United States General Accounting Office

GAO

Report to the Subcommittee on the Census, Committee on Government Reform, House of Representatives

October 2000

2000 CENSUS

Headquarters Processing System Status and Risks

GAO-01-1

Contents

GAO
Accountability * Integrity * Reliability

United States General Accounting Office
Washington, D.C. 20548

October 17, 2000

The Honorable Dan Miller
Chairman
The Honorable Carolyn B. Maloney
Ranking Minority Member
Subcommittee on the Census
Committee on Government Reform
House of Representatives

As you know, the U.S. Census Bureau is conducting the 2000 decennial census. The accuracy of this census depends in part on the proper functioning of 10 interrelated information systems, one of which is the Bureau's headquarters (HQ) processing system. Given the criticality of this system, you asked us to (1) identify the nature and status of the HQ processing system and (2) assess the quality of the system and the risks facing the Bureau if effective quality controls are not in place.

This report summarizes the information presented at our September 14, 2000, briefing to your staff. A copy of our briefing is included in appendix I. We performed our work from July through September 2000 in accordance with generally accepted government auditing standards.

Results in Brief

The HQ processing system consists of 48 applications, all developed internally by the Bureau, that support a variety of census operations such as updating address files, creating a file of census responses, and preparing data for tabulation and dissemination. The Bureau reported that, as of July 20, 2000, 31 of the 48 applications were complete and supported or are supporting census 2000 operations.

The Bureau lacks effective, mature software and system development processes to control development of its HQ processing system applications. It relies instead on the efforts of individuals to deliver applications on time and within budget—an approach that increases the risk that the applications will not be available when needed and/or perform as intended. As a result, the Bureau lacks adequate assurance that the functions performed by the HQ processing system applications—such as ensuring accurate and complete address files and identifying the correct households for enumerators to contact—are properly executed.

Given the short amount of time remaining before the results of the decennial census will be used, the Bureau will need to take immediate steps to mitigate the near-term risks it faces with the quality of the applications that these process weaknesses may have caused. Longer-term improvements will also be required.

HQ Processing System Nature and Status

The HQ processing system comprises 48 applications that support census 2000, grouped into the following three functional areas. *Address list capture operations* consists of 10 applications that refine, update, and edit the Bureau's address files and database prior to conducting census 2000. *Decennial management controls* includes 22 applications that create the Decennial Response File and control census 2000 operations, such as identifying households for follow-up. *Post-response processing systems* comprises 16 applications that resolve inconsistencies, code handwritten responses, perform edits, add missing data, and prepare data for tabulation and dissemination.

According to the Bureau, as of July 20, 2000, 17 applications were complete and supported census 2000 operations that are complete; 14 applications are complete and support ongoing census 2000 operations; and 17 applications are under development to support future census 2000 operations. All of the applications are developed internally, and are mainframe-based at the Bureau's computer center in Bowie, Maryland, and its National Processing Center in Jeffersonville, Indiana.

HQ Processing System Quality and Risks

The Software Engineering Institute's (SEI) Capability Maturity Model[1] and our test management guide[2] define effective software/system development processes. The risk of poor software/system performance, late delivery, and cost overruns rises when effective processes are not in place.

The Bureau lacks effective processes to control development of its HQ processing system applications in three particularly important process areas—requirements management, risk management, and test management.

[1]Capability Maturity Model[SM] is a service mark of Carnegie Mellon University, and CMM® is registered in the U. S. Patent and Trademark Office.

[2]*Year 2000 Computing Crisis: A Testing Guide* (GAO/AIMD-10.1.21, November 1998).

- *Requirements management* involves establishing and maintaining an agreement between the system/software development organization and the customer (i.e., users) on the requirements for the software. The practices the Bureau uses to develop the HQ processing system partially satisfied each of the four specific criteria for requirements management, including the need for documented, reviewed, validated, baselined, and controlled requirements (i.e., the Bureau sometimes—but not always—implements the requisite management control).
- *Risk management* involves proactively identifying potential problems as opposed to reacting to problems only after they materialize. However, the practices the Bureau uses to develop the HQ processing system software did not satisfy each of the five specific criteria for risk management, including the need for risk management planning; risk identification, documentation, and analysis; and risk mitigation strategy development, tracking, and reporting.
- *Test management* entails the planning and conduct of structured and disciplined system testing to provide reasonable assurance that systems perform as intended. The Bureau's test management practices for the HQ processing system did not satisfy the criteria for documenting testing policies and procedures; did satisfy the criteria for documenting and correcting problems resulting from testing; and partially satisfied the remaining three criteria for planning test activities, documenting test cases, and reporting test results.

Conclusions

The Bureau does not have adequate assurance that the functions performed by the HQ processing applications, such as having accurate and complete address files and identifying the correct households for enumerators to contact, are properly executed. While Bureau management has implemented some practices to promote HQ processing applications' quality, the Bureau does not have effective and mature software and system development processes, such as those specified in SEI's CMM and our test management guide. Instead, Bureau management is counting on the efforts of individuals to deliver quality applications on time and within budget. This approach unnecessarily increases the risk that these applications will not be available when needed and will not perform as intended. In particular, the Bureau's approach to testing does not provide the necessary analytical and documented basis for knowing whether applications will function correctly.

Recommendations for Executive Action

Correcting its many HQ processing system development process weaknesses should not be the Bureau's near-term focus. Rather, given the short amount of time remaining before the decennial census results will be used, the Bureau needs to take immediate steps to understand the near-term risks it faces with the quality (i.e., availability and performance) of HQ processing system applications that these process weaknesses may have caused. Complicating this situation is the fact that at this late juncture, the staff who can quickly and effectively assess and address risks are essentially the same individuals who are fully engaged in developing the applications. Thus, we are making both near-term and longer-term recommendations to the Director of the Census Bureau.

We recommend that the Director of the Census Bureau require the Associate Director for Decennial Census to

- collaborate with the Bureau's chief information officer (CIO) to identify Decennial Systems and Contracts Management Office (DSCMO) staff who, in light of workloads and competing priorities, can be diverted on a short-term basis to work with CIO staff to (1) assess the actual scope and coverage of testing performed to date and planned on each HQ processing system application (vis-à-vis application functional, performance, and interface requirements), (2) assess each application's potential impact on the quality of the decennial census if it does not perform as intended (either because it is unavailable or does not function correctly), and (3) use these assessments of the probability and impact of HQ processing system application problems to prepare a risk profile for each application;
- use the risk profile to establish a plan(s) for thoroughly testing the applications on a priority basis, including the use of documented test conditions and cases that are traceable to requirements, documented test results, and documented resolution of defects; and
- execute the plan(s) and report to the Director on the results.

We further recommend that following the completion of the 2000 decennial census, that the Director require the Associate Director for Decennial Census to correct each of the HQ processing system applications development weaknesses that we identified and require that future development efforts be conducted in accordance with mature and effective management processes.

