U.S. DEPARTMENT OF COMMERCE
Office of Inspector General

PUBLIC RELEASE

BUREAU OF THE CENSUS

Data Capture System 2000 Requirements and Testing Issues Caused Dress Rehearsal Problems

Inspection Report No. OSE-10846 / January 1999

Office of Systems Evaluation

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Census Bureau is conducting the 1998 Dress Rehearsal to test methods and systems that will be employed for the 2000 Decennial Census. Beginning in the spring of 2000, the Census Bureau will collect and process data from approximately 120 million households, and by the end of the year, will deliver to the President state level population counts for allocating seats among the states in the House of Representatives. These operations will require the capability to capture data from an estimated one billion pages of census forms within a four-month period. To accomplish data capture, a state-of-the-art system, Data Capture System 2000 (DCS 2000), will be employed. In a marked departure from any previous decennial census, DCS 2000 is being developed by a contractor, rather than internally by the bureau. The DCS 2000 contract is divided into two overlapping phases. Phase I was the design and development of a pre-production version of the system for use in the dress rehearsal and is the subject of this report. Phase II is the development of the full-scale production version, which will be used for the 2000 Decennial Census.

We conducted this evaluation to determine whether the performance of DCS 2000 during the dress rehearsal indicates that it is likely to be able to accurately capture data within the required time constraints during the 2000 Decennial Census. In addition, we assessed problems encountered during the dress rehearsal to identify improvements that are needed to help ensure a successful decennial census.

We found that during the dress rehearsal, DCS 2000 experienced numerous serious problems in processing dress rehearsal forms resulting from inadequate control of requirements and insufficient testing.[1] Despite the problems, the data capture system met all of its processing deadlines. However, the size, complexity, and performance requirements of the decennial census mean that similar problems during the decennial would introduce a high risk of not being able to complete data capture operations on time and could produce data of questionable accuracy. We believe that with strict requirements management, comprehensive testing, and sufficient funding, the problems experienced with DCS 2000 during the dress rehearsal can be solved, and the system will be capable of performing as needed during the decennial census.

The data capture contractor had planned to use rigorous and well-defined system engineering procedures, but continuing growth and change in requirements caused the contractor to abandon its planned approach. Instead, concurrent development, testing, and deployment activities were

[1] Requirements are the functional and performance capabilities that a system has to provide and are the basis for system development. Requirements growth and change must be controlled so that system capabilities can be implemented within an established cost and schedule.

performed on a short cycle that did not allow enough time to consistently apply sound system engineering practices, including systematic software and system testing. (See page 6.)

The bureau also had developed a comprehensive and rigorous test program for DCS 2000 that was designed to identify and correct problems early and to validate that the system functioned properly before dress rehearsal. However, funding shortfalls and the disruption to the system engineering approach caused by requirements instability made it necessary to reduce the size and scope of the test program. As a result, many problems that should have been identified before dress rehearsal were not found until dress rehearsal. (See page 8.)

In view of the above problems associated with DCS 2000 requirements and testing, the Census Bureau and the contractor are to be commended for their efforts in analyzing and correcting the problems that occurred, spending many more hours on problem resolution than originally anticipated. Efficient review teams, rapid software development, effective analysis and communication of the issues, and a committed staff were all key factors in correcting the problems and making the system work.

At our exit conference, bureau officials told us that they recognize the need to improve the management of requirements not only for data capture but for all aspects of the decennial census, and they have convened a steering group comprising the decennial operational managers to implement a requirements control process. Improving requirements management is imperative, and the bureau has taken an essential step to do so. Nonetheless, the bureau expects continued and intense pressure on requirements from parties both within and outside the bureau seeking refinements, additions, and changes to planned operations and procedures. Many proposed changes will have merit individually but may be prohibitive from a cost or schedule perspective or may have unanticipated effects on other operations. We believe that control of requirements is an absolute necessity for achieving a successful decennial census and that bureau senior management officials must continue to support and strengthen the requirements control process.

We recommend that the bureau strengthen the requirements management process for DCS 2000. We also recommend that the bureau establish schedules with sufficient time and provide adequate funding to perform complete and improved testing of DCS 2000, including operational testing. Our complete recommendations begin on page 12.

The Census Bureau has agreed with the conclusions of this report, and will implement our recommendations as detailed in its response, which appears as Appendix A.

INTRODUCTION

The Census Bureau is currently conducting the 1998 Dress Rehearsal to test methods and systems that will be employed for the 2000 Decennial Census. The dress rehearsal has three major phases: data collecting, data capture, and data reporting. Because of the importance of the dress rehearsal to the success of the decennial census, we are evaluating major information technology components used to conduct the dress rehearsal. As a result of our initial work, we have selected the following major components of the census process to evaluate: data capture, decennial field interface, and headquarters processing. This report addresses data capture.

PURPOSE AND SCOPE

The objective of this evaluation is to determine if the performance of DCS 2000 during the 1998 Dress Rehearsal indicates that it is likely to be able to accurately capture data, within the required time constraints, during the decennial census. The evaluation also assesses the reasons for problems encountered during the dress rehearsal and identifies improvements that are needed to help ensure a successful census.

During our evaluation, we interviewed Census Bureau and contractor personnel involved in the development and testing of DCS 2000 and its operation for the dress rehearsal. We met with census staff within the Decennial Systems and Contracts Management Office at bureau headquarters and within the Data Preparation Division in Jeffersonville, Indiana, as well as with contractor personnel supporting these offices, including the DCS 2000 contractor. During the dress rehearsal, we observed the operation of DCS 2000 at Jeffersonville. We also reviewed system development documentation, trouble reports, and program management review data.

Our work was performed in accordance with the *Standards for Inspections* issued by the President's Council on Integrity and Efficiency.

BACKGROUND

Beginning in the spring of 2000, the Census Bureau will collect and process data for approximately 120 million households, and by the end of the year, will deliver to the President state level population counts needed to determine the allocation of seats among the states in the House of Representatives. These operations will require the capability to capture the data from

an estimated one billion pages of census forms within a four-month period commencing on March 8, 2000. In order to capture this amount of data with such severe time constraints, a state-of-the-art data capture system, DCS 2000, will be employed.

In a marked departure from any previous decennial census, DCS 2000 is being developed by a contractor rather than internally by the bureau. A contract for the design and development of DCS 2000 was awarded on March 21, 1997. The DCS 2000 contract is divided into two overlapping phases. Phase I was the design and development of a pre-production version of the system for use in the dress rehearsal and is the subject of this report. Phase II is development of the full-scale production version that will be used for the 2000 Decennial Census. DCS 2000 development is being performed at the bureau's central computing facility in Bowie, Maryland.

Data Capture System 2000

DSC 2000 is designed to provide a complete data capture solution for the Census Bureau, encompassing the check-in of forms received from the public, census enumerators and other sources, high-speed direct electronic imaging of the data on those forms, and preparation of that data for tallying and statistical processing. As shown in Figure 1, the system must:

- Check-in forms,
- Perform high-speed electronic imaging to digitally capture and process the forms,
- Automatically convert the image data into ASCII[2] files,
- Provide corrective actions for data that fails automatic conversion, and
- Prepare output files of the captured data.

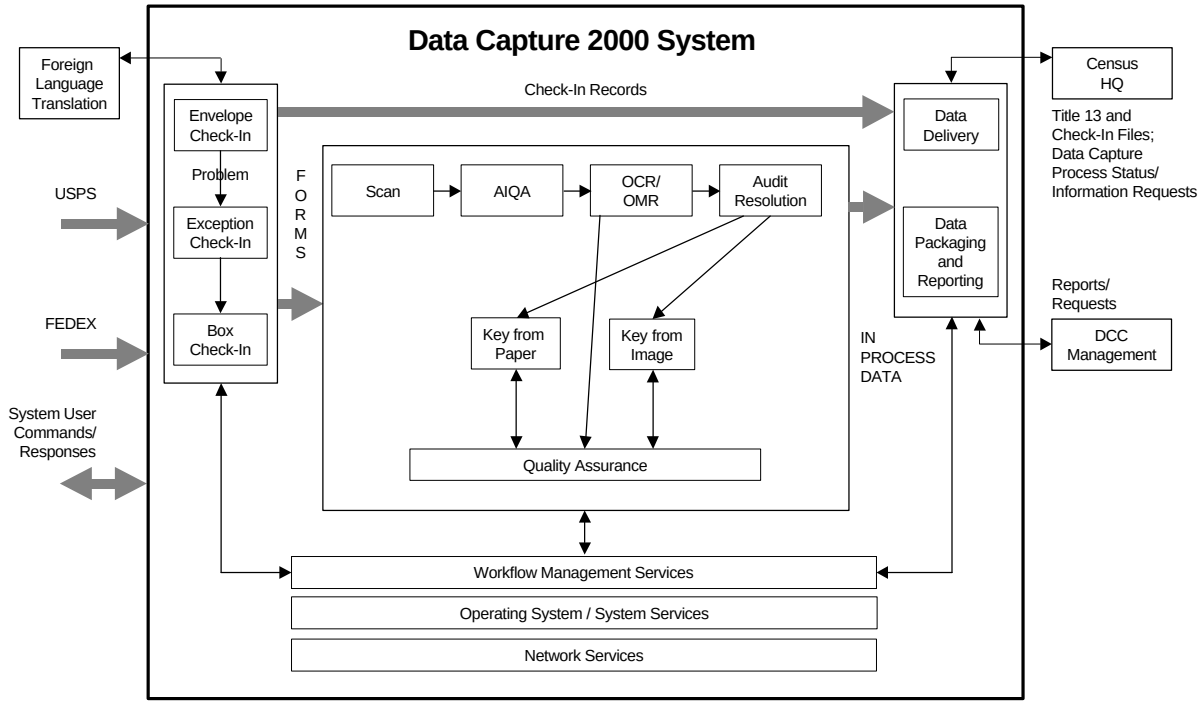
The data capture process begins with the check-in of returned census-related forms (e.g., respondent, continuation, enumerator-filled, group quarters, and Be Counted). Forms in languages other than English or Spanish are sent for translation prior to processing. The check-in step is performed by a special purpose machine that slices open the envelopes containing the forms and sorts them into pockets in priority order based on information encoded in a pre-printed bar code. Check-in data, indicating which forms have been received, is developed during this step and delivered to the Census Bureau headquarters for use with non-response follow-up. This procedure is used by Census to count those households that do not provide a timely response by sending a census enumerator to collect the information in person.

