database under the appropriate master variable names. In addition, the CGI scripts perform a series of formatting operations that serve to adjust the rendering of the Web questionnaires to conform to NASS business rules and Section 508 compliancy guidelines.

The EDR database stores the Web samples (name, address and other identification information) and the data the Web respondents provide. NASS' proprietary Survey Management System (SMS) provides the user interface for populating the Web samples and for retrieving Web provided survey data. The Web-provided survey data are then combined with data obtained from other modes and processed through post data collection editing and summary systems.

The EDR system relies on an edit repository for the edit checks that act on Web respondent provided data. These edits are coded in PERL and are typically data consistency and range checks, but may be virtually any logical operations to the data. The Sybase database that contains the edit repository stores the actual PERL syntax for the edit along with any error messages the respondent would receive. NASS personnel knowledgeable with PERL create the edits and error messages through a Web browser interface. The edit logic uses master variables, and once created, the system automatically applies the appropriate edits to all Web questionnaires that contain all of the necessary questions. For example, if an edit's logic contains the variables A and B (e.g., A < B), then this edit is applied every time the questions that contain the variables A and B appear in a Web questionnaire. Hence, edits are built at the variable level instead of for specific questionnaires. This makes programming edits more efficient since each specific edit need only be built once. In addition, this approach ensures more consist editing across surveys.

Although the edit repository is currently only used for Web questionnaires, NASS intends to expand its application to include CATI and post-data collection edits as well.

It should be noted that the NASS EDR system also contains various firewalls and other security measures. These topics were intentionally omitted in this paper because they are not only beyond the scope of the paper, but also to avoid disclosing any information that may jeopardize their effectiveness.

5. QRS SCREEN CAPTURES

Figures 2 and 3 contain screen captures of two Question Repository System interfaces.

Figure 2 illustrates the most basic view of the user interface client. As can be seen, the interface is laid out similar to Microsoft's Windows Explorer in that the screen contains a left pane and a right pane. The left pane contains a navigation tree that is used to scroll through existing

questionnaire components (as illustrated), or surveys (i.e., questionnaires). Buttons above the left pane switch between displaying the components or surveys. Other buttons above the pane allow users to toggle different sorts of the components or surveys.

The right pane contains a detailed view of what is selected in the left pane. In Figure 2 the right pane contains a pre-view of a specific question (for illustration purposes, the list of questions that would normally be displayed in the left pane was collapsed). The right pane also contains various metadata for the particular selection – in this case, a list of states where the question is valid, the user who created the component, when the component was created, and indicators that identify in which data collection modes the question will be used.

Figure 3 illustrates the Microsoft Word interface with the VBA enhancements. The added features appear as several toolbars that were added to standard MS Word. These toolbars are identified with a bracket and arrow in Figure 3. As with other Word toolbars, users may select which added features they wish to view at any given time. Note: NASS is currently restructuring the toolbars to display them in a more logical, user-friendly way.

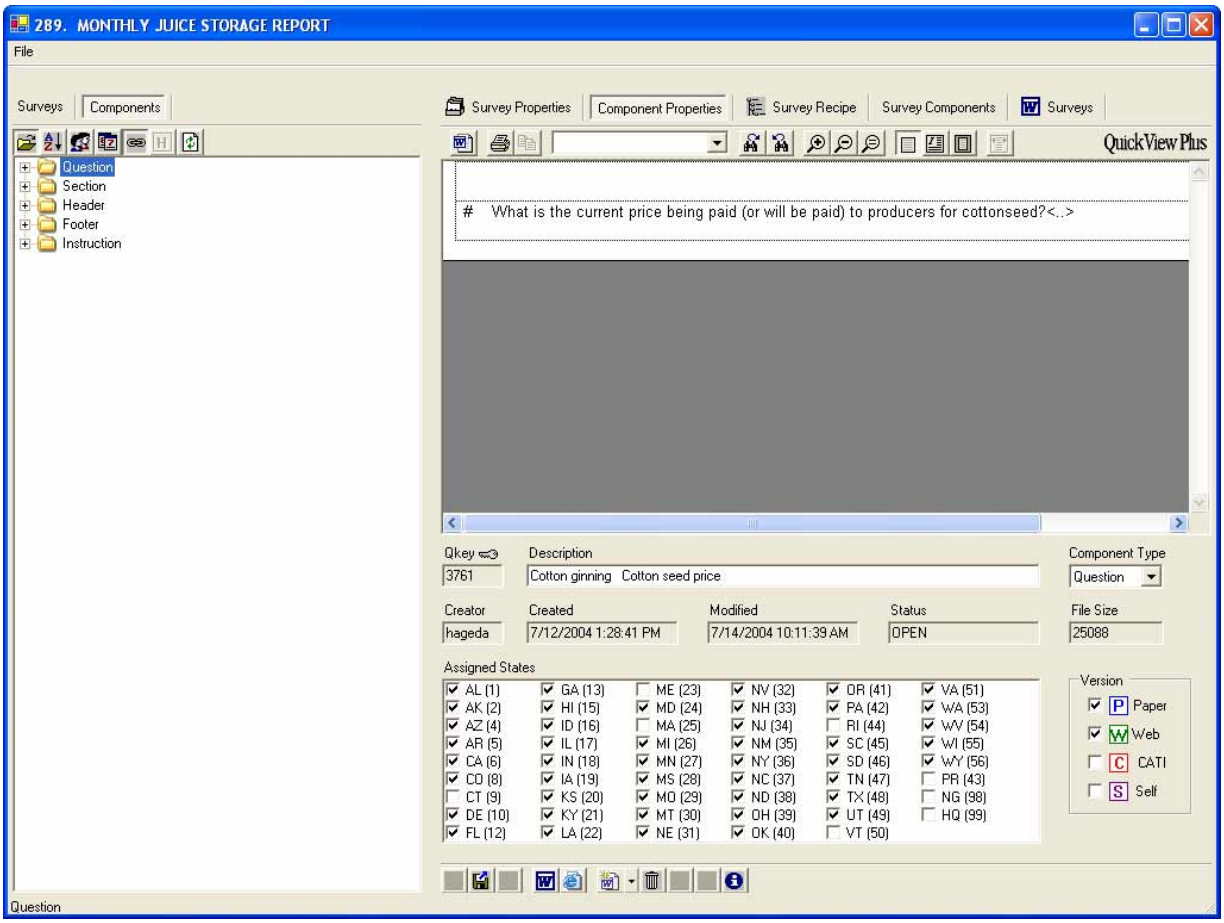

Figure 2

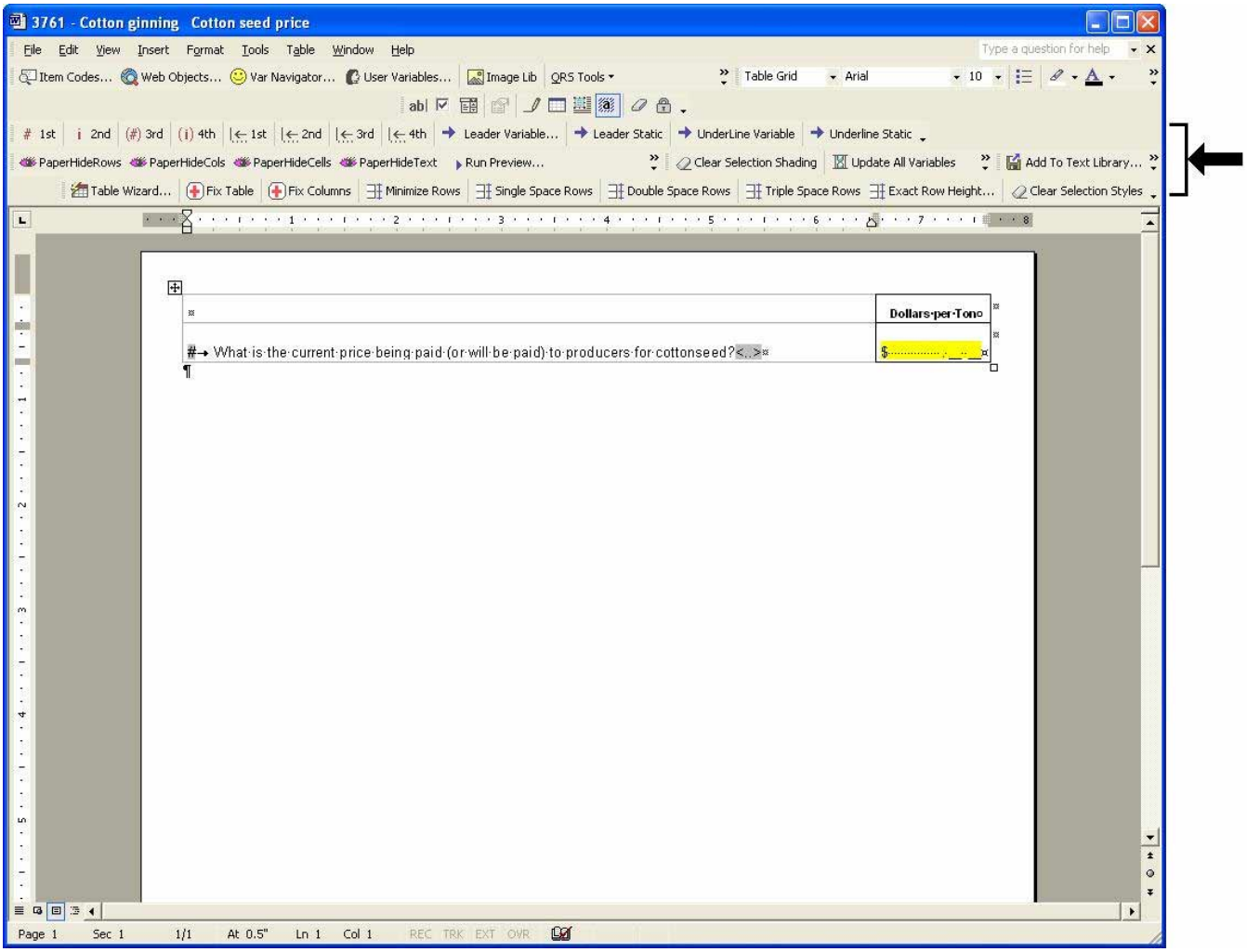

Figure 3

6. CONCLUSION

The NASS Question Repository System provides an efficient, flexible platform to generate both paper and Web survey questionnaires. In addition to paper and Web questionnaires, the system is adaptable for producing questionnaires for other delivery modes, such as computer assisted telephone interview (CATI) and portable electronic devices. NASS is currently exploring how the QRS may best benefit the organization for these other modes.

Although NASS' QRS is a great tool for producing questionnaire instruments for multiple modes, it does not reduce the need for good design. Thorough pretesting and usability testing are still critical steps in designing new questionnaires and in reformatting existing ones. Indeed, this is especially true with the increasing number of multi-mode surveys that may impart unknown modal differences.

7. REFERENCE

House, C. (2000), "Integrating Paper and Web Instrument Development to Enhance Efficiency and Standardization," paper presented at the 2002 International Conference on Improving Surveys, University of Copenhagen, Denmark.