

U.S. DEPARTMENT OF COMMERCE
Office of Inspector General



**PUBLIC
RELEASE**

BUREAU OF THE CENSUS

*Improvements Needed in Multiple Response
Resolution to Ensure Accurate, Timely
Processing for the 2000 Decennial Census*

Inspection Report No. OSE-10711 / September 1999

Office of Systems Evaluation



Table of Contents

| | |
|---|----|
| EXECUTIVE SUMMARY | i |
| INTRODUCTION | 1 |
| PURPOSE AND SCOPE | 2 |
| BACKGROUND | 4 |
| OBSERVATIONS | 8 |
| I. Dress Rehearsal Indicates That the Bureau Can Improve MRR Accuracy | 8 |
| A. Dress Rehearsal Evaluation Raises Accuracy Issues | 8 |
| B. MRR Must Address Questionnaire Data Quality Problems | 11 |
| C. MRR Rules and Software Warrant Review | 13 |
| II. A More Structured Development Approach Would Improve MRR Software | 15 |
| A. Bureau Software Engineering Standards Are Available but Not Widely Used .. | 16 |
| B. The Bureau Should Define MRR Requirements More Explicitly | 16 |
| C. MRR Software Testing Needs to Be More Complete | 19 |
| D. Beginning-to-End Testing Is Needed to Ensure Consistent Results | 20 |
| III. Improvements Are Needed to Make MRR Processing Timely for the 2000 Census ... | 21 |
| CONCLUSION | 23 |
| RECOMMENDATIONS | 24 |

EXECUTIVE SUMMARY

To make it easier for people to be counted in the 2000 Decennial Census, the Census Bureau will provide several ways for households to respond. In addition to collecting data through mailed questionnaires and enumerator operations, the bureau will enable the public to initiate responses on forms available in public places, by telephone, or the Internet. While multiple methods of responding increase the opportunities for people to be counted, they also increase the likelihood of receiving multiple responses for some housing units, with some responses including the same people and others reporting people for the first time. The bureau has devised a set of rules to count individuals at each address and has automated this process, called multiple response resolution (MRR).

We conducted this evaluation to determine whether MRR is likely to resolve multiple responses correctly—that is, when multiple responses are received, MRR will include people at an address who resided there on Census Day (April 1, 2000) and omit people who did not reside there on Census Day or whom MRR already counted via another response. We also sought to determine whether MRR could accomplish its work in the allotted time during the decennial. Beginning in February 1998, we observed the bureau's preparations for using MRR for the dress rehearsal and reviewed results from its processing of dress rehearsal response data.

According to the Census Bureau, advances in computer technology provide the flexibility to offer multiple response options without incurring undue risk to the accuracy of the resulting census data. The bureau's *Census 2000 Dress Rehearsal Report Card* asserts error rates as low as 0.3 percent for MRR. However, this statistic reflects the impact of MRR errors on the total enumeration. It does not reflect the accuracy of MRR in resolving multiple responses, which in the bureau's dress rehearsal evaluations showed significantly higher error rates (see page 8).

We found several conditions that diminished the accuracy of MRR in the dress rehearsal. Issues associated with questionnaire design, the way in which the public completed the questionnaire, and automated data capture sometimes caused inaccurate data to be sent to MRR processing, which then could not produce accurate results. In addition, the MRR software sometimes produced questionable results in selecting the persons who make up a particular household and in determining whether duplicate sets of data represented the same person. MRR analysts are reviewing the dress rehearsal and evaluation data to determine what modifications are needed to the rules and associated software for resolving multiple responses (see page 11).

The Census Bureau cannot ensure that MRR is implemented correctly because it did not use a sufficiently structured approach in developing the MRR software. A more structured approach, using software engineering standards, would help ensure correct implementation of MRR. In particular, conformance to a software engineering standard for specifying requirements would produce a more accurate, complete requirements specification that would benefit users and developers alike during software development and testing. The bureau has established such a

standard, but it has not yet been widely adopted by decennial staff. In addition, software testing needs to be expanded to address more diverse response data and assure that MRR can reliably handle a wide array of potential input conditions. Finally, the bureau needs to undertake a coordinated effort to perform beginning-to-end testing of questionnaire data processing to ensure that the output from each step is accurate and can be input into the next step without error (see pages 15 and 20).

According to the Census Bureau, MRR processing for the decennial census will need to be completed within 30 days to ensure timely availability of data for follow-on processing operations. However, dress rehearsal results indicate that unless improvements are made, MRR processing for the decennial census will require approximately 87 days. The bureau needs to modify MRR software to reduce excessive processing time and obtain additional headquarters computing resources (see page 21).

As a result of these issues, important decisions concerning MRR requirements and design must be made and significant development and testing work accomplished. The bureau has recognized many of the problems that we have discussed and is analyzing the dress rehearsal data and working on improvements to MRR. The bureau has also made substantial improvements to the data capture system since the dress rehearsal. However, the bureau has not fully defined the activities needed to refine MRR and complete its development. To ensure that MRR is correctly implemented and well tested for the 2000 Decennial Census, the bureau should define these activities and develop and implement a plan for their completion (see page 23).

Our complete recommendations begin on page 24.



In its response to our draft report, the bureau stated that, with one exception, it concurs with, or had already acted upon, our recommendations. The exception regards specifying the software requirements necessary to address data quality issues, such as how to handle incomplete, inconsistent, or erroneous data (Recommendation No. 4d). The bureau requested clarification of this recommendation, which we provide on pages 18-19. Although the bureau concurred with our recommendation to perform beginning-to-end testing of the questionnaire data processing stream (Recommendation No. 5f), we believe that the planned testing, while valuable and needed, will not fully address the concerns we identify in this report. Our comments on pages 20-21 describe the type of testing that would be responsive to our recommendation. The bureau also provided comments on several aspects of our observations. Based on these comments, we amplified our discussion of questionnaire design and modified the title of Observation II. B. to avoid any confusion between establishing standards for MRR accuracy and writing software requirements specifications according to software engineering standards. The bureau's response is included in its entirety as the appendix to this report.

INTRODUCTION

The Bureau of the Census conducted the 1998 Dress Rehearsal to test methods and systems that are planned for the 2000 Decennial Census. Because of the importance of the dress rehearsal to the success of the decennial census, we have been evaluating critical information technology components used to conduct the dress rehearsal, including data capture, the personnel and payroll processing systems, and headquarters processing. This report addresses multiple response resolution (MRR), a component of headquarters processing. Our evaluation of data capture was presented in a report entitled, *Data Capture System 2000 Requirements and Testing Issues Caused Dress Rehearsal Problems*.¹ We found that the data capture system experienced serious problems in dress rehearsal resulting from insufficient testing and from difficulties in controlling and communicating requirements.

The availability of multiple methods of responding to the census has significantly increased the potential for persons and households to submit multiple responses. In addition to bureau-initiated enumeration, the bureau will enable the public to initiate responses on forms available in public places or via telephone or the Internet. While multiple methods of responding increase the opportunities to ensure that people have been counted, they also require the capability to resolve situations in which more than one response for a housing unit² is received.³ In the past, the bureau has tested some operations requiring resolution of multiple responses and has developed automated and clerical methods to process them. New operations such as the Be Counted program⁴ create the prospect of receiving forms from unknown residences and obtaining fragmented household responses—adding new complexity to data processing development tasks.

The Census Bureau developed MRR software to consolidate questionnaire data submitted for a residence on multiple forms. The bureau used MRR software to process 1998 Dress Rehearsal response data. MRR is one step in the multi-step process within headquarters computer processing that culminates in the apportionment counts and other products of tabulating census

¹*Data Capture System 2000 Requirements and Testing Issues Caused Dress Rehearsal Problems*, OSE-10846, January 1999.

²A housing unit is a house, an apartment, a mobile home or trailer, a group of rooms, or a single room intended for occupation as separate living quarters to which there is direct access from outside the building or through a common hall.

³Resolving more than one response for a housing unit means choosing the appropriate data from multiple census questionnaires identified with that housing unit to represent that housing unit in the census.

⁴The Be Counted program targets areas that are traditionally undercounted by placing census questionnaires at accessible sites so people who believe they were not counted can pick them up, complete them, and mail them to the bureau.

data. MRR processes all housing unit questionnaire data by comparing data on all forms received for an address, identifying duplicate persons, and choosing members of each household. Since this processing affects the counts, it is vital that MRR software operate correctly and reliably.

PURPOSE AND SCOPE

In November 1997 we reported that the bureau was developing software for decennial headquarters processing without a well-defined software development process based on software engineering principles.⁵ To follow up on that work, we conducted this evaluation to study one system in depth, MRR. Our inspection objectives were to: (1) evaluate MRR requirements specifications to determine if they are appropriate, sufficiently defined, and clearly communicated, (2) determine if the data and software design will result in efficient, reliable processing, (3) determine if a comprehensive set of test cases that fully tests the requirements and a definitive method that correctly evaluates the outcome have been articulated, and (4) assess plans to capture lessons learned from the 1998 Dress Rehearsal and apply them to the 2000 Decennial Census.

During our evaluation, we reviewed requirements specifications, design documentation, software code, test plans, test data, test results, and Census Bureau evaluation reports pertaining to the dress rehearsal. We also reviewed relevant documentation and operational assessments from the 1995 and 1996 census tests, and available dress rehearsal data processing results. In addition, we met with representatives from the Census Bureau's Decennial Systems and Contracts Management Office, Decennial Statistical Studies Division, Statistical Research Division, Systems Support Division, and the MRR team.

With respect to our objectives to evaluate the requirements specification and testing, we analyzed two key documents⁶ regarding the two principal MRR elements, Within-Block Search (WBS) and Primary Selection Algorithm (PSA). The remainder of this report refers to these documents as the specification and the test plan, respectively:

- *Revised Specifications for the Within-Block Search and Primary Selection Algorithm in the 1998 Dress Rehearsal*, Bureau of the Census, May 27, 1998.⁷

⁵*Headquarters Information Processing Systems for the 2000 Decennial Census Require Technical and Management Plans and Procedures*, OSE-10034, November 1997.

⁶These documents are *Census Confidential*, which means that they are available to personnel involved in developing, testing, and required evaluations.

⁷An algorithm is a set of steps for solving a particular problem.

- *Census 2000 Dress Rehearsal Test and Acceptance Plan for the Within-Block Search (WBS) and Primary Selection Algorithm (PSA)*, Bureau of the Census, September 11, 1998.

To assist in our evaluation of the bureau's software testing, we also reviewed input and output data made available to us from MRR testing efforts. We used the bureau's documentation on applicable data files to understand this data.

To determine whether the data and software design would result in efficient, reliable processing, we reviewed design documentation, code, the structure of the decennial response file, and logs from the September 2, 1998, production processing of dress rehearsal data.

To assess the ability of the software to determine matches of person data accurately, we reviewed output of the Service-Based Enumeration (SBE)⁸ component of dress rehearsal processing, which employs the same matching software as MRR.

To assess how the bureau plans to capture lessons learned from the 1998 Dress Rehearsal and apply them to the 2000 Decennial Census, we reviewed the MRR evaluations:⁹

- Evaluation Memorandum F1c and F2b, *The Within-Block Search and Primary Selection Algorithm Operational Evaluation*, January 31, 1999,
- Census 2000 Dress Rehearsal Evaluation Results Memorandum Series #F2a, *Within-Block Search Expansion Evaluation Draft*, March 11, 1999, and
- Census 2000 Dress Rehearsal Evaluation Results Memorandum Series #F1b, *Results of the Evaluation of the Primary Selection Algorithm*, [undated].

We also evaluated the bureau's efforts to modify the implementation of these elements resulting from lessons learned from the dress rehearsal.

The purpose of this report is to inform the bureau of difficulties that we observed in implementing MRR. The team working to develop MRR has done a commendable job on a complex problem under difficult circumstances. We urge bureau management to support the team in refining and thoroughly testing MRR to ensure high quality results in the decennial census.

⁸The SBE operation is designed to enumerate people at facilities where they might receive services, such as shelters, soup kitchens, health-care facilities and other selected locations. This operation targets the types of services that primarily serve people who have no usual residence.

⁹The bureau restricts distribution of these dress rehearsal evaluation reports regarding MRR.

This inspection has been conducted in accordance with the Inspector General Act of 1978, as amended, and the *Quality Standards for Inspections*, March 1993, issued by the President's Council on Integrity and Efficiency.

BACKGROUND

Because the 2000 Decennial Census will provide several ways for households to respond, it is highly probable that there will be multiple responses for some housing units. Therefore, the complete set of census response data for housing units must be viewed as potentially including (a) multiple responses for a housing unit having person data in common, (b) multiple responses for a housing unit not having person data in common, and (c) person data in common among responses of different housing units. The Census Bureau has implemented MRR computer processing to select the residents of a housing unit that has submitted multiple responses and to prevent data resulting from multiple responses for the same person from being erroneously included in the census. Each housing unit has a unique identifier called a census housing unit ID, which processing prior to MRR uses to associate responses with an address. Then, MRR uses two elements to search for person matches among multiple returns. WBS resolves duplicate person data among returns at the same or different addresses within a defined area when those persons have been reported on Be Counted Form Equivalents (BCFEs),¹⁰ while PSA resolves multiple responses within the census housing unit ID.

Earlier Experiences with Multiple Responses for a Census Housing Unit ID

In the 1990 Decennial Census, multiple responses occurred for a census housing unit ID because 1990 had overlap between mail and non-response follow-up (NRFU) enumerations. The bureau consequently received both mail and enumerator returns for some households. Also, data capture allowed corrections to the paper form which was then recaptured. This recycling of the paper form through data capture could occur repeatedly. To select the "best" questionnaire record for an ID, the bureau designed and implemented the PSA program.¹¹ This PSA version selected the data of one primary questionnaire (with possible supplemental questionnaires for large households) with all persons intact to represent the census housing unit ID.

Additional opportunity for multiple responses arose in the 1995 Census Test, which was conducted at three sites. In addition to mail returns, the public could respond by using widely

¹⁰BCFEs include the paper "Be Counted" questionnaire forms, Telephone Questionnaire Assistance (TQA) interviews initiated by the respondent, paper mail return questionnaires sent by TQA upon request of the respondent, and Individual Census Questionnaires (ICQ), Individual Census Reports (ICR), and Military Census Reports (MCR) if the ICQ, ICR, or MCR have been assigned a census housing unit ID.

¹¹*Description of the Decennial Census Algorithm for the Selection of the Primary and Supplemental Records from the 1990 FOSDIC Data Capture Files*, Susan P. Love, November 27, 1990.

distributed “Be Counted” forms and by telephone. Data capture keying methods input persons’ names as well as population data (e.g., sex, age, and race). The bureau designed a more complex PSA to resolve multiple responses for each ID by selecting unique persons and marking duplicate person entries. An automated matching algorithm compared person data on pairs of responses and scored resultant pairs of persons based on the weights of matching data it found. Multiple responses requiring computer matching occurred for 2.64 percent of the total 178,680 households, or some 4,717 households. Under some conditions, clerks reviewed the multiple responses for an ID to minimize matching errors. Clerical review resolved 17 percent of cases submitted to computer matching.¹² A slightly modified 1995 version of PSA also processed data collected during the 1996 Community Census. However, the bureau eliminated the clerical review step because of insufficient time to conduct such a review.

The MRR Team

A team of analysts has defined the operational concepts and system requirements for MRR for the 1998 Dress Rehearsal and 2000 Decennial Census. Since the dress rehearsal and decennial census employ operations for obtaining responses that are new or expanded from the mid-decade tests, the likelihood of multiple responses for households and persons at census housing unit IDs has increased. The MRR team is responsible for producing the specifications that define the software requirements. Permanent members of the MRR team include personnel from the Decennial Statistical Studies Division and Decennial Systems and Contracts Management Office. Members from the former office are statisticians who provide the information on the functions that MRR must perform and review the input data and results for correctness. The latter office contains the processing systems group, which is responsible for developing the software for several processing components including MRR and running them operationally to produce official results. The software developers have also worked on defining the functional and software requirements. Since requirements specifications reflect decisions regarding statistical methods, Decennial Statistical Studies Division management signature is required to indicate that consensus has been reached before releasing the specifications to software developers.

The MRR team has had to contend with issues such as determining the geographic boundaries within which to search for duplicate responses for a person reported on a BCFE, deciding the criteria to use to define a household and to select the primary household, and specifying what data are required to qualify a person entry on a questionnaire as having information sufficient for matching. For the dress rehearsal, the team also had responsibility for developing the Invalid Return Detection requirements. However, because detection of invalid (or fabricated) returns is not a multiple response issue, it is not a responsibility of the MRR team for the Census 2000 program. The team also engineered and conducted system testing, which included developing

¹²1995 Census Test Results Memorandum No. 50, Theresa Leslie and Maureen Lynch, Bureau of Census, Appendix 4, pp. 4-5.

the test plan, preparing test decks to model various multiple response scenarios, conducting the testing, analyzing the results, and correcting and re-testing the software as necessary.

Major Elements of MRR

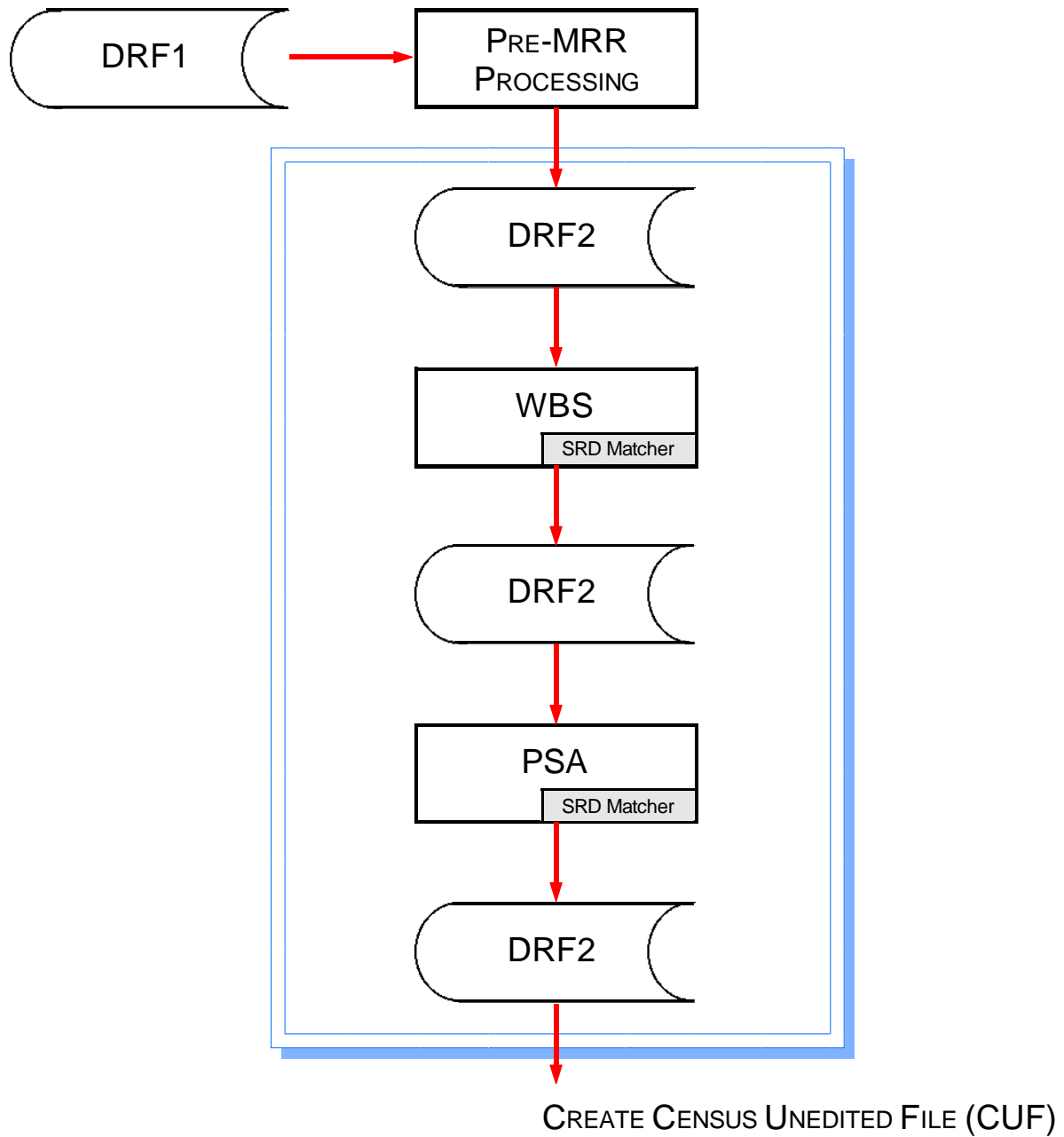
MRR consists of a sequence of processing steps that is executed in batch (non-interactive) mode. Figure 1 shows a simplified view of headquarters processing leading up to creation of the Census Unedited File (CUF),¹³ with the scope of MRR processing outlined in blue. The elements of MRR are described as follows:

- **Decennial Response File - Stage 2 (DRF2)**¹⁴ is both the input and output file for MRR and contains data fields whose values are compared to determine matching person records.
- **WBS** is a software program that performs a search and match operation for DRF2 person data records submitted on BCFEs when there is more than one response for a census housing unit ID. The search is conducted among census housing unit IDs within a defined area. If a match is found, WBS chooses one person entry and marks the other for deletion. WBS was developed and used operationally for the first time in the dress rehearsal.
- **PSA**, whose implementation is based on the software program developed for the 1995 Census Test, determines which DRF2 person records for a given census housing unit ID are to be counted. Since it is possible for the same person to be included on different census forms for the same census housing unit ID, PSA performs a search and match operation between the person records on pairs of response records for an ID. The PSA uses the presence or absence of person matches along with a set of rules to create groups of person records that are considered households by the PSA process. Other rules are used to determine exactly which persons are to be included at that ID.
- **Statistical Research Division (SRD) Matcher** is a software program invoked by both the WBS and PSA elements to match pairs of person records. The matcher assigns

¹³The output from PSA retains the name of DRF2 and is used to create the CUF, which merges census housing unit ID information from the Decennial Master Address File (DMAF) with data associated with each census housing unit ID from the DRF2. The CUF also contains standard codes for fields that may contain write-in information such as race and relationship for easy tabulation.

¹⁴The DRF is processed in two stages. Stage 1 creates DRF1 with the receipt of the first response records and is conducted on a continuing basis as response records are delivered to bureau headquarters from data capture operations. Selected DMAF variables are added to the DRF1 as it is sorted by block and census housing unit ID and before it is input to DRF2 creation. DRF1 is input to stage 2 which includes MRR and runs once to create a final DRF2.

Figure 1. Scope of MRR Processing



weights based on probabilities to each characteristic compared, and adds them to arrive at a score. WBS and PSA compare the score to a cut-off weight to determine whether to treat the pair as a match. WBS and PSA invoke the same version of the matcher program; however, WBS requires more restricted conditions for matching persons from different census housing unit IDs than PSA uses for matching persons in the same ID.

OBSERVATIONS

I. Dress Rehearsal Indicates That the Bureau Can Improve MRR Accuracy

The bureau has issued its dress rehearsal report card,¹⁵ which assesses the performance of six dress rehearsal operations. One of those operations is MRR, the focus of this evaluation. Despite the high accuracy rating attributed to MRR by the report card, we have found areas of concern to include (1) the report card rating of MRR versus evaluation results, (2) accuracy of the questionnaire data, and (3) unexpected results from MRR processing that warrant review.

A. Dress Rehearsal Evaluation Raises Accuracy Issues

The bureau's August 1997 *Report to Congress—The Plan for Census 2000* states, "Advances in computer technology in the areas of computer storage, retrieval, and matching, along with image capture and recognition, have now given the Census Bureau the flexibility to provide multiple response options without incurring undue risk to the accuracy of the resulting census data."¹⁶ This statement prompts the question of how accurately multiple responses actually are resolved. The Census 2000 Dress Rehearsal Evaluation Program has planned and implemented evaluations on some aspects of MRR and reported the results. The bureau has issued the *Census 2000 Dress Rehearsal Report Card*, which states the following accuracy levels for MRR processed data:

- For the type of error of including persons that a follow-up interview determined were not residents, the percentage was 0.3 percent for both the Sacramento and Columbia, South Carolina dress rehearsal sites.
- For the type of error of omitting persons that a follow-up interview determined were residents, the percentage was 0.4 percent for Sacramento and 0.3 percent for Columbia.

The bureau's dress rehearsal evaluation plan had set accuracy standards for MRR based on the percent of erroneous enumerations in the 1990 Census as measured by the Post Enumeration

¹⁵*Census 2000 Dress Rehearsal Report Card—Evaluation of the Standards for Success*, Bureau of the Census, February 1999.

¹⁶*Report to Congress—The Plan for Census 2000*, Bureau of the Census, August 1997, p.13.

Survey (PES). These standards were 4.6 percent for erroneously included persons (overcount) and 1.3 percent for erroneously omitted persons (undercount).¹⁷ Since the reported error rate for erroneously including persons for both sites was 0.3 percent, the report card claimed that this error rate met the standard. Similarly, reported error rates for erroneously excluding persons also met the standard.

The report card obtained these error rates from the draft report for the Evaluation F1b, *Evaluation of the Primary Selection Algorithm*.¹⁸ This evaluation report puts these error rates in perspective by stating that if the evaluation method had found that all persons selected by the PSA were selected in error, then the selected in error rate would be 3.1 percent in Sacramento and 2.3 percent in South Carolina. Similarly, if the evaluation method had found that all persons excluded by the PSA were excluded in error, then the excluded in error rate would be 1.0 percent in Sacramento and 0.8 percent in South Carolina.

So, even if the PSA element of MRR operated completely in error (excluded all residents and included all non-residents found on multiple responses), the values that would then be used in the report card would still meet the standards set in the evaluation study plan. The report card concluded that MRR did not create a problem relative to the overall accuracy of dress rehearsal enumerations.¹⁹

Bureau statisticians who define and develop MRR disagree with the use of PES percentages of erroneous inclusions and exclusions as the standard to which MRR results should be held. The percentage values are obtained by comparing response data from the census to data gathered during the PES. This comparison provides the numbers of matching and non-matching responses from the two operations. Matching and non-matching values support the dual-system estimation method²⁰ used to estimate the true population. Bureau statisticians working on MRR state that calculating percentages of matches and non-matches and using them as the standard for MRR is invalid because these discrepancies occur for many reasons in addition to incorrect decisions by MRR. Recognizing the controversy within its own staff, the bureau later amended the standard to read, “small relative to 4.6 percent” for the number of persons included in error and “small relative to 1.3 percent” for the number of persons excluded in error. The bureau should develop an appropriate standard of measuring the success of MRR processing.

¹⁷*Census 2000 Dress Rehearsal Evaluation Program Draft*, Bureau of the Census, March 10, 1998, revised May 15, 1998, Appendix E.

¹⁸*Census 2000 Dress Rehearsal Evaluation Results Memorandum Series #F1b, Results of the Primary Selection Algorithm Draft*, Bureau of the Census, [undated].

¹⁹*Report Card—Evaluation of the Standards for Success*, p. 8.

²⁰Tommy Wright, “Sampling and Census 2000, The Concepts,” *American Scientist*, May-June 1998, p. 252.

The report card identified the impact of MRR errors on the total enumeration, but it did not report the accuracy of MRR in resolving multiple responses. The manner in which the error rates in the report card were calculated resulted in substantial understatement of both the erroneous inclusions and erroneous exclusions produced by MRR. For example, in Sacramento, the number of persons belonging to households was about 300,000—which was the value used in the denominator of the erroneous inclusion rate. However, PSA obtained data from multiple responses for approximately 10,000 persons of that population. When calculating the error rate with the denominator equal to this value, the result is 9 percent. This is 30 times the error rate presented by the report card. Table 1 shows a summary of error rates for the two types of errors that the bureau evaluated for the dress rehearsal sites

Table 1
Dress Rehearsal Multiple Response Rates

| Site | Type of Error | PES Standard | Report Card | Total Possible MRR Error | Multiple Response Error Rate |
|------------|-------------------|--------------|-------------|--------------------------|------------------------------|
| Sacramento | Included in error | 4.6% | 0.3% | 3.1% | 9% |
| | Excluded in error | 1.3% | 0.4% | 1.0% | 11% |
| Columbia | Included in error | 4.6% | 0.3% | 2.3% | 14% |
| | Excluded in error | 1.3% | 0.3% | 0.8% | 11% |

The bureau’s dress rehearsal evaluation provides data that the MRR team is analyzing to adjust the algorithm for the decennial. This effort is important because the decennial will increase the number and variety of conditions under which multiple responses occur. Modifying MRR to process multiple response data more accurately will aid in reducing the overcount and the undercount that could potentially be caused by multiple responses.

- The third column, *PES Standard*, shows the PES standard as previously discussed.
- The fourth column, *Report Card*, shows the error rates presented in the report card.
- The fifth column, *Total Possible MRR Error*, shows the percentage value of the ratio of person data originating on multiple responses to the site’s total population. These values indicate the percentage of error MRR would contribute to the total count if MRR resolved all multiple responses erroneously.
- The sixth column, *Multiple Response Error Rate*, shows the error rates calculated by dividing the weighted number of persons in error by the total weighted number of persons whose data was obtained from multiple responses. We believe that the error rates shown

in the sixth column give a more valid measure of how well MRR software made the decision to include or exclude persons in the household. These equal the error rates calculated by the Census Bureau's evaluation team.

The report card's use of error rates which were so low as to suggest that MRR worked nearly perfectly is misleading. Inferences from these error rates would suggest that any efforts to improve MRR are unnecessary. However, the evaluation data indicates weak areas where improvement is needed. The MRR team is using the evaluation data to improve MRR processing, and we encourage the bureau to support the team in this effort. In reporting on MRR performance in the future, the bureau should identify both the accuracy of MRR in resolving multiple responses and the impact of MRR errors on the total enumeration.

B. MRR Must Address Questionnaire Data Quality Problems

Data quality affects the accuracy of any processing. Obtaining accurate input data starts with the respondent filling out the census form legibly and with complete and correct information properly placed on the form. Next, the data capture operation submits these forms to the automated system, DCS 2000, which converts write-in and check box data from the questionnaires into standard computer format. If DCS 2000 inaccurately converts this data to computer format, those errors propagate through all downstream processing including MRR. Missing, incomplete, or erroneous data in certain questionnaire fields affects the ability of MRR to match person data.

The DCS 2000 program has made significant improvements to the accuracy of data capture since the dress rehearsal and is continuing to do so.²¹ Errors will remain, however, because no method of capturing data from census questionnaires will be error free. We also recognize that at this late date, it is not advisable or feasible to make changes to the questionnaires. However, when refining and improving MRR, it is important for the bureau to be aware of problematic conditions that the software must handle. The following discussion is intended to raise the visibility of these conditions.

The questionnaires' lack of instructions or questions to obtain necessary information, phrasing of instructions and questions, and layout of blank fields where respondents write their data have confused some respondents. The quality of some responses indicates the following questionnaire issues:

²¹The most recent analysis conducted by the RIT Research Institute showed accuracy rates for short form mail return write-in data of between 99.36 percent and 99.56 percent including blank fields and between 97.33 percent and 98.14 percent excluding blank fields. For check box data, the accuracy rate including all fields was 99.47 percent and excluding blank fields was 98.82 percent. See *Joint R&D Project to Advance Technology for Data Capture System*, Management Summary for February 1999, RIT Research Corporation, Rochester, NY, March 15, 1999.

- The questionnaires do not explain that an automated system will read and interpret the responses and that for this automated process to work accurately, respondents must print well-formed characters in the spaces provided.
- The forms do not include a warning that the automated process cannot capture responses in red ink.
- The forms do not show how to mark an answer indicator box so it is machine-readable. Also, some questions require exactly one choice while others allow one or more. Some respondents appeared to be confused and supplied more than one response when exactly one was required.
- The forms do not provide separate space for a generational suffix field (such as Jr., Sr., III). The generational suffix helps distinguish between two person entries with identical first and last names, and lack of this information has led to suspected errors.

Some MRR test case results demonstrated that since there is no provision for the respondent to indicate that more than one household cohabits the same residence, when two families at the same census housing unit ID each submit their own non-BCFE questionnaire response, MRR will choose one family over another when both should be recorded at that address.

If the respondent omits certain information, the data cannot be compared to other more complete entries to detect matching responses for a census housing unit ID. As noted in the OIG report, *Columbia Dress Rehearsal Experience Suggests Changes to Improve Results of the 2000 Decennial Census*,²² there were numerous instances in SBE operations in which the same person was counted multiple times by enumerators but the questionnaires provided incomplete data. These duplicate responses were not detected because of insufficient information. Similarly, incomplete person responses on housing unit records, particularly from enumerator operations, have been observed by census personnel to impair correct resolution of multiple responses. The bureau should stress during enumerator training how incomplete data impacts the accuracy of census results.

Also, bureau analysts have expressed concern about the reliance of DCS 2000 on optical character recognition (OCR) and optical mark recognition (OMR), and are exploring ways to improve the keying of data that the system cannot accurately process. DCS 2000 was designed to identify questionable data and send it for human review and correction through the process

²²*Columbia Dress Rehearsal Experience Suggests Changes to Improve Results of the 2000 Decennial Census*, Inspection Report No. ESD-10783-8-0001, September 1998.

known as key from image (KFI).²³ During dress rehearsal, numerous instances of converted data were sent to KFI because of low confidence in OCR and OMR results. However, limitations in the types of review and corrections permitted during dress rehearsal did not always enable complete and efficient correction of erroneous data. In some cases, fuller portions of the questionnaire needed to be viewed for the KFI operator to know what the correction should be. Also, in dress rehearsal, the operators were instructed to key what they saw on the form without using human judgement, even if what was written was clearly invalid. For example, a respondent would write “May” in the month of birth field instead of the acceptable numeric value “05.” According to instructions, the operator was not to convert the letter version to the number. Discussions are underway on how best to develop “intelligent keying” guidelines to mitigate this limitation for the decennial.

The forms filled out by the public repeat the questions for each person. Not only does that require more paper for each questionnaire, it slows the respondent down by inhibiting an efficient way of answering the questions, and it complicates the data capture operation by requiring more sheets and folds of sheets of paper to be handled. During dress rehearsal, there were mismatched sheets among questionnaires as a result. The enumerator questionnaire for the short form has the most efficient format where data for the whole household can be completed and viewed in rows and columns across one open sheet. At this late date, we are not advocating questionnaire redesign, but in future censuses, developing questionnaires and instructions that better facilitate accuracy should be a top priority.



In its response, the bureau maintains that the questionnaires’ design is based on extensive cognitive research and is constructed to lead people to respond as accurately as possible. However, processing problems did arise from the questionnaires’ design as discussed in this section. Further, the bureau maintains that research has found that extensive instructions do not increase the accuracy of the responses. An approach that does not require extensive instructions might be to include a brief caveat at the beginning of the questionnaire that explains that the form is machine-read. This would alert the respondent to confine the answers to spaces provided. More balance in weighing the trade-offs between “user-friendly” and technically feasible could improve the quality of the processed response data.

C. MRR Rules and Software Warrant Review

Several members of the MRR team are analyzing results from MRR processing of dress rehearsal data because of questionable resolution of multiple responses. The two elements that they are

²³Data Capture System 2000 resorts to KFI, and Key-From-Paper (KFP) under specific conditions that indicate probable OCR and OMR errors.

primarily analyzing are the order in which PSA applies its selection criteria and the correctness of the matcher's rating of pairs of person data to determine if they represent the same person. This section discusses the rationale motivating this analysis and subsequent modification of the software to improve its accuracy in performing these functions.

During dress rehearsal, there were instances where the matcher results were not what the bureau analysts expected and did not appear to be accurate. Questionable matches were identified during processing of SBE data, which uses the same version of the matcher as MRR and the same constraints as the outside-census housing unit ID portion of WBS. In addition, the matcher has given persons sharing birthdays or close in age and having similar names scores that indicate they are duplicates. However, upon close examination of such duplicate pairs, we questioned whether such persons might be twins, cousins, or related in some other way. We also questioned removal of persons who had some data in common, but other data that was considerably different. Removing one of such matching pairs erroneously would make the final results less accurate. Review of MRR testing revealed similarly questionable results. The "Results and Analysis" section of the test plan acknowledges these instances and notes that a clerical procedure might preclude such questionable matches.²⁴

MRR uses the version of the matcher which the bureau modified for the 1995 test to accommodate commonly made typographical errors. The bureau now utilizes DCS 2000, which primarily uses OCR to capture handwritten responses. However, the bureau has not analyzed OCR errors that the matcher could also accommodate. Because OCR introduces different errors from keyed data, we are concerned that some of the decisions made by the matcher in its scoring process may not be appropriate. Since the input data is obtained through OCR, and keying when DCS 2000 has calculated a low confidence in the OCR data, the bureau needs to review matcher performance against these conditions to ensure accurate matching.

Because of concerns about the matcher software's performance in operations during the dress rehearsal, the lead programmer of MRR has initiated an effort to assess the matcher's performance on dress rehearsal data. The assessment will use a subset of the dress rehearsal data consisting of all households for which multiple returns were received. After establishing guidance for each member to use to determine whether persons from different returns match, the team is manually identifying all person matches across all returns for a household. This effort will establish a truth set to serve as the basis for assessing matcher performance. The team plans to use the matcher software to identify matches across these same returns and then determine how closely the matcher results correspond to the truth set established manually. Current plans are to adjust the parameter file used by the matcher to score the data until results agree closely with the results of the manual matching.

²⁴*Census 2000 Dress Rehearsal Test and Acceptance Plan for the Within-Block Search (WBS) and Primary Selection Algorithm (PSA)*, Bureau of the Census, September 11, 1998, p. 81.

Bureau management from the Decennial Systems and Contracts Management Office have made output data from DCS 2000 testing available to bureau personnel. Bureau management should continue this sharing of data to the extent necessary to improve the matcher and DCS 2000 in unison. For example, bureau personnel have developed the capability of viewing the questionnaire image in conjunction with the resulting computer standard format for each field. Use of such a tool would facilitate the debugging of both MRR and DCS 2000. We further discuss coordinated testing that checks the correctness of each step that processes the questionnaire data in a later section on beginning-to-end testing.

In addition to issues associated with the matcher element of MRR, questions remain on how effectively the PSA element applies criteria to choose members of a household. As noted above in the discussion of the report card and evaluation results, evaluation data provides the analysts a basis for determining how well the criteria performed and assessing whether the order in which they are used by the algorithm is appropriate. We support this effort, which will result in a revised set of software requirements that will aid in modifying the software to produce more accurate results. In a section to follow that addresses developing the software requirements specification according to software engineering standards, we cite specific areas of ambiguity that need to be addressed in a modified specification. The bureau's efforts to analyze the data and determine modifications to the algorithm that remove errors and ambiguities, together with a more rigorous approach to defining the software requirements and implementing that algorithm should result in improved MRR processing.

II. A More Structured Development Approach Would Improve MRR Software

Successful software development depends on defining requirements as early in the project as feasible. To ensure that software correctly implements the intentions of the users, requirements specifications must be complete, consistent, unambiguous, and verifiable. Use of a software engineering standard to guide subject matter experts through the process of defining requirements helps this effort. Applying such a standard to the development of the MRR specification would also be beneficial in determining how MRR will compensate for questionnaire design issues, clarifying expectations for input data quality, assessing the impacts of changes in requirements, and guiding the testing effort.

The bureau is fortunate to have a team of experienced and dedicated analysts and developers who are implementing MRR for the decennial census. The team hired a contractor who has assisted them in applying software engineering principles, which has had a very positive effect. For instance, in spite of an already compressed schedule, the team submitted new software to peer review and extended testing to a wider variety of cases. This section highlights the need for additional efforts of this type.

A. Bureau Software Engineering Standards Are Available but Not Widely Used

The bureau's software standards branch in the Office of the Associate Director for Information Technology has written a manual that describes how to write software specifications by accepted software engineering principles. Entitled *The Census Software Development Life Cycle*,²⁵ this manual unfortunately has not been used by the MRR team for its specifications. Members of the team who wrote the manual are still involved in software development for the bureau and are available to provide consulting support in applying the manual to the development process. They can help a development team tailor the manual to fit the needs and constraints of an organization. For instance, we believe applying the third section of the manual, "Software Requirements Definition and Analysis," which provides guidance on how to state rigorously what the software must accomplish, would substantially clear up MRR issues. Two other standards available within the bureau are:

- *Programming Standards and Guidelines Manual*, Bureau of the Census, March 1991, and
- *Decennial System and Contract Management Office Decennial Processing Systems Software Development Process*, Bureau of the Census, December 11, 1997.

The programming standards manual presents standards for how to write software code. It includes sample code and guidelines for conducting peer reviews, which can be invaluable to the process. It also cites excellent references. The second manual was developed by the on-site contractor and provides extensive and detailed instructions for many software engineering methods to be applied throughout development. The MRR code and testing efforts show evidence that developers have followed some methods advocated by these documents. However, conformance to a standard occurred because of individual preference as opposed to a strategy established by bureau management and supported by policy and training.

B. The Bureau Should Define MRR Requirements More Explicitly

To determine the appropriateness and quality of the requirements specification, we reviewed the key requirement of handling multiple responses associated with various form types and source operations, e.g., mail return, update/leave, and Be Counted program. For any household, MRR must handle multiple responses occurring in various combinations of form type and source operation. During the dress rehearsal, MRR had to handle a total of 20 distinct form types—9 long form types and 11 short form types—associated with 11 operations. Table 2 shows which forms were distributed by each operation, grouped by long and short form.

MRR processing takes several steps to determine how household and person data should be extracted from responses returned for a census housing unit ID. Decisions are based upon

²⁵*The Census Software Development Life Cycle*, Bureau of the Census, November 7, 1994.

Table 2

Operations and Form Types Handled in Dress Rehearsal

| Dress Rehearsal Source Operations | Long Form Types | Short Form Types |
|-----------------------------------|-----------------|------------------|
| Coverage edit follow-up | DX-2 | DX-1 |
| Mail return (MR) | DX-2 | DX-1 DX-1 |
| Update/leave | DX-2(UL) | DX-1(UL) |
| Update/leave ADD | DX-2A(UL) | DX-1A(UL) |
| MR Replacement | DX-2 | DX-1 |
| NRFU (barcoded) | DX-2(E) | DX-1(E) |
| | DX-2(E)(cf) | DX-1(E)SUPP |
| NRFU (not barcoded) | DX-2(E) | DX-1(E) |
| | DX-2(E)(cf) | DX-1(E)SUPP |
| Be Counted program (paper) | | DX-10 |
| Group quarters/special places | DX-15B | DX-15A |
| | DX-20B | DX-20A |
| | DX-21 | DX-21 |
| TQA Be Counted | TQA-10L | TQA-10S |
| Large household follow-up orphans | DX-2(HF) | DX-1(HF) |

NOT INCLUDED IN TEST DATA

TEST DECK TO DETERMINE MOST QUALIFIED RESPONSE FOR A HOUSEHOLD: ID1, ID2, ID3

values of variables such as the number of responses for a household, the number of people in a household, the completeness of the 100 percent population data, whether the response is on a long or short form, the source operation generating the response, and the date that the response was received. To describe how decisions are to be made by the software, the specification sets up hierarchical rules, where lower rules act as “tiebreakers” for cases where both returns rank the same according to all higher rules. However, because the specification does not explicitly list a representative set of cases of multiple responses, present the resolution logic, and specify the output, it is difficult to determine what the correct resolution of multiple responses should be under various conditions. For example, it is unclear when:

- Two person records on different returns are compared to see if they belong to the same household, whether a person record that has already been marked for deletion should still be utilized in the comparison.
- A person entry on a short form response contains more complete data than the corresponding entry on a long form response, how that data is consolidated into the long form response.
- There are more than two responses for a census housing unit ID, which pairs of responses are compared and how to reconcile inconsistent linking results.

Additionally, the specification cites the 100% Data Quality Index which measures the completeness of the population data for persons reported on the response, but the specification does not explain how to calculate the index. Also, it is unclear why more complete data for a household, as measured by the index, does not carry more weight than its relatively low position in the rule hierarchy and processing steps currently credit it. Finally, the specification does not define contingencies in processing data that does not conform to items listed in the section titled “Operational Constraints and Assumptions.” Omissions such as those cited here leave the specification incomplete for purposes of software development. A more comprehensive specification written according to the “Software Requirements Definition and Analysis” chapter in the bureau’s life cycle manual would help resolve these issues, as well as facilitate testing.



In addition to its current efforts to develop rules and specifications for processing incomplete data, the bureau needs to implement error handling procedures. Robust software is designed to handle input conditions consisting of data values that fall outside permitted values. These error handling procedures may attempt to fix the data to conform to permitted range values when feasible. When data is recognized as erroneous, the software may also report these conditions for further analysis.

Bureau standards that were cited on page 16 of this report offer some guidance about specifying and testing erroneous data conditions:

- The Census Software Development Life Cycle: On pages 94 through 98 of this manual, the contents of the specification of a requirement are described. According to this manual, the processing section should include a description of “responses to abnormal situations.” The analyst could describe how to handle (fix or omit this data and report action) in this portion of the specification.
- Decennial System and Contract Management Office Decennial Processing Systems Software Development Process: Appendix B of this manual contains a description of various types of testing such as “equivalence partitioning” and “boundary condition testing.” The requirements specification would describe values or ranges of values that would be used for this testing.

A. MRR Software Testing Needs to Be More Complete

The MRR team conducted testing from April through September 1998 in preparation for the dress rehearsal. The testing began by breaking the specification’s narrative description into individual requirements and documenting this information in a test plan. The test plan describes the requirements to be tested, test cases and the associated procedures, as well as the test data (referred to as test decks) for executing the procedures.

The test plan identifies 59 testable requirements. In addition, the plan describes test procedures to cover 43 test cases, where each procedure addresses one or more requirements. The test plan also provides guidance for the development of test decks, which are intended to satisfy the conditions of one or more of the test cases. It also includes a section in which results from running the software with each test deck are recorded with an indication of success or failure, along with an explanation of the nature of any failures. The final test plan, which indicated successful processing of each test deck, was released on September 11, 1998.

The test plan served as the basis for a commendable testing effort that was performed with great care under severe time constraints. The test decks developed by the test team afforded valuable input data for the developer to correct software errors in preparation for testing with dress rehearsal data when it became available. A more complete set of test cases would further enhance this testing effort. In Table 2, entries highlighted in red denote conditions of form type and source operation that were not tested. For instance, the operation of coverage edit follow-up was not tested for either the short or long form. Additionally, the operations update/leave and update/leave ADD were not tested for the short form, and so on. In all, 14 operation-form type combinations out of a possible 30 were not tested. More tests addressing these omissions and ensuring that the software handles the test data as intended by the requirements are needed to verify the correctness of the software.

As a further example, the test deck designed to test the choice of the most qualified response for a household could include many more cases of multiple responses. This test deck uses the data for three census housing unit IDs. The instances of form type and operation included are highlighted in Table 2 in green, yellow, and blue; i.e., there are two instances each for ID1 and ID2 and four instances for ID3. We suggest that multiple responses representing several more combinations of form type and operation be tested to verify that the software chooses the most qualified response for a household.

B. Beginning-to-End Testing Is Needed to Ensure Consistent Results

In addition, beginning-to-end testing should be conducted with data containing comprehensive multiple response instances for the questionnaire data path from DCS 2000 to creation of the CUF. Beginning-to-end testing starts with scanned images of the completed forms and proceeds with formatting the data contained in DRF1. The next step is the linking of continuation forms and other processing that produce DRF2. The results of MRR processing of DRF2 data should be analyzed for correctness. This testing is particularly important in ensuring that the results of the matcher software are correct with respect to captured data and would check the correctness of all pre-edit files.²⁶ If the quantity and variety of data are large enough, this effort will also stress test the software.²⁷

Enabling headquarters processing personnel to view the scanned questionnaires' image data would provide a further opportunity for quality assurance, and hence improved accuracy of the scanning procedures. The results of beginning-to-end testing should be used, as appropriate, to adjust matcher parameters to fine-tune the scoring process or, if necessary, to modify matcher software. Beginning-to-end testing should be repeated as modifications to DCS 2000 affect the captured data.



The bureau has concurred with our recommendation for beginning-to-end testing and cited that the Decennial Integrated System Test (DIST) is designed to fulfill this recommendation. However, this testing, although valuable and needed, will start in October, 1999, which is considerably earlier than MRR's development and testing time frame. In addition, documentation indicates that the DIST will treat headquarters processing of response data as a single entity. The limited amount of input data planned for the DIST may not sufficiently test the conditions that MRR is specified to handle. To meet this recommendation fully, the bureau needs to conduct beginning-to-end testing with completed questionnaires that reflect a sufficient variety of multiple response scenarios, submit the questionnaires to the processing stream starting

²⁶Pre-edit files contain all questionnaire data. Edited files contain subsets of the questionnaire data.

²⁷Stress testing is designed to overload the system in various ways.

with DCS 2000 and continuing with headquarters processing including MRR. This testing would not treat headquarters processing as a single entity but would test each component of headquarters processing with testing tools that allow analysts and developers to monitor processing and isolate software errors. By using these tools, bureau personnel would be able to track processing decisions within each step, as well as to review the results of each step to find errors.

III. Improvements Are Needed to Make MRR Processing Timely for the 2000 Census

The Census Bureau has projected that MRR processing for the decennial census will need to be completed within 30 days to ensure timely availability of data for follow-on processing operations. Results for dress rehearsal indicate that unless improvements are made, MRR processing for the decennial census will require approximately 87 days. If MRR processing cannot be completed within the required time constraints using available computing resources, follow-on operations will be impacted and census results will be late. Therefore, the bureau needs to streamline MRR software design and data organization, upgrade computer resources, or institute some combination of streamlining and upgrading to ensure timeliness of MRR results.

To assess the efficiency of MRR software design and data organization, as well as the adequacy of the computer systems planned to run MRR software for the decennial census, we reviewed MRR processing times for the three dress rehearsal sites. Table 3 shows the total elapsed time required on a dedicated computer system to execute the MRR software for each dress rehearsal site. The bureau planned to use the same headquarters computer system for MRR processing during the decennial census. As noted, the bureau has allotted 30 days for MRR processing during the decennial census. The total time to perform MRR processing for the approximately 400,000 households in the dress rehearsal was 6.95 hours. Under the same conditions, it would take roughly 300 times as long—2,085 hours or 87 days—to perform the MRR processing for the 120 million households expected in the decennial census. The bureau clearly will not have nearly this much time available. Even if the bureau were to eliminate the second mailing of questionnaires, the time required to perform MRR processing would not be significantly affected since fewer than 6 percent of households responded to both the initial and second mailings for the dress rehearsal.

The MRR team has identified several ways to reduce the processing time. For example, developers are considering using binary files, rather than ASCII files, to represent DRF2, the data

file which serves as input to the WBS and PSA software.^{28, 29} In general, binary files provide a more compact representation of numeric data (e.g., age, number of persons in a household, year of birth) than ASCII files. Additionally, binary files provide computational efficiency over ASCII files in that numeric data does not need to be converted from its ASCII representation before computations can be performed using data read from the file. Developers have also found that using vendor-provided file compression software reduces both processing times and the amount of disk storage used.

Table 3
MRR Processing Times in Dress Rehearsal

| | Menominee | Columbia | Sacramento | Total (all three sites) |
|----------------|-----------|----------|------------|-------------------------|
| WBS processing | .0190 | 3.1830 | 1.9405 | 5.142 |
| PSA processing | .0165 | 1.1365 | .6550 | 1.808 |
| Total | .0355 | 4.3190 | 2.5950 | 6.950 |

Table entries represent processing time in hours.

The bureau has decided not search for matches for any persons outside the census housing unit ID for which they are reported, which is a function performed by WBS for dress rehearsal. Based on the evaluation studies for this element, bureau personnel have observed that few matches were found and that the block-wide search conducted for dress rehearsal accounts for 70 percent of all the processing time for MRR. Therefore, simplifying the design by eliminating searching for person matches outside the census housing unit ID would realize considerable processing-time savings. If the bureau determines that the trend observed in the dress rehearsal will likely be the same for the census, then this would be a valid design change.

Bureau managers and staff have expressed concern not only about the adequacy of headquarters computing resources to support MRR, but other decennial census operations as well. As a result, we believe that the MRR team should continue to pursue ways to streamline MRR software design and data organization and that bureau managers should upgrade headquarters computer

²⁸Binary pertains to a number system that has just two unique digits, 0 and 1. Computers are based on the binary number system.

²⁹ASCII is the acronym for the American Standard Code for Information Interchange. ASCII is a code for representing English characters as numbers, with each character assigned a number from 0 to 127. Most computers use ASCII codes to represent text, and files used to store ASCII coded text are commonly referred to as ASCII files.

resources to ensure timely processing for MRR and other required operations for the decennial census.

CONCLUSION

As a result of the foregoing issues, the bureau must make important decisions concerning MRR requirements and design and perform significant development and testing work. The bureau has recognized many of the problems that we have discussed and is performing analysis of the dress rehearsal and evaluation data and working on improvements to MRR, as well as automated data capture. However, the bureau has not fully defined the activities needed to refine MRR and complete its development. Because of the complexity and amount of work remaining to be done in time for the decennial, the bureau needs to define the remaining tasks and plan carefully for their completion.

The bureau needs to produce a rigorous and complete MRR requirements specification and should use its software engineering standard, *The Census Software Development Life Cycle* manual to do so. The bureau also needs to develop a comprehensive test plan following the same template used for the dress rehearsal to ensure that the software correctly implements the MRR requirements. Finally, the bureau needs to complete software development and testing in accordance with the specification and test plan. To accomplish these tasks successfully, the bureau should prepare a plan that identifies all of the decisions to be made and work to be performed, along with the responsible organizations, a realistic schedule, and a complete list of milestones.

MRR is only one of many processing elements that manipulates questionnaire data from the public and is used to create the final decennial tabulations. The same structured software development process that is needed for MRR, including clear and complete requirements and thorough testing, is essential for all software. Ensuring the correctness of the software also requires beginning-to-end testing of all processing elements. Use of more rigorous software development methods and sharing data among processing elements in a coordinated fashion will help the bureau secure accurate results for the 2000 Decennial Census.

RECOMMENDATIONS

We recommend that the Director of the Census Bureau direct senior management for the 2000 Decennial Census to take the necessary actions to:

1. Establish an appropriate standard of measuring the success of MRR processing.

The bureau concurred with this recommendation.

2. Identify both the accuracy of MRR in resolving multiple responses as well as the impact of MRR errors on the total enumeration when reporting on the success of MRR.

The bureau concurred with this recommendation.

3. Define and implement “intelligent keying” guidelines to process data sent to the KFI component of DCS 2000.

The bureau stated that it has developed and is reviewing such guidelines and will report to the OIG as soon as a decision is made.

4. Develop a rigorous and complete revision of the MRR software requirements specification by:

- a. Obtaining assistance from the Office of the Associate Director for Information Technology in using Chapter 3 of the *Census Software Development Life Cycle* manual to develop a revised software requirements specification.
- b. Specifying a revised set of software requirements necessary to implement PSA with the highest probability of resolving multiple responses accurately as determined by the MRR team’s analysis of the evaluation data.
- c. Specifying a revised set of software requirements for use by the matcher by creating a team of DCS 2000 contractor and bureau analysts to analyze data produced by DCS 2000 and corresponding matcher output.
- d. Specifying the software requirements necessary to address data quality issues, such as how to handle incomplete, inconsistent, or erroneous data.

The bureau concurred with Recommendations No. 4a through 4c and requested clarification on No. 4d, which is provided on pages 18-19.

5. Plan, schedule, and implement all tasks needed to complete MRR definition, development, and testing, including:
 - a. Developing the revised MRR software requirements specification in accordance with Recommendation No. 4.
 - b. Evaluating approaches to streamline MRR design and data organization.
 - c. Developing a test plan based upon the software requirements, using the same template that the MRR team used for the dress rehearsal test plan.
 - d. Developing the software in accordance with the specification.
 - e. Testing the software in accordance with the test plan.
 - f. Performing beginning-to-end testing of the questionnaire data processing stream from DCS 2000 through MRR using available debugging tools.

The bureau concurred with Recommendation Nos. 5a through 5f. Further description of beginning-to-end testing that would more fully satisfy Recommendation No. 5f is provided on pages 20-21.

6. Assess whether selective use of clerical quality assurance procedures would be cost effective as part of MRR.

The bureau stated that this recommendation has been resolved.

7. Determine the additional headquarters computing resources needed to ensure timely processing of multiple response resolution and all other required operations for the decennial census, and obtain these resources.

The bureau stated that this recommendation has been resolved.

8. Revise enumerator training to emphasize the importance of obtaining complete and accurate input data because poor response data precludes accurate processing.

The bureau stated that this recommendation has been resolved.

The bureau's complete response is included as the appendix of this report.



UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

OFFICE OF THE DIRECTOR

AUG 9 1999

Appendix

MEMORANDUM FOR Judith J. Gordon
Assistant Inspector General for Systems Evaluation

Through: Robert J. Shapiro *RJS*
Under Secretary for Economic Affairs

From: Kenneth Prewitt *Kenneth Prewitt*
Director

Subject: *Improvements Needed in Multiple Response Resolution to Ensure Accurate, Timely Processing for the 2000 Decennial Census.*
Draft Inspection Report No. OSE 10711

This is in response to your memorandum dated June 11, 1999, which was received at the Census Bureau on June 17, 1999. In reviewing this draft report, Census Bureau officials have expressed concerns about the following observations and conclusions reached by the auditors:

- Page 9 – The Census Bureau agrees that there needs to be an appropriate standard to which the MRR results should be held. However, the report implies that the Census Bureau amended the accuracy standards included in the *Census 2000 Dress Rehearsal Report Card* only after “recognizing the controversy within its own staff.” Rather, these changes occurred after a more careful analysis of the results of the Dress Rehearsal and resulted from the type of discussion and review that constitutes a normal part of Census Bureau operations.
- Page 10 – With respect to the data presented in *Table 1--Dress Rehearsal Multiple Response Rates*, it is important to stress that the columns in the table should not be viewed in isolation. The whole table is important. For instance, even a 10% error would have little impact on the census if there were a small number of multiple responses.
- Page 11 – The report states that “the questionnaire’s lack of instructions or questions to obtain necessary information, phrasing of instructions and questions, and layout of blank fields where respondents write their data have confused some respondents.” The Census Bureau agrees that instructions for answering each question were not included in the questionnaire. However, the design of the questionnaire was the result of extensive cognitive research to ensure that it was constructed in such a manner as to lead people to respond as accurately as possible. This research indicates that extensive instructions do not increase the accuracy of the responses.

- Page 12 – The report states that “some MRR test case results demonstrated that since there is no provision for the respondent to indicate that more than one household cohabits the same residence, when two families at the same census housing unit ID each submit their own non-BCFE questionnaire response, MRR will choose one family over another when both should be recorded at that address.” This has nothing to do with questionnaire design. It results from the application of the rule that the Census Bureau developed to address this issue. This rule is currently being evaluated for efficacy based on data from the Dress Rehearsal. It can and should be changed if that evaluation indicates that a change is necessary. However, at this stage, Census Bureau staff believe that such a change would lead to more error.
- Page 14 – The report states that “removing one of such matching pairs erroneously would make the final results less accurate.” Although this may be true, it addresses only one of two types of potential error. The Census Bureau is just as concerned that leaving in such pairs erroneously (that is, when there is a true match) also has an important adverse effect on the accuracy of the results.
- Page 16 – The report states that “the Bureau should define MRR requirements according to its standard.” The Census Bureau will hold to whatever standards we develop. However, it is not possible to specify a unique standard for each distinct population under a probabilistic matcher algorithm. The probabilistic matcher aggregates information from a variety of variables to determine whether individuals match. Therefore, we apply an overall standard to the accuracy of a probabilistic matcher.
- Page 19 – The report states that “enabling headquarters processing personnel to view the scanned questionnaires’ image data would provide a further opportunity for quality assurance, and hence improved accuracy of the scanning procedures.” The Census Bureau agrees that the quality of scanned data is important for MRR, as well as other uses, and adequate quality assurance is an important component. Decennial Systems and Contracts Management Office (DSCMO) and Decennial Statistical Studies Division (DSSD) are working with the contractors on this issue. However, the need for headquarters quality control as a necessary component of the production MRR system has not been established.

In response to the specific recommendations contained in the draft audit report, the Census Bureau submits the following.

1. Establish an appropriate standard of measuring the success of MRR processing.

The Census Bureau Concurs: The Census Bureau recognizes the value of establishing a standard for MRR processing. However, there is great difficulty in developing a methodology for defining empirical variables to use in the evaluation of MRR processes.

Census Bureau staff are continuing to evaluate the results of the Dress Rehearsal in order to develop the most effective standards possible. Once standards are finalized, Census Bureau staff will need to conduct an evaluation after the census to ascertain whether the standards were met. The Census Bureau will report to the Office of Inspector General (OIG), as staff continues to explore this issue.

2. *Identify both the accuracy of MRR in resolving multiple responses as well as the impact of MRR errors on the total enumeration when reporting on the success of MRR.*

The Census Bureau Concur: Census Bureau staff are currently exploring the possibility of conducting evaluations based on reinterviews after Census Day. Staff are also looking into the possibility of conducting an evaluation based on Accuracy and Coverage Evaluation data. Draft plans for both procedures are under development.

3. *Define and implement "intelligent keying" guidelines to process data sent to the KFI component of DCS 2000.*

A specification for providing data entry staff with more interpretive guidelines for entering data has been developed and is currently under consideration. The Census Bureau will report to the OIG as soon as a decision is made regarding this matter.

4. *Develop a rigorous and complete revision of the MRR software requirements specification by:*

- a. *Obtaining assistance from the Office of the Associate Director for Information Technology in using Chapter 3 of the "Census Software Development Life Cycle" manual to develop a revised software requirements specification.*

The Census Bureau Concur: As appropriate, Census Bureau staff have obtained and will continue to seek assistance from the Office of the Associate Director for Information Technology in decennial census operations, including MRR.

- b. *Specifying a revised set of software requirements necessary to implement PSA with the highest probability of resolving multiple responses accurately as determined by the MRR team's analysis of the evaluation data.*

The Census Bureau Concur: A sub-team of staff members from the MRR team is currently revising software requirements based on data from the Dress Rehearsal. The team plans to complete preparation of a draft of these requirements in fall 1999.

- c. *Specifying a revised set of software requirements for use by the matcher by creating a team of DCS 2000 contractor and Bureau analysts to analyze data produced by DCS 2000 and corresponding matcher output.*

The Census Bureau Concurs: Census Bureau staff from DSSD and DSCMO are analyzing Dress Rehearsal data and testing matcher output software

- d. *Specifying the software requirements necessary to address data quality issues, such as how to handle incomplete, inconsistent, or erroneous data.*

The Census Bureau seeks clarification from the OIG regarding this recommendation. Census Bureau staff are developing rules and specifications for processing incomplete data and will work with the OIG to ensure that any concerns regarding software requirements are addressed.

- 5. *Plan, schedule, and implement all tasks needed to complete MRR definition, development, and testing, including:*

- a. *Developing the revised MRR software requirements specification in accordance with the Recommendation #4.*

The Census Bureau Concurs: As mentioned above, Census Bureau staff plan to complete preparation of a draft of the software requirements in fall 1999.

- b. *Evaluating approaches to streamline MRR design and data organization.*

The Census Bureau Concurs: Census Bureau staff have developed revised procedures to streamline MRR design and data organization. The OIG will be provided with a report summarizing these procedures.

- c. *Developing a test plan based upon the software requirements, using the same template that the MRR team used for the dress rehearsal test plan.*

The Census Bureau Concurs: A sub-team to develop a test plan is in place. They are currently awaiting the specifications. Some members of this team will monitor the development of the requirements as they are being finalized to get a feel for them. Review and acceptance of the test results is scheduled for April 3, 2000 through August 4, 2000.

d. Developing the software in accordance with the specification.

The Census Bureau Concur: Software development is scheduled to begin in September 1999 and will continue through the formal testing phase to production in early fall 2000. Revisions of the Dress Rehearsal software are being made now to test changes designed to speed up and optimize the general software. Specification changes for 2000 should not change the basic design of the software, only the rules.

e. Testing the software in accordance with the test plan.

The Census Bureau Concur: See response to Recommendation #5(c).

f. Performing beginning -to-end testing of the questionnaire data processing stream from DCS 2000 through MRR using available debugging tools.

The Census Bureau Concur: The Decennial Integrated System Test is designed to fulfill this recommendation and will be implemented throughout the data processing operation.

6. Assess whether selective use of clerical quality assurance procedures would be cost-effective as part of MRR.

Resolved: Census Bureau staff have explored this and concluded that it is not possible to implement a clerical review under the current schedule.

7. Determine the additional headquarters computing resources needed to ensure timely processing of multiple response resolution and all other required operations for the decennial census, and obtain these resources.

Resolved: The Census Bureau upgraded the DEC processors to the 8200 series and tripled existing disc space. In addition, processing will be done with binary (not ASCII) files. These enhancements will now provide for timely processing of multiple-response resolution and all other required operations for the decennial census.

8. Revise enumerator training to emphasize the importance of obtaining complete and accurate input data because poor response data precludes accurate processing.

Resolved: Training already emphasizes this. The training manual has been revised to ensure that enumerators are given clear directions to collect the most accurate data possible.

cc: US/EA