[2]ASCII 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.

Once checked in, the forms are manually prepared for scanning. An electronic image of the form is then captured by a scanner. Workflow software guides the captured image through automatic processing by passing it through a series of steps. This processing begins with the automated image quality assessment (AIQA) operation, which checks image quality and registration (correct alignment of the document during scanning). Following quality assessment, the image is passed in succession to optical character recognition (OCR) and optical mark recognition (OMR) for conversion of handwritten data on the forms into ASCII data. Images with errors generated by character or mark recognition failures are sent to audit resolution for corrective action. Corrective actions include key from image, which allows operators to verify imaged fields that have a low accuracy confidence score, and key from paper for images that fail.

At this point in the process, DCS 2000 will merge all ASCII data generated from data capture and produce ASCII files, which are transmitted on a scheduled basis to headquarters, via Census Bureau network services, for statistical processing.

Figure 1. Processing Flow of DCS 2000



OBSERVATIONS

During the dress rehearsal, DCS 2000 experienced many serious problems resulting from inadequate control of requirements and insufficient testing. A similar number and severity of problems during the 2000 Decennial Census would introduce a high risk of not being able to complete data capture operations on time and could produce data of questionable accuracy. Problems encountered during the decennial census will be magnified as compared to the dress rehearsal because of the decennial's much greater size and complexity. Whereas dress rehearsal data capture operations were conducted at one site (Jeffersonville), which processed forms from approximately 400,000 households, these operations for the decennial census will be conducted concurrently at four geographically dispersed sites, which must process forms from approximately 120 million households—a three-hundredfold increase. Moreover, each center will be staffed by significant numbers of inexperienced, temporary workers.[3]

In April 1998 we visited Jeffersonville to observe the operation of DCS 2000. Although the bureau had expected some problems to occur with DCS 2000 during dress rehearsal, it had planned for the system to perform in a fashion similar to the decennial census. However, at the time of our visit, the installed data capture system was experiencing many more problems than the bureau expected in processing dress rehearsal forms. Bureau and contractor personnel reported 250 critical trouble reports[4] before dress rehearsal, and nearly half were still open when the dress rehearsal began. A total of 398 critical trouble reports was recorded through the completion of dress rehearsal. Examples of critical trouble reports that were open *during dress rehearsal* included the following:

- Pipeline manager failure, unable to notify workflow of new batch
- Sorter generates duplicate identification numbers
- Validation of identification numbers not properly functioning
- Pipeline manager crashes after 40 batches
- Automatic backup of Title 13 data fails
- Key controller goes into infinite loop
- English forms being treated as foreign language
- Cannot generate check-in files
- Keying install deletes in-progress batches

[3] Three of the data capture centers will be operated under the contract. The fourth will be operated by the Census Bureau at Jeffersonville.

[4] Trouble reports are assigned a priority indicating the urgency with which a resolution is required. The priority order is low, important, and critical. The DCS 2000 program defines a critical trouble report as a problem that prevents the essential functioning of the system.

- Production Control hangs if processed tray has missing forms
- Pipeline manager crashes (out of memory)
- Missing records in Jeffersonville databases

All but 2 critical trouble reports were closed (i.e., the problem reported was corrected as of November 1998).

During our visit, we observed problems with check-in, scanning, and automatic image processing. Check-in problems included double-feeding and missorting of envelopes and misreading of bar codes caused by inconsistencies in their specification and lower than expected photo eye capability.

Poor paper quality, resulting in extremely high levels of paper dust, caused significant problems for the scanning operation. The high level of dust impaired the feeding of paper into the scanner by sticking to the rollers, slowly making them slick and slightly larger, and by adhering to the sensors that detect more than one form being fed into the scanner. With diminished ability to detect double feeds, loss of friction, and increased roller size, a higher than expected error rate resulted from feeding the forms into the scanner. The paper dust also collected on the scanners' optical guide bars causing extraneous marks and lines to be introduced on many of the captured images. The level of dust was so high that a periodic maintenance operation that was to be performed every four hours had to be performed at 15-to-20 minute intervals. The need to perform routine maintenance with much greater frequency was distracting and may have increased the possibility that some forms were missed during scanning operations[5].

Once the image was captured, workflow software that was designed to move the image through automatic processing steps (i.e., AIQA, OMR, OCR) failed periodically, as did some of the steps themselves. These failures required manual intervention by the contractor's on-site personnel to correct or, in many instances, restart the entire process. As a consequence, backlogs developed in the data capture system that made it unable to keep pace with the scanners. Scanning was then forced to wait until downstream automatic operations were able to process the backlogs. According to bureau officials, this precluded the use of one of the three scanners the system was designed to support due to the downstream performance problems.

The problems noted above interacted to introduce another problem—unaccounted forms. An unaccounted form results when its bar code is captured at only one of two steps, either check-in or scanning. According to the contractor, approximately 4,000 forms are "missing" from the dress rehearsal. A variety of errors at check-in and scanning caused this problem. These errors included:

[5] Although paper quality caused significant problems for data capture, analysis of this problem is outside the scope of this report.

- Invalid bar codes being entered into the system at check-in,
- Valid bar code being entered for blank forms and not tagged as such,
- Double-feeding of forms at scanning,
- Mishandling of forms at the scanner due to the high rate of operator intervention, and
- Scanning of forms from envelopes that had been missed at check-in due to double-feeding of envelopes.

Many of the problems with unaccounted forms have been corrected; however, an exact accounting of all forms must be assured prior to any processing for the census.

The principal reasons for the data capture problems encountered during dress rehearsal are discussed below.

I. Inadequate Control of Requirements Disrupted System Development

The success of software intensive projects, such as DCS 2000, depends on the precision and completeness of the understanding between the user and the implementor as documented by requirements. Software errors are frequently attributable to problems with or misunderstandings of requirements, and errors related to requirements generally are the most expensive to fix. Consequently, every reasonable effort should be made to precisely define system requirements as early in the project as feasible.

However, on DCS 2000, a significant number of requirements were late, added, or inconsistent. In particular, (1) requirements for forms and reports were not defined until late in the development cycle, (2) bar code and output file specifications were inconsistent, and (3) tasks for map scanning and data capture for projects unrelated to the dress rehearsal were added. While it is not uncommon for requirements to be ill-defined and volatile during the early project phases, this problem persisted on DCS 2000 until the dress rehearsal and continued even during dress rehearsal data capture operations. The fact that the DCS 2000 project had little flexibility to compensate for requirements problems or to manage risks exacerbated the problem: Contract funding levels were not increased to accommodate the additional development that the new and changing requirements demanded, and funds initially intended for testing had to be used for development to satisfy these requirements. Moreover, because the dress rehearsal date was fixed, the schedule could not be extended to provide more time for additional development and testing.

The DCS 2000 contract originally sought to deliver a fully-functioning data capture system at the start of the 1998 Dress Rehearsal. Because many key requirements were unknown at contract award in March 1997, the contract left them undefined. The bureau planned to define those requirements within six months of contract award through a series of internal meetings and to

communicate them to the contractor to enable production of a functional and performance specification. The contractor produced the *DCS 2000 Functional and Performance Specification* on September 12, 1997. This specification document was based on the requirements that were known at that time, which were defined with varying degrees of precision. Many requirements remained unidentified and thus undefined.

For example, requirements for the different types of census forms to be automatically processed were late and grew in number. The number of form types increased from approximately four in initial planning to 30 for the dress rehearsal. While the bureau provided general guidelines on the nature and types of forms that DCS 2000 would have to process early in the development cycle, it did not provide actual forms to the contractor when originally planned. The first seven forms were not provided to the contractor until February 9, 1998, with the remaining 23 forms being provided between February 10, 1998, and April 2, 1998. Access to actual forms is essential in order to program the system to accurately recognize the position, type, and range of the data to be read. However, forms were received from five to seven months late, and with Census Day for the 1998 Dress Rehearsal on April 18, 1998, little time was left to verify system processing and perform any necessary reprogramming of DCS 2000.

Another example of late requirements was related to the status reports. The format and content of DCS 2000 performance and production status reports that the contractor was required to produce were not fully defined until after the start of dress rehearsal. As a result, the contractor had to employ a time-consuming trial-and-error approach to produce reports to meet the bureau's requests.

Additional examples of late requirements involved the Data Capture Audit and Resolution (DCAR) process. DCAR required the contractor to determine and report additional information related to the completeness of data capture and to the population count for each processed questionnaire. The additional data was included in the data files that were electronically sent from Jeffersonville to Census Bureau headquarters for subsequent processing. The specifications provided to the contractor were ambiguous, and considerable effort had to be expended by bureau and contractor personnel to finally determine the form and content of the additional data required to support DCAR.

Examples of inconsistent requirements pertained to how the bar codes would be pre-printed on each form and how the output file of the captured data would be formatted. In the case of bar codes, it was not apparent until returned forms arrived at Jeffersonville and were processed that the printing contractor and the DCS 2000 contractor were given conflicting specifications. This conflict resulted in incorrect processing of forms and required reprogramming of DCS 2000. Also, the initial format specified by the bureau for the output file of the captured data was inconsistent with the data that was available from the returned forms. The format specified data fields that did not exist on the forms that were to be processed, while ignoring some data fields

that did exist. As a result, Census Bureau headquarters systems encountered problems with processing data capture files provided by Jeffersonville because of the unexpected variations in the content and format of the data.

The data capture contractor had proposed a rigorous and well-defined system engineering approach in which specific activities had to be performed and completion criteria met before the project could proceed to the next phase of development, test, or deployment. However, requirements growth and change, which were not accompanied by funding growth or schedule relief, caused the contractor to abandon its planned approach and perform concurrent development, testing, and deployment activities on a short cycle that did not allow sufficient time to consistently apply sound system engineering practices. This problematic approach was adopted in order to deliver a system in time for dress rehearsal, an event whose scheduled start was fixed.

Although the bureau had planned to deliver a fully tested and functioning data capture system to Jeffersonville before the dress rehearsal began, it became necessary to deliver the system in increments, with incremental deliveries continuing after dress rehearsal data capture operations had begun. As a result, the contractor and Census Bureau implemented a two-week software upgrade delivery cycle during dress rehearsal. In all, seven software upgrades and 12 emergency corrections to DCS 2000 were made during dress rehearsal to add capabilities and correct deficiencies. This approach allowed critical processing capabilities to be delivered just in time to support the schedule of operations and to correct problems, but it did not allow the contractor to follow well-defined development and test procedures. We believe that the system engineering shortcuts that the contractor had to take as a result of the requirements issues were a major contributor to the number and severity of problems encountered during dress rehearsal data capture operations.

II. Insufficient Testing Caused Problems During Dress Rehearsal

Errors detected early in the development of a system are generally easier and less costly to fix than those found later in the development cycle. Hence, the bureau had developed a comprehensive and rigorous test program for DCS 2000 that was designed to validate that it functioned properly and to identify and correct problems before dress rehearsal. The contractor's proposed test program was designed to support that of the bureau.

The DCS 2000 contract specified that the data capture system would be developed in four increments, with each increment to include additional capabilities. The contract also specified that the first three increments were to result in interim releases of the system and that each release would be subject to a demonstration (referred to in the contract as level B, C, and D demonstrations) that would thoroughly test its capabilities using census forms provided by the Census Bureau.

The first interim release, which supported the Level B demonstration, was to include basic scanning and forms processing capabilities. The second release, which supported the Level C demonstration, was to include the additional capabilities associated with the check-in and systems administration subsystems, as well as the processing of 12,000 forms to be filled out by the public in a planned 1997 census test. The third release was supposed to include the capabilities to support the dress rehearsal, which was the Level D demonstration. The fourth release will be the data capture system that will be delivered to the bureau's data capture centers for the decennial census.

Although both the bureau and contractor had a well-defined testing program, their planned approach to testing could not be followed because of several factors, including funding problems and the effect of requirements instability as discussed earlier. The Level B and Level C demonstrations had to be reduced in scope and delayed due to fiscal year 1997 funding constraints. The funding shortfall resulted from the bureau's lack of planning for many program requirements, as well as increased contractor overhead rates and unplanned local travel expenses because of delayed occupancy of the Bowie computer center. As noted above, for the Level C demonstration, the bureau had intended to provide the contractor with a test deck of approximately 12,000 forms from a planned 1997 census test. However, the bureau canceled the test because of a lack of funding, and an alternative test deck had to be developed. Thus, instead of the more realistic test deck comprising 12,000 forms filled in by the public, a limited test deck of 2,000 forms filled in by census staff was substituted. Testing with a reasonable quantity of forms filled out by the public helps ensure a level of correctness of many aspects of the system that cannot be achieved by any other means. Besides improved testing of the physical imaging of the forms, the OMR, OCR, quality check, and key-from-image functions would have received a more thorough check-out if the originally-planned test deck had been available.

According to the bureau's plans, the dress rehearsal was to be the first operational test of the system. Operational testing exercises the complete system in an actual operational environment or in an environment as close to the operational environment as possible. The Level B and Level C demonstrations were designed to ensure that dress rehearsal data capture operations would proceed smoothly. However, the limited testing of the system during the previous demonstrations did not uncover many of the problems that should have been identified before dress rehearsal.

Testing was limited not only by the reductions in the size and scope of the demonstrations, but also by the use of a test environment that did not replicate the operational environment. The original DCS 2000 plan had called for the installation of two separate systems, one for development and another for testing. However, because of funding shortfalls, the test system never had all of the equipment designated for a DCS 2000 operational system, as had been planned. For example, the test system had only two scanners instead of the planned three. In

addition, the number of workstations used for performing the automatic processing steps, while not intended to completely mirror an operational suite, was significantly less than planned (AIQA—1 versus 17, OMR—4 versus 12, OCR—5 versus 10).[6] Contractor and bureau management officials did not expect these differences to significantly affect performance. In reality, these differences allowed the very serious and disruptive performance backlogs that occurred during dress rehearsal to go undetected during system testing. Had the full complement of equipment and test materials been available for the Level C demonstration as planned, we believe that most of these problems would have been identified and corrected before the dress rehearsal.

In addition to supporting the Level B, C, and D demonstrations, the contractor had also designed an internal test process that emphasized identifying and eliminating errors systematically as development proceeded in order to lessen propagation of defects to later phases. The contractor's planned approach includes unit testing, which is conducted by the software developers on the basic software components; software integration testing, which tests integrated suites of software components and is also carried out by the software developers; and system integration test, which comprises tests conducted by test engineers in a laboratory environment that is to be configured identically to the operational system. The test process also includes acceptance testing, which will be conducted by the contractor and witnessed by the bureau.

As noted previously, due to growth and change in requirements, the software was not developed or tested as Census planned, and as additional requirements were identified, the contractor had to quickly develop the software, abandoning its systematic test process. More than one-half of the nearly 1,700 trouble reports were discovered at system integration testing or later. We believe that many of the problems encountered during system integration testing would have been discovered and corrected earlier during unit testing or software integration testing if the contractor's original test program had been followed.

In order to ensure more effective testing during the remainder of the DCS 2000 development contract, the contractor and the bureau have proposed or initiated several courses of action. First, the contractor has met with members of all the software development teams and stressed the importance of following the original test plan and ensuring that effective unit and software integration testing is performed during software development. This will not be possible, however, unless the bureau controls requirements changes. In addition, the contractor has proposed that separate systems, identical to those that will be used at the data capture centers for the decennial census, be installed at the Bowie computer center for development and testing. Also, the contractor has proposed that its systems engineers work with bureau developers on the design and development of the decennial forms. This action is intended to help finalize form definitions

[6] Seventeen workstations were allocated to AIQA after problems were encountered during dress rehearsal.

more quickly, leaving more time for testing than was possible for dress rehearsal. Finally, in order to obtain forms earlier for use in testing, the contractor and the bureau's Data Capture System Program Office are exploring the feasibility of using a test deck of forms based on an electronic representation of the form's definition.

CONCLUSION

In view of the above problems associated with DCS 2000 requirements and testing, the Census Bureau and the contractor are to be commended for their efforts in analyzing and correcting the problems that have occurred, spending many more hours on problem resolution than originally anticipated. Efficient review teams, rapid software development, effective analysis and communication of the issues, and a committed staff were all key factors in correcting the problems and making the system work.

Despite the problems encountered during dress rehearsal, the data capture system met all of its processing deadlines. However, given the size, complexity, and performance requirements of the actual decennial census, additional personnel resources, dedication, and hard work will not be able to compensate if serious system deficiencies are encountered. Therefore, requirements and testing for Phase II of DCS 2000 must be carefully managed to ensure that the systems delivered to the data capture centers for the decennial census are capable of handling the expected volume of forms and can deal with unexpected problems that are likely to occur. We believe that strict management of requirements, comprehensive testing, and sufficient funding will allow the problems experienced during dress rehearsal to be solved and DCS 2000 to be capable of performing as needed during the decennial.

At our exit conference, bureau officials told us that they recognize the need to improve the management of requirements not only for data capture but for all aspects of the decennial census. As a result, they have developed the *Census 2000 Management Plan,* which addresses control of requirements for the entire census. The plan calls for careful management of requirements by a steering group comprising the decennial's operational managers. This mechanism will provide the bureau with a much-needed tool for change control by allowing impacts to budget, schedule, and performance to be analyzed and understood before altering decennial census plans.

Improving requirements management is imperative for a successful decennial census, and we believe that the bureau has taken an essential step to do so. Nevertheless, the bureau expects that there will be continued and intense pressure on requirements from parties both within and outside the bureau seeking refinements, additions, and changes to planned operations and procedures. Many proposed changes will have merit individually but may be prohibitive from a cost or schedule perspective or may have unanticipated effects on other operations. The more credibly

and quickly the change control process can determine these impacts, the better chance the bureau has of controlling requirements. We encourage senior management officials at the bureau to continue to support and strengthen the requirements control process and to communicate the process both internally and externally, emphasizing the absolute need to control requirements in order to achieve a successful 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. Strengthen the requirements management process for DCS 2000 to ensure requirements are adequately specified in a timely fashion by:

    a. Identifying all outstanding requirements and defining methods, responsible organizations, and a schedule for completing their definition.

    b. Involving contractor personnel earlier in the process of forms definition and report format determination.

    c. Enforcing cutoff dates for inclusion of requirements.

    d. Ensuring new requirements are added only if sufficient resources are available and they do not adversely impact DCS 2000 development and testing.

    e. Ensuring that the same requirements are defined and communicated in a uniform and consistent manner to all relevant parties.

2. Establish schedules with sufficient time and provide adequate funding to perform complete testing of DCS 2000, including operational testing.

3. Improve DCS 2000 testing for the decennial census by:

    a. Ensuring that the test environment is identical in configuration and workload to the operational systems.

    b. Pursuing testing with simulated forms.

c. Requiring the data capture services contractor to perform sufficient operational testing at each data capture center to assure correct system operation and adequate performance.

The Census Bureau has agreed with the conclusions of this report, and will implement our recommendations as detailed in its response, which appears as Appendix A.

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

OFFICE OF THE DIRECTOR

JAN 26 1999

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

Through: Robert J. Shapiro
Under Secretary for Economic Affairs

From: Kenneth Prewitt
Director

Subject: ***Data Capture System 2000 Requirements and Testing Issues Caused Dress Rehearsal Problems***
Draft Inspection Report No. OSE-10846

This is in response to your memorandum dated November 30, 1998, transmitting the above referenced draft audit report. The Census Bureau agrees with the fundamental conclusions reached by the Office of Inspector General (OIG) in this report and recognizes that improved procedures for requirements management are essential for a successful decennial census. Senior management officials at the Census Bureau will continue to support and strengthen the requirements control process and communicate that process as described in response to the OIG's recommendations which follow:

1. *Strengthen the requirements management process for DCS 2000 to ensure requirements are adequately specified in a timely fashion by:*

    a. *Identifying all outstanding requirements and defining methods, responsible organizations, and a schedule for completing their definition.*

    The Decennial Systems and Contracts Management Office (DSCMO) has established a Requirements Change Request Management Process for Data Capture (memorandum dated 10/30/98). DSCMO is working closely with the Decennial Management Division (DMD) to ensure that the data capture change process feeds directly into the Decennial Directorate's change process.

    The approval, documentation, and communication of new or changing requirements are handled through one or both of two existing groups: Census Operational Managers (COM) and the Issue Resolution/Change Control (IR/CC) Board, both of which have been functioning since July 1998. The COM is comprised of a Chair, the Assistant Division Chiefs (ADCs) from DMD, and lead census managers from all divisions and offices with major responsibility for census operations. Its charter is to provide

requirements to participating divisions and offices, facilitate discussion of the operations designed to meet those requirements, and resolve issues with meeting the requirements. Therefore, changes in requirements are assessed by COM, so that the full impact of any changes on any system, including DCS 2000, is analyzed. If COM determines that a particular change in requirements has acceptable operational impact and little or no budget impact, it may decide to accept a change; the change is then documented and the individual division representatives are responsible for implementation within their division. In the case of proposed changes with major budgetary, timing, and/or other operational impact as determined by COM, the new requirements are further considered by the IR/CC Board.

This Board consists of all ADCs in DMD and the Chair of COM and is chaired by the Chief of DMD. The Board reviews the proposed change in requirements, evaluates options, and assesses budgetary impacts and operational risks. For issues where a solution/change is found acceptable without intractable cost, schedule, or strategic implications, the IR/CC Board will approve the change. If the Board determines that the budgetary and/or operation risks are unacceptable, it may reject the change or may refer the issue to the Assistant to the Associate Director (AAD) for Decennial Census. The AAD is responsible for making the final decision based on consultation with census officials as he deems appropriate, including the Division Chiefs, Associate Director for Decennial Census, other Associate Directors, the Deputy Director, or the Director.

At each stage of consideration, the operational, quality, cost, and other aspects of any change in requirements are documented. The final decision to accept or reject a change, and the rationale for doing so, is documented and widely disseminated.

The Data Capture Programing Office (DCPO), along with Lockheed Martin (LM), has developed an inventory of all outstanding requirements. They have been categorized as (1) to be defined, (2) already defined, and (3) no longer government requirements. DCPO staff are constantly acting on and updating the lists with LM. They have also provided DMD with those requirements that are not included in baseline funding and have provided several cost estimates for some of those efforts.

Currently, schedules for completing these requirements are being defined. If agreements have been reached with LM about certain requirements, schedules associated with their design, development and implementation are included in LM's master program schedule.

b. Involving contractor personnel earlier in the process of forms definition and report format determination.

LM personnel are now active members of a joint data capture/forms design "Forms 2000 Tiger" team. This team meets weekly and ensures that form specifications are known by all relevant parties. Additionally, a reports working group was formed to determine

report format and structure. The group consisted of Headquarters, National Processing Center, DCPO, LM, and TRW staff. This group prepared an all encompassing reports requirements document that was used by the Census Bureau to determine actual DCS 2000 report requirements.

c. *Enforcing cutoff dates for inclusion of requirements*

LM has provided the Census Bureau with cutoff dates for all known requirements. The dates are very aggressive, and the Census Bureau is working diligently to meet these dates. DMD is working with all decennial divisions to ensure that DCS 2000 requirements are defined in time to meet the cutoff dates.

In a memorandum dated November 4, 1998, Deputy Director Bradford R. Huther established a new procedure for controlling requirements and specifications to facilitate the completion of critical systems for Census 2000. Effective November 13, 1998, any request for change in system requirements must be directed to the Assistant to the Associate Director for Decennial Census. Such requests must be accompanied by a formal analysis of cost and other impacts, initiated by the requesting division in coordination with DMD Program Management Staff. If significant resources are involved, the request will first need to be reviewed by the COM and IR/CC Board as outlined above. The decision about whether to accept such a system change will then be made, with the advice of these bodies and staff analysis, by the Assistant to the Associate Director for Decennial Census.

d. *Ensuring new requirements are added only if sufficient resources are available and they do not adversely impact DCS 2000 development and testing.*

The COM and IR/CC processes, in conjunction with the memo described above, will ensure that new requirements are added only if sufficient resources are available and they do not adversely impact DCS 2000 development and testing.

e. *Ensuring that the same requirements are defined and communicated in a uniform and consistent manner to all relevant parties.*

Requirements to DCS 2000 are handled in a consistent manner using the COM-IR/CC Board and the DSCMO Requirements Change Request management Process as previously explained.

At times, requirements are generated both to and from DCS 2000. Consistency and uniformity of communication are maintained. For example, during the Dress Rehearsal coding operations, the Population (POP), Housing and Household Economic Statistics (HHES), and Geography (GEO) Divisions voiced their concern to DSCMO regarding the quality of data that were captured. The coding areas (geography, language, ancestor

heritage, race, industry and occupation, and Spanish origin) of these divisions met with DSCMO regarding their joint concern. They jointly communicated their requirement that the data quality be improved for Census 2000. The solution identified was to implement data dictionaries for Census 2000. POP, HHES, and GEO were requested to assist DSC 2000 by preparing these dictionaries, which will evaluate write-in responses.

These dictionaries will be implemented and tested during the Operational Dry Run (ODR) in Baltimore during the summer of 1999. DMD supports this solution and is working with DSCMO in the oversight of this as well as all other requirements.

2. *Establish schedules with sufficient time and provide adequate funding to perform complete testing of DCS 2000, including operational testing.*

The Census Bureau has implemented a joint testing program with both TRW and LM, and staff are in the process of developing detailed schedules. This joint testing program will avoid duplication of effort and maximize efficiencies between the two contractors. Through ODRs at the Data Capture Centers (DCCs), the Census Bureau will ensure adequate operational testing of DCS 2000.

3. *Improve DCS 2000 testing for the decennial census by:*

   a. *Ensuring that the test environment is identical in configuration and workload to the operational systems.*

The Census Bureau has included a technology upgrade for the Bowie Computer Center in Phase II negotiations.

   b. *Pursuing testing with simulated forms.*

The DCPO is working closely with the Rochester Institute of Technology (RIT), its joint partner, to develop digital test data, and a draft plan is in place. This will lead to the printing of adequate forms for testing. The forms will also ensure that the varied test cases are tested thoroughly. DCPO and RIT will use these to test the system requirements, test system operations (including the human interface with the systems), test cluster daily readiness, and test consistency between scanners, clusters, and sites.

   c. *Requiring data capture services contractor to perform sufficient operational testing at each DCC to assure correct system operation and adequate performance.*

The Census Bureau plans to perform sufficient operational testing at each DCC through the ODRs. The ODRs will assure correct system operations and adequate performance of DCS 2000.