Agency Comments and Our Evaluation

The Bureau's Director provided written comments on a draft of this report. In the comments, the Bureau agreed that its software and system development procedures do not provide the kind of rigor and discipline advocated in SEI and GAO guidance and cited in this report. The Bureau also agreed that decennial census operations could have benefited from earlier implementation of our recommendations, and it stated that it welcomes the opportunity to work with GAO in enhancing the Bureau's procedures prior to commencement of decennial census 2010 operations.

However, the Bureau disagreed with our recommendations that it needs to take immediate steps to assess and understand the near-term risks it faces with HQ processing system applications supporting decennial census 2000, and to thoroughly test these applications on the basis of the priorities established by this risk assessment. The Bureau's key points supporting its position are discussed below along with our response to each.

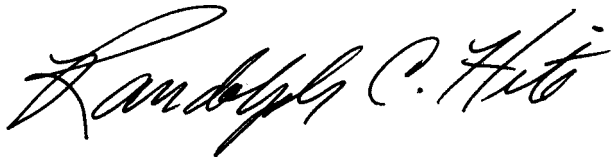
- First, the Bureau stated that it believes that procedures are in place to meet all risk management, system testing, and quality-assurance requirements. We do not agree. Our report cites the results of our analysis of these procedures against SEI and GAO guidance, which describe the kind of rigorous and disciplined processes that mature development organizations employ. These results show that the Bureau lacks effective processes.
- Second, the Bureau stated that since it is impossible to control the nature of incoming census data, effective and essential testing of HQ processing system applications occurs once actual processing of census data is underway. It also stated that to permit assessment and correction of system problems discovered during actual operations, it builds sufficient time into its census operations schedule to accomplish this. In our view, this is not a cost-effective and efficient way to test systems because it postpones efforts to identify the presence of system problems or defects to later in the system development cycle when correcting defects is more difficult and expensive. In contrast, effective test management processes provide for structuring testing activities in an incremental fashion, so that problems and defects with system increments can be discovered earlier, and correcting the problems demands less time and resources. Moreover, given the Bureau's corporate knowledge and experience with prior decennial census operations and the associated incoming data, constructing a meaningful set of test cases that are representative of incoming data, while challenging, is not impossible.

- Third, the Bureau stated that the 31 HQ processing system applications that were completed as of July 30, 2000, have successfully performed the operations for which they were designed. The Bureau further states that these applications' operational performance is the best indication that the Bureau's software and system development procedures are in place and functioning successfully. We do not agree. Without having thoroughly tested the 31 applications' functionality, we do not understand what basis the Bureau is using to make this claim of successful operational performance. The fact that the 31 applications have been operationally executed and have produced an output is not evidence of success (i.e., does not provide adequate assurance that the output is complete and correct). To obtain reasonable assurance, application functions need to be tested using defined test cases with expected outcomes that are designed to provide a predetermined level of function coverage. As we state in the report, the Bureau has not consistently performed such testing.
- Fourth, it stated its commitment to meeting the December 31, 2000, deadline for producing apportionment counts and the April 1, 2001, deadline for producing other census data, and it stated that diverting staff to implementing our near-term recommendation at this late stage would jeopardize its ability to produce census data on time, and would require changes in the existing schedule for dissemination of the census 2000 counts. It also stated that staff and resources have been appropriately committed to control the development, and ensure the accuracy, of HQ processing system applications, and that insufficient funding was provided for early requirements definition and testing against those requirements. We do not question the Bureau's commitment. However, the Bureau's chosen course of action provides for meeting the deadlines using applications that are at risk of not performing as intended. Therefore, we question the Bureau's decision to not apply its staff and resources in a way that mitigates the risk of this occurring. In light of the risk and uncertainty associated with these applications due to the Bureau's lack of effective and mature development processes, we continue to believe, as we recommended, that the prudent step for the Bureau at this juncture is to quickly, and with minimal investment of time and resources, understand which applications are at greatest risk and thus deserving of priority attention, and to take appropriate actions in light of these risk-based priorities to ensure that applications are adequately tested.

The full text of the Bureau's comments is reprinted in appendix II.

We are sending copies of this report to the Honorable Norman Y. Mineta, Secretary of Commerce; the Honorable Kenneth Prewitt, Director of the U.S. Census Bureau; the Honorable Jacob J. Lew, Director, Office of Management and Budget; and other interested parties. Copies will also be made available to others upon request.

Should you have any questions on matters discussed in this report, please contact me at (202) 512-6240 or by e-mail at *hiter.aimd@gao.gov*. Other key contributors to this report were Mark Bird, Garry Durfey, Michael Fruitman, and Richard Hung.



Randolph C. Hite
Director, Information Technology

Briefing on Census' Headquarters Processing System

Bureau of Census Headquarters Processing System

Briefing to the
Subcommittee on the Census Staff,
Committee on Government Reform,
House of Representatives

September 14, 2000

GAO
Accountability * Integrity * Reliability

Overview

- Objectives, Scope, and Methodology
- Background
- Nature and Status of Headquarters (HQ) Processing System Applications
- Quality of Development Processes for HQ Processing System Applications and Associated Risks
- Conclusions and Recommendations

2

Objectives, Scope, and Methodology

Objectives:

- What is the nature and status of Census' HQ processing systems?
- What is the quality of HQ processing systems and what risks does Census face if effective quality controls are not in place for developing these systems?

Scope and Methodology:

- To determine nature and status of systems and effectiveness of controls, we
  - Identified HQ processing system applications (purpose, relationships, development and execution status, environments) developed by Decennial Systems and Contracts Management Office (DSCMO);

3

Objectives, Scope, and Methodology

- Analyzed Census documentation/oral descriptions of key management controls used by DSCMO (requirements management, risk management, test management) to develop these systems as well as test management controls used by user organizations (Decennial Statistical Studies Division and Decennial Management Division) and an independent test organization (DSCMO's Beta Site);
- Compared these management controls to recognized best practices, such as those specified in the Software Engineering Institute's (SEI) Capability Maturity Model (CMM)[1] and GAO's test management guide;[2]

[1]Capability Maturity Model[SM] is a service mark of Carnegie Mellon University, and CMM® is registered in the U.S. Patent and Trademark Office.
[2]Year 2000 Computing Crisis: A Testing Guide (GAO/AIMD-10.1.21, November 1998).

4

Objectives, Scope, and Methodology

  - Examined documentation to see how these controls were implemented on one system application that the Bureau identified as illustrative of current development practices; and
  - Assessed risks, if any, in light of management controls in place.
- Performed work between July and September 2000 in accordance with generally accepted government auditing standards.
- As agreed with Subcommittee staff, deferred reviewing specific system applications to determine their quality until after meeting with the Subcommittee staff to discuss our work to date.

5

Objectives, Scope, and Methodology

- Requested comments on a draft of this briefing from the Commerce Department and the Bureau and incorporated the comments they provided where appropriate.

6

GAO
Accountability * Integrity * Reliability

Background

- An accurate census depends in part on the proper functioning of ten interrelated systems, one of which is HQ Processing. The other nine systems are listed below and described in appendix I.
  - Data Capture System (DCS) 2000
  - Geographic Support System (GSS)
  - Operations Control System (OCS) 2000
  - Pre-Appointment Management System/Automated Decennial Administrative Management System (PAMS/ADAMS)
  - Telephone Questionnaire Assistance and Coverage Edit Follow-Up (TQA/CEFU)

7

GAO
Accountability * Integrity * Reliability

Background

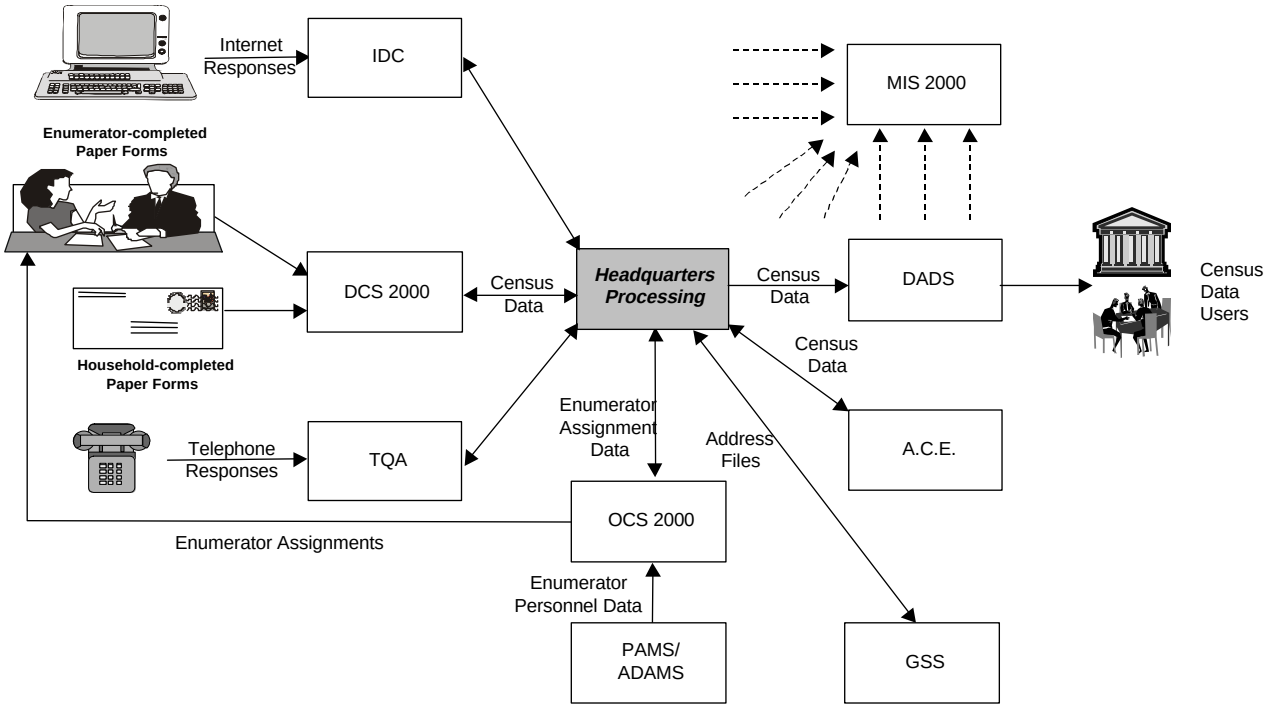
- Internet Data Collection/Internet Questionnaire Assistance (IDC/IQA)
- Accuracy and Coverage Evaluation System (ACE)
- Management Information System (MIS) 2000
- Data Access and Dissemination System (DADS)

8

GAO
Accountability * Integrity * Reliability

Background

Simplified Diagram of Decennial Census Systems



9

Nature and Status of HQ Processing
High-Level Application Functions

- The HQ processing system consists of 48 applications that are grouped into 3 functional areas:
  - Address List Capture Operations (ALCO) -- 10 applications that include a series of operations to refine, update, and edit the bureau's Master Address File (MAF), Decennial Master Address File (DMAF), and Topographically Integrated Geographic Encoding and Referencing (TIGER) database prior to conducting the Census.

10

Nature and Status of HQ Processing
High-Level Application Functions (cont.)

- Decennial Management Controls (DMC) -- 22 applications that collect and process census data from a variety of sources (e.g. DCS 2000, TQA) to create the Decennial Response File (DRF) and control subsequent census 2000 operations, such as Non-Response Follow-Up (NRFU).

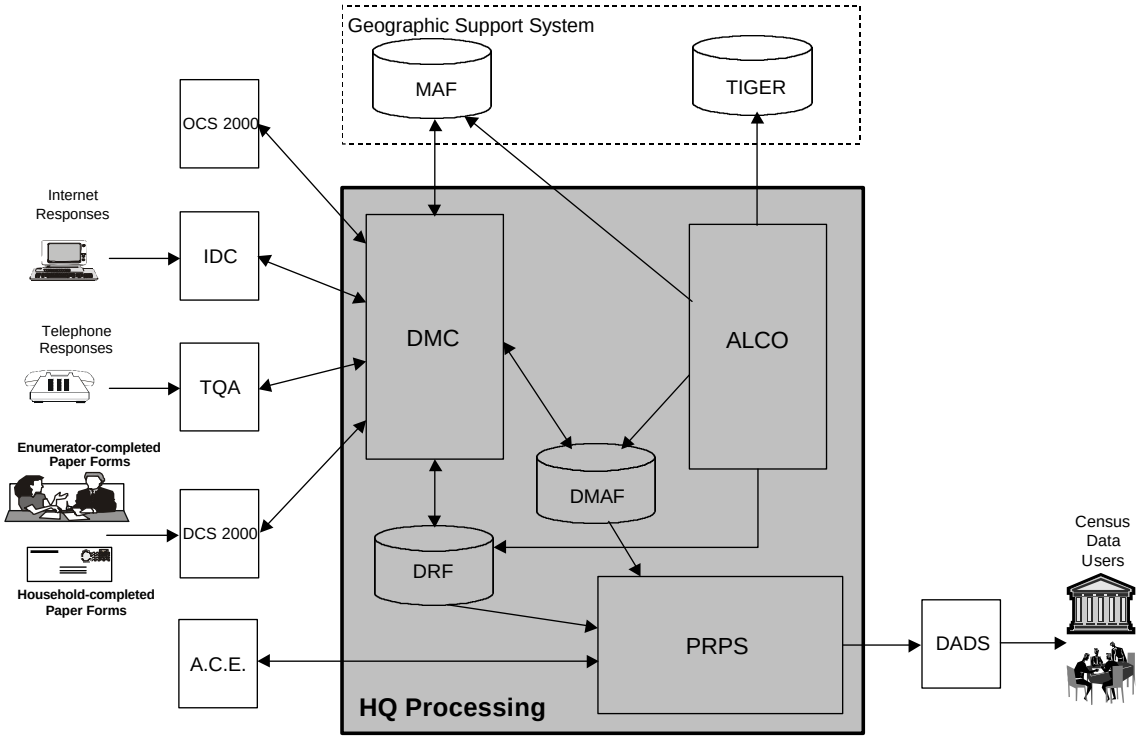
- Post Response Processing Systems (PRPS) -- 16 applications that resolve inconsistencies, code hand-written responses, perform edits, add missing data, and prepare for tabulation and dissemination.

11

GAO
Accountability * Integrity * Reliability

Nature and Status of HQ Processing

Simplified Diagram of HQ Processing Architecture



12

GAO
Accountability * Integrity * Reliability

Nature and Status of HQ Processing
Application Status

- All HQ processing system applications are developed internally by the Bureau.
- According to the Bureau's 7/20/00 status report:

Status	ALCO	DMC	PRPS
Applications are complete and supported Census 2000 operations that are complete.	7	10	0
Applications are complete and support Census 2000 operations that are ongoing.	2	10	2
Applications are under development to support future Census 2000 operations.	1	2	14

13

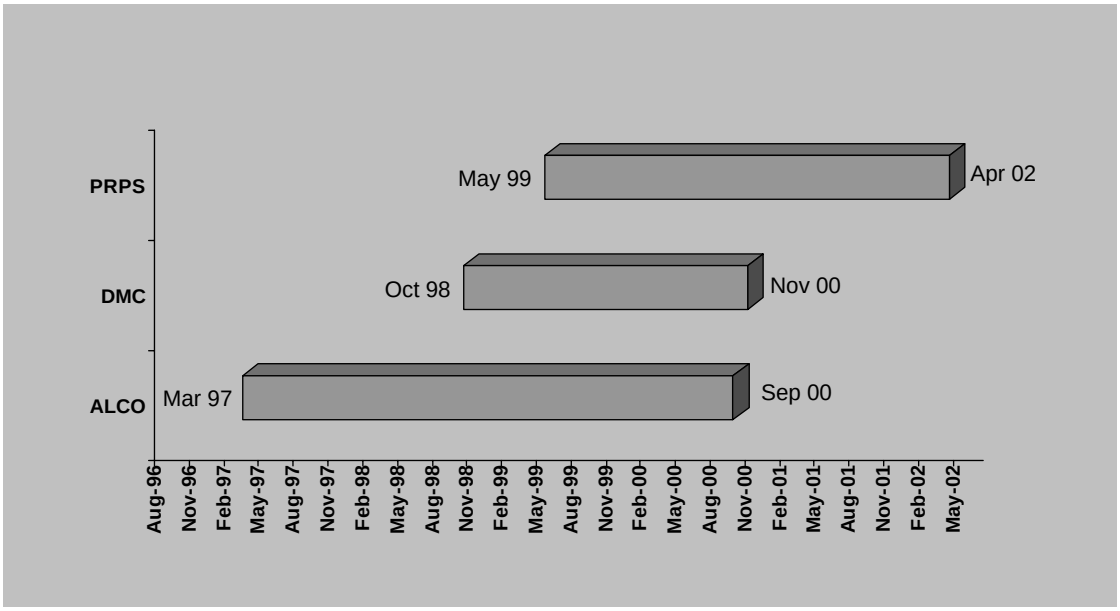
Nature and Status of HQ Processing
Application Characteristics

- Applications are mainframe-based at the Computer Center in Bowie, Maryland and the National Processing Center in Jeffersonville, Indiana.
- Applications are written in the FORTRAN, Pascal, C, and C++ programming languages.
- The Bureau estimates the size of the HQ processing system applications at about 1,250,000 lines of code.
- The HQ processing system applications' status, functions, and relationships are described and depicted in greater detail in appendix II.

14

GAO
Accountability * Integrity * Reliability

Nature and Status of HQ Processing
Application Use Timeline



15

GAO
Accountability * Integrity * Reliability

Quality of Development Processes and Risks
Reported Application Problems

- Earlier this year, the Bureau reported two problems associated with the NRFU HQ processing system application within the DMC functional area. The bureau devised workarounds to minimize the impact of these problems. The two reported problems were:
  - Omission of surnames from address lists used by census workers during the NRFU operation. The Bureau created supplemental surname lists for census workers and identified the cost of this workaround as nominal.
  - Miscalculation of bar code digits for address labels used on census forms as part of the NRFU operation. Because of the miscalculation, DCS 2000 required modification before enumerator forms could be processed.

16

GAO
Accountability * Integrity * Reliability

Quality of Development Processes and Risks
Actions to Address reported Application Problems

- As a result of these problems, the Bureau reports that it has strengthened its HQ processing application development practices, beginning with the Coverage Improvement Follow-Up (CIFU) application as identified in the following table.

Software/System Development Practice	Used for NRFU	Used for CIFU
Requirements walkthrough	No	Yes
External review of output against input files	No	Yes
End user review	No	Yes
Independent external code validation	No	Yes
Internal software quality review	Yes	Yes
Peer code review	Yes	Yes
External review of output	Yes	Yes
Post production evaluation	Yes	Yes
Test deck evaluation	No	No[1]

[1]According to Bureau documentation, time does not permit the use of a formal test deck evaluation for CIFU. However, future software development efforts will include test deck evaluation.

17

Quality of Development Processes and Risks

Effective System Development Processes

- The Clinger-Cohen Act requires establishment of effective IT management processes. Such processes for managing software/system development are defined in various published models and guides, such as SEI's CMM and GAO's testing guide.
- Risk of system/software development product problems (i.e., poor performance, late delivery, cost overruns) is higher when system/software development processes (i.e., management controls) are ad hoc and chaotic.
- Requirements management, risk management, and system/software testing are three process areas important to ensuring that systems/software are delivered on time, within budget, and perform as intended.

18

Quality of Development Processes and Risks (cont.)

Requirements Management

- Requirements management involves establishing and maintaining an agreement between the system/software development organization and the customer (i.e., users) on the requirements for the software project.
- Requirements management includes the analysis of business needs and their translation into documented, validated, and baselined functional, performance, and interface requirements, as well as formally controlling changes to the requirements baselines.
- Effectively managed system/software requirements provide the essential specifications against which systems/software are built and tested. Without them, the chances of development efforts not performing as intended and costing more and taking longer than planned is increased.

19

GAO
Accountability * Integrity * Reliability

Quality of Development Processes and Risks (cont.)

Criteria	Finding	
Policies and procedures for requirements management should be documented.	Partially satisfied	The Bureau has a documented policy and procedures for managing census operational requirements. According to Bureau officials, they have an undocumented practice to derive detailed requirements for use by developers.
Requirements should be documented.	Partially satisfied	According to Bureau officials, their undocumented practice is to document functional, performance, and interface requirements based on user-defined business needs, known system interface definitions, and expected workload estimates. However, they acknowledged that requirements have not been documented for all applications, and that they are sometimes documented after the application has been operationally used. For the application we reviewed, requirements were documented in a memo with e-mail, other memo attachments, and hand written notes.
Requirements should be reviewed and validated.	Partially satisfied	According to Bureau officials, their undocumented practice is for a designated official in the developing organization to review and approve documented requirements. However, they stated that this is not always done. For the application we reviewed, requirements were approved by this official.
Requirements should be baselined and changes to requirements should be controlled.	Partially satisfied	The Bureau has a documented procedure for baselining and controlling census operational requirements changes. According to Bureau officials, they have an undocumented practice to baseline and control detailed requirements for use by developers. For the application we reviewed, the requirements memo had a numbering scheme that could be used to identify the baselined version, however, we found that changes to the detailed requirements baseline were not controlled.

20

Quality of Development Processes and Risks (cont.)

Risk Management

- Risk is the possibility of a problem or loss. Risk management is a proactive approach to avoiding loss or problems before they occur.
- Risk management includes formal identification of potential problems, analysis of their impacts and probability of occurrence, definition of risk mitigation plans, accountability for implementing the plans, and tracking and reporting of progress against the plans.
- Without effective risk management, an organization is forced to react to problems only after they materialize. Such a reactive approach provides little assurance that projects will meet cost, schedule, and performance expectations.

21

GAO
Accountability * Integrity * Reliability

Quality of Development Processes and Risks (cont.)

Criteria	Finding	
Policies and procedures for risk management should be documented.	Not satisfied	The Bureau does not have documented policies and procedures for risk management. Instead, the Bureau relies on informal practices that developers have learned over time and understand.
Risk management activities should be planned.	Not satisfied	According to Bureau officials, risk management plans are not prepared. For the application we reviewed, we found no evidence of a risk management plan.
Risks should be identified and documented.	Not satisfied	According to Bureau officials, risks are not systematically and proactively identified and they are not documented. For the application we reviewed, we found no evidence of identified risks.
Risks should be analyzed and appropriate risk mitigation strategies developed.	Not satisfied	According to Bureau officials, their undocumented practice is to take action to address a risk if one is informally surfaced. However, these actions (the risk mitigation strategy) are not documented. For the application we reviewed, we found no evidence of risk action planning/mitigation strategies.
Implementation of mitigation strategies should be tracked and reported.	Not satisfied	According to Bureau officials, progress in addressing identified risks may be documented as part of regular project reporting and tracking reports and meetings. For the application we reviewed, we found no evidence that implementation of plans for addressing risks were tracked and reported.

22

Quality of Development Processes and Risks (cont.)

Test Management

- Complete and thorough system testing is essential to provide reasonable assurance that systems perform as intended.
- To be effective, testing should be planned and conducted in a structured and disciplined fashion that includes processes to control and direct each incremental level of testing, including testing of individual software units or modules, the integration of these units or modules in creating an application, and the integration of related applications in creating a system.
- Testing processes should include assignment of authority and responsibility for testing, development of test plans and procedures, execution of plans and procedures, documentation of results, and correction of problems.

23

GAO
Accountability * Integrity * Reliability

Quality of Development Processes and Risks (cont.)

Criteria	Finding	
Policies and procedures for unit, integration, and system testing should be documented.	Not satisfied	The Bureau does not have documented policies and procedures for unit, integration, and system testing. Instead, the Bureau relies on practices that the development organization has used over time and understands.
Testing activities should be planned.	Partially satisfied	According to Bureau officials, various levels of testing/checks occur, but not all applications are subjected to all levels of testing. The levels are: (1) programmers perform checks/reviews on their own units or modules of code and the integration of these units/modules; (2) an independent test group tests the entire application; (3) and two user organizations (either the Decennial Management Division or the Decennial Statistical Studies Division) test the entire application. Neither the programmers test activities nor the Decennial Statistical Studies Division tests follow written plans. In contrast, the independent test group and the Decennial Management Division have written test plans. For the application we reviewed, we found no test plans for either the programmer checks/reviews or the Decennial Statistical Studies Division tests. The Decennial Management Division had a test plan, and the independent test group did not perform testing.

24

GAO
Accountability * Integrity * Reliability

Quality of Development Processes and Risks (cont.)

Criteria	Finding	
Test cases should be traceable to requirements, documented, and executed.	Partially satisfied	According to Bureau officials, test cases are documented for the independent test group tests, but since requirements are not always documented, traceability is not always possible. However, test cases are not documented for the checks/reviews performed by each programmer and for the testing performed by one of the user groups. Instead, the Bureau will either leave the scope and nature of programmer checking and review to the discretion of the development organization, or in some cases where time and resources permit, it will employ a technique referred to as "double programming" whereby two programmers write code for the same requirements, compare their results, and address discrepancies. For applications where user testing is performed, this technique is sometimes employed, meaning that the user organization writes its own code (i.e., "triple programming") for the same requirements, compares its results with the results of the developer's executed code, and addresses discrepancies. User testing can, but does not always, include development of test cases. For applications where independent testing is performed, test cases are documented prior to test execution. For the application we reviewed, we found there were no documented programmer or Decennial Statistical Studies Division test cases. The Decennial Management Division had test cases.

25

GAO
Accountability * Integrity * Reliability

Quality of Development Processes and Risks (cont.)

Criteria	Finding	
Problems resulting from testing should be documented and corrected.	Satisfied	According to Bureau officials, problems resulting from testing are documented in e-mail messages and they are corrected. For the application we reviewed, problems resulting from programmer and Decennial Management Division testing were documented and the Decennial Statistical Studies Division reported that there were no problems.
Test results should be reported.	Partially satisfied	According to Bureau officials, programmer and Decennial Statistical Studies Division test results are not documented and reported. Independent test organization and Decennial Management Division test results are documented and reported. For the application we reviewed, a programmer test report was not prepared. A Decennial Management Division test report was prepared and incorporated a Decennial Statistical Studies Division test report

26

Quality of Development Processes and Risks (cont.)

- In addition to the previously described software/system practices, after development is completed the Bureau performs reviews of production data resulting from executing some of the PRPS applications. These reviews are conducted by two user organizations (i.e., the Population Division and the Housing and Household Economic Statistics Division) to check the reasonableness and consistency with other data sources before approving data for subsequent processing for creation of census data products.

27

Conclusions and Recommendations

Conclusions

- The quality (availability and performance) of the Bureau's HQ processing applications is a function of the effectiveness of the processes used to build and test the applications. While process weaknesses do not mean that product problems in fact exist, they are a predictor. Leading software and system authorities have shown a correlation between immature processes and late product delivery and poor product performance. In the Bureau's case, this means that the Bureau does not have adequate assurance that the functions performed by the HQ processing applications, such as having accurate and complete address files and identifying the correct households for enumerators to contact, are properly executed.

28

Conclusions and Recommendations (cont.)

- While Bureau management has implemented some practices to promote HQ processing applications' quality, the Bureau does not have effective and mature software and system development processes, such as those specified in SEI's CMM and GAO's test management guide. Instead, Bureau management is counting on the heroic efforts of individuals to deliver quality applications on time and within budget. This approach unnecessarily increases the risk that these applications will not be available when needed and will not perform as intended. In particular, the Bureau's approach to testing does not provide the necessary analytical and documented basis for knowing whether applications will function correctly.

29

Conclusions and Recommendations (cont.)

- Correcting its many HQ processing system development process weaknesses should not be the Bureau's near-term focus. Rather, given the short amount of time remaining before the decennial census results will be used, the Bureau needs to take immediate steps to understand the near-term risks it faces with the quality (availability and performance) of HQ processing system applications that these process weaknesses may have caused. Complicating this situation is the fact that at this late juncture, the persons that can quickly and effectively assess and address risks are essentially the same individuals that are fully engaged in developing the applications. Accordingly, we are making both near-term and longer-term recommendations to the Director of the Census Bureau.

30

GAO
Accountability * Integrity * Reliability

Conclusions and Recommendations (cont.)

Recommendations

- We recommend that the Director of the Census Bureau require the Associate Director for Decennial Census to:
  - Collaborate with the Bureau's Chief Information Officer (CIO), to identify DSCMO staff who, in light of workloads and competing priorities, can be diverted on a short-term basis to work with CIO staff to:
    - Assess the actual scope and coverage of testing performed to date and planned on each HQ processing system application (vis-à-vis application functional, performance, and interface requirements);

31

GAO
Accountability * Integrity * Reliability

Conclusions and Recommendations (cont.)

- Assess each application's potential impact on the quality of the decennial census if it does not perform as intended (either because it is unavailable or does not function correctly); and
- Use these assessments of the probability and impact of HQ processing system application problems to prepare a risk profile for each application;

- Use the risk profile to establish a plan(s) for thoroughly testing the applications on a priority basis, including the use of documented test conditions and cases that are traceable to requirements, documented test results, and documented resolution of defects; and
- Execute the plan(s) and report to the Director on the results.

32

Conclusions and Recommendations (cont.)

- We further recommend that following the completion of the 2000 decennial census that the Director require the Associate Director for Decennial Census to correct each of the HQ processing system applications development weaknesses that we identified, and require that future development efforts are conducted in accordance with mature and effective management processes.

33

GAO
Accountability * Integrity * Reliability

Appendix I: Description of Decennial Census Systems

System Name	Description
Geographic Support System (GSS)	Provides the basic census address list, maps, and geographic reference files for all census programs, including the 2000 decennial census.
Pre-Appointment Management System/ Automated Decennial Administrative Management System (PAMS/ADAMS)	Supports temporary bureau employee applicant tracking and processing, temporary employee selection records, recruiting reports, personnel and payroll processing, and archiving of historical employment data.
Operations Control System (OCS) 2000	Supports, manages, and controls all field operations for Census 2000.
Management Information System (MIS) 2000	Provides the official source of management information on Census 2000 and will provide information on scheduling, progress to date, cost to date compared to budget, and performance anomalies.
Telephone Questionnaire Assistance (TQA) and Coverage Edit Follow-Up (CEFU)	Uses both automated interactive voice response technology and live operator responses to ensure that the public's needs are addressed (TQA). Flags edit failures on census forms and sends them to a telephone follow-up operation where a generic follow-up interview will be conducted (CEFU).
Internet Data Collection/Internet Questionnaire Assistance (IDC/IQA)	Provides respondents, on a limited basis, the ability to complete their short form questionnaire on-line over the Internet and provides respondents with answers to questions via the Internet.
Accuracy and Coverage Evaluation (ACE)	Supports a follow-up survey for a representative sample of 300,000 housing units across the nation.
Data Capture System (DCS) 2000	Checks in, digitally images, and optically reads data from census forms and converts these data into files that will be transmitted to headquarters for tabulation and analysis.
Data Access and Dissemination System (DADS)	Provides access to census results contained in two primary subsystems, American FactFinder and Data Products Production, through the Internet.

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications

#	Application Name	Description	Census 2000 Development and Operations Complete	Census 2000 Development Complete and Ongoing Operations	Under Development for Future Census 2000 Operations
1	Decennial Master Address File (DMAF) Creation	Controls and tracks census operations. Initially created from the Master Address File (MAF).	✓		
2	Language Form Determination	Responds to requests for foreign language forms.	✓		
3	Form Type Sampling	Identifies the households to receive the long forms.	✓		
4	Surname Determination	Identifies the households for priority processing capability to capture surnames for follow-up operations.	✓		
5	Accuracy and Coverage Evaluation (A.C.E.) Sample Identification	Selects block clusters with housing unit counts for A.C.E. sample determination.	✓		
6	Creation of Address File Tapes	Extracts files from the DMAF for creation of addresses, bar codes, and other control information for production of census forms.	✓		
7	DMAF Updates	Updates the DMAF from MAF refresh files.			✓
8	Decennial Response File (DRF) Processing	Stores every census response from continuous (daily) updates from census operations.		✓	
9	Check-in of Undeliverable As Addressed (UAA) Returns	Captures the census ID and reason code and loads information into the DMAF.	✓		

Note: Development and operational status of applications is as of July 20, 2000.

35

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications (cont.)

#	Application Name	Description	Census 2000 Development and Operations Complete	Census 2000 Development Complete and Ongoing Operations	Under Development for Future Census 2000 Operations
10	Receive Check-in Files of Mailback and Enumerator Returns.	Captures the IDs of mail-back responses and enumerator response forms for housing units (not including GQ) and updates the DMAF.	✓		
11	Receive Data Capture Files of Mailback and Enumerator Forms	Receives data capture files from mail responses and enumerator response forms for housing units (not including GQ) and loads the response data into the DRF.		✓	
12	Receive Check-in of Group Quarters/Special Place (GQ/SP) Responses	Captures the IDs of GQ/SP forms and updates the DMAF.	✓		
13	Receive Check-in of TQA Responses	Receives daily transfers of TQA responses.		✓	
14	DMAF/MAF Interface	Sends responses that do not have an address identifier to Geography for assignment and updates the DMAF or the DRF accordingly.		✓	
15	Decennial Field Interface (DFI)/OCS2000	Controls field collection activities at Regional Census Centers (RCCs) and Local Census Offices (LCOs).		✓	
16	Computer Assisted Telephone Interview (CATI)/Internet Interface	Provides joint check-in/data capture output for census internet responses.		✓	

Appendix II: Description of HQ Processing System Applications (cont.)

#	Application Name	Description	Census 2000 Development and Operations Complete	Census 2000 Development Complete and Ongoing Operations	Under Development for Future Census 2000 Operations
17	DCS 2000/DCSC Interface	Transfers the data capture files and control information from the data capture centers to HQ.		✓	
18	Address Verification/Fraud Detection	Identifies new addresses requiring field verification, such as BCF/TQA responses.			✓
19	Non-Response Follow-up (NRFU) Identification	Identifies NRFU universe based on DMAF data.	✓		
20	Coverage Edit Processing	Selects failed responses, based on coverage edit indicators in the DRF, for coverage edit follow-up.		✓	
21	Large Household Processing	Identifies housing units with more than 6 persons from the DMAF for a continuation follow-up package.		✓	
22	Coverage Improvement Follow-up (CIFU)	Identifies, following NRFU, vacant and deleted housing units, blank mail returns, MAF adds after earlier operations.		✓	

Appendix II: Description of HQ Processing System Applications (cont.)

#	Application Name	Description	Census 2000 Development and Operations Complete	Census 2000 Development Complete and Ongoing Operations	Under Development for Future Census 2000 Operations
23	Multiple Response Processing	Identifies and flags for removal redundant person and housing unit records. The system uses the Primary Selection Algorithm (PSA) to unduplicate the DRF.			✓
24	Create Census Unedited File (CUF)	Merges the DMAF control file with PSA results on the DRF.			✓
25	Edit and Imputation for the 100% Data	Creates the Census Edited File (CEF) by applying a series of line-item edits to the CUF file and imputing for missing data.			✓
26	Service Based Enumeration (SBE) Estimation	Performs statistical estimations based on information collected during the SBE operations.			✓
27	A.C.E. Estimation Preparation	Not applicable for Census 2000.			
28	A.C.E. Estimations	Automates statistical equations to incorporate the results of the A.C.E. program into the census results. Application of dual system estimation.			✓
29	Create Apportionment Counts	Tabulates final unadjusted person counts by state for use in computing the assignment of congressional seats to the states.			✓
30	Estimation Review System	Not applicable for Census 2000.			

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications (cont.)

#	Application Name	Description	Census 2000 Development and Operations Complete	Census 2000 Development Complete and Ongoing Operations	Under Development for Future Census 2000 Operations
31	Disclosure Avoidance Processing	Applies necessary disclosure avoidance techniques to insure privacy of the respondents.			✓
32	Sample Data Weight Processing	Defines and produces the weighted counts from the sample data.			✓
33	Edit and Imputation for the Sample Data	Creates the Sample Census Edited File (SCEF) by applying a series of line-item edits to the Sample Census Unedited File (SCUF) file and imputing for missing data.			✓
34	Compute Variances	Calculates confidence intervals that represent the bounds of the census data/counts.			✓
35	Tabulation Recoding	Applies tabulation recodes to convert the data from raw values to publication values in preparation for loading into DADS/American Factfinder System.			✓
36	Geographic Extract	Codes the write-in responses pertaining to Place of Birth (POB), Place of Work (POW) and Migration (MIG).			✓
37	General Coding	Automates process to code the write-in responses (Race, Relationship, Spanish Origin, Ancestry and Language) to selected questions on the census form and then allows clerical code response entries.		✓	
38	Industry & Occupation (I & O) Coding	Automates process to code the write-in responses to questions pertaining to industry and occupation on the census form and then allows clerical code response entries.			✓

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications (cont.)

#	Application Name	Description	Census 2000 Development and Operations Complete	Census 2000 Development Complete and Ongoing Operations	Under Development for Future Census 2000 Operations
39	Special Place and Group Quarters Data System (SPGQDS)	Collects all SP and GQ information. All GQ data must be represented on the MAF and DMAF.			✓
40	Decennial/Beta Site/NPC Interface	Supports clerical operations conducted in the NPC from the Beta Site.		✓	
41	A.C.E. Housing Unit Computer Matching	Not part of HQ processing system.			
42	A.C.E. Matching Review & Coding System (MaRCS) Housing Unit Clerical Matching	Not part of HQ processing system.			
43	A.C.E. Post Housing Unit Match Processing	Not part of HQ processing system.			
44	A.C.E. Person/ Dual System Estimation (DSE) Computer Matching	Not part of HQ processing system.			
45	A.C.E. MaRCS DSE/Person Matching	Not part of HQ processing system.			
46	A.C.E. Final Housing Unit Matching	Not part of HQ processing system.			

40

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications (cont.)

#	Application Name	Description	Census 2000 Development and Operations Complete	Census 2000 Development Complete and Ongoing Operations	Under Development for Future Census 2000 Operations
47	Address Listing Data Capture	Captures over 23 million addresses in address listing books. Operation performed by NPC.	✓		
48	Address Listing Map Spot Digitizing	Captures map spot indicator data from the address listing books.	✓		
49	Block Canvas (BC) Data Capture	Captures city-style address changes from block canvas listings.	✓		
50	Postal Casing Check (P.C.) Data Capture	Not applicable for Census 2000.			
51	Address List Review (ALR98 and ALR99) and "New Construction" Data Capture	Captures data after local and tribal government review and updates of the MAF.		✓	
52	Independent Listing Book (A.C.E.) Data Capture	Captures over 1.9 million addresses from the A.C.E. address listing project at NPC.	✓		
53	Independent Listing Book (A.C.E.) Map Scanning	Scans the A.C.E. Address Listing maps to provide electronic copies to the housing unit matching clerks.	✓		
54	Update/Leave (U/L) Address Book and Map Check-in	Captures address changes from the U/L address books.	✓		

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications (cont.)

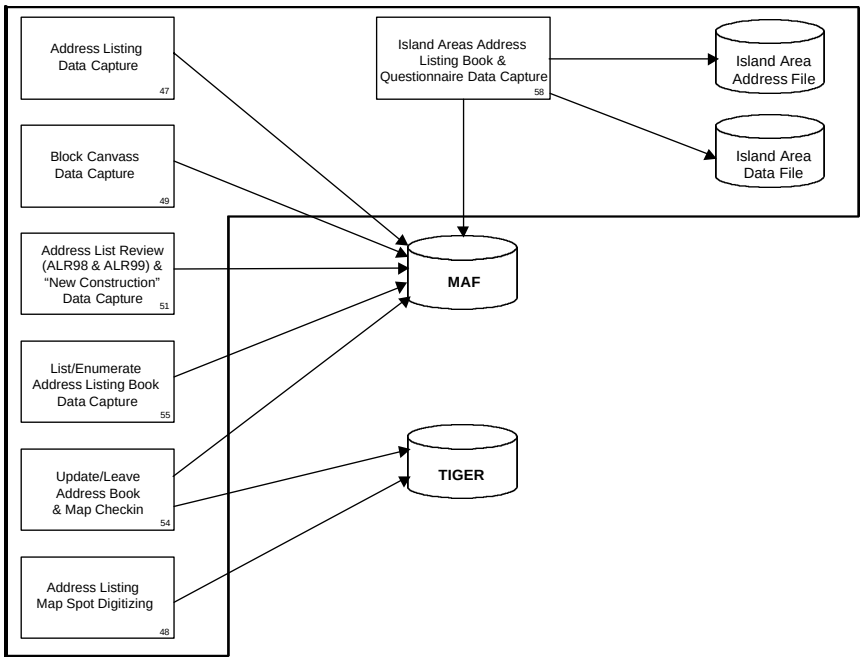
#	Application Name	Description	Census 2000 Development and Operations Complete	Census 2000 Development Complete and Ongoing Operations	Under Development for Future Census 2000 Operations
55	List/Enumerate (L/E) Address Listing Book Data Capture	Captures data from the L/E address books.		✓	
56	NRFU U/L Updates Data Capture	Not applicable for Census 2000.			
57	Quality Improvement Program (QIP) Address Listing Data Capture	Not applicable for Census 2000.			
58	Island Areas Address Listing Book and Questionnaire Data Capture	Captures data from the Island Areas address listing books. Captures data from questionnaires via key-from-paper.			✓
59	Group Quarters Questionnaire Data Capture	Captures GQ data at NPC.	✓		

Note: Application numbers 41, 42, 43, 44, 45, and 46 are not part of the HQ processing system but are included to illustrate the relationships between other HQ processing system applications. Application numbers 27, 30, 50, 56, and 57 are part of the HQ processing system but are not used in the 2000 decennial system. We did not include these 11 applications in determining that the HQ processing system consists of 48 applications.

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications (cont.)

Address List Capture Operations

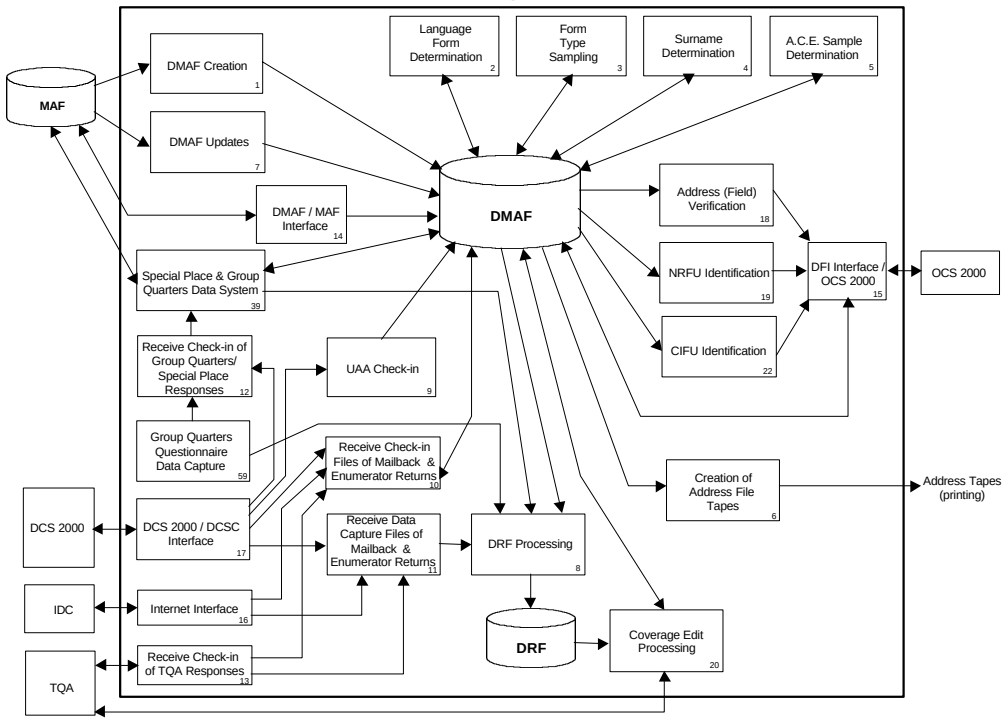


43

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications (cont.)

Decennial Management Controls

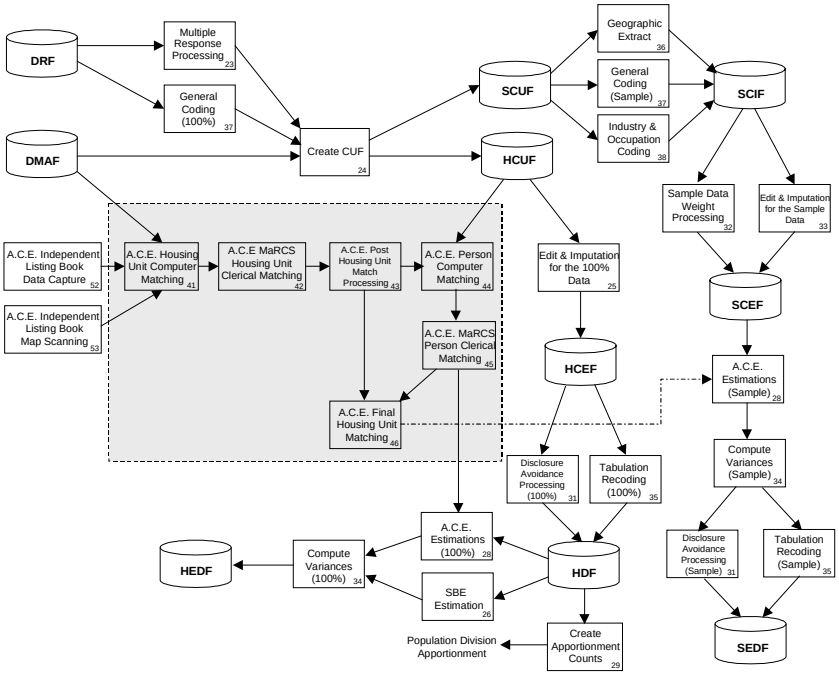


44

GAO
Accountability * Integrity * Reliability

Appendix II: Description of HQ Processing System Applications (cont.)

Post Response Processing System



Comments From the Department of Commerce

UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001
OFFICE OF THE DIRECTOR

OCT 12 2000

Mr. Randolph C. Hite
Associate Director
Governmentwide and Defense Information Systems
United States General Accounting Office
Washington, DC 20548

Dear Mr. Hite:

The U.S. Department of Commerce appreciates the opportunity to comment on the General Accounting Office's (GAO) report entitled *Status of Headquarters Processing System* to the House Subcommittee on the Census.

The U.S. Census Bureau agrees that procedures must be in place to ensure the proper functioning of its interrelated information systems and to assess the quality of each operation. While Census Bureau staff acknowledges that these procedures do not strictly follow the guidelines outlined by the Software Engineering Institute, including the documentation of each procedure, it is important to stress that staff and resources have been appropriately committed to control the development and ensure the accuracy of processing systems in the areas underscored by GAO. Indeed, as noted in the report, 31 out of 48 applications were completed as of July 30, 2000, and each successfully performed the operations for which they were designed. The remaining applications are either under development or currently being implemented, and the Census Bureau is on schedule to meet the federally mandated deadlines for producing Census 2000 data. There is no better indication that the quality-assurance procedures developed by the Census Bureau are in place and functioning successfully.

Census Bureau staff continues to work under strict deadlines to implement each operation necessary for the tabulation of Census 2000 data, as they implement the integrated operations in place to produce the apportionment counts by December 31, 2000, and the data required by Public Law 94-171 by April 1, 2001. Each of the remaining operations is undergoing some or all of the following measures to define requirements and specify the programming system or application in order to internally and externally test their functionality prior to and during implementation:

- Assessing programming requirements – staff examine existing specifications and requirements and then package them for reference during software development.
- Peer code review – the lead software developer for an application shares his/her design and resulting code with peer programmers and supervisors for review, making appropriate adjustments as necessary.

- Internal product review – for most applications, a product is specified in the programming requirements. This product is reviewed by the specifier of the programming requirements for compliance and completeness.

- External division review – live, test products or files are generated and processed in a test mode for review and operational assessments.

- Ongoing production reviews – production products are reviewed to determine anomalous situations that were not discovered prior to production reviews.

Given GAO's appreciation of the strict, federally mandated deadlines in place for the dissemination of Census 2000 data, the report's suggestion that staff should be diverted from their current responsibilities to fulfill the short-term recommendations outlined in the report is particularly troubling. Census Bureau staff believes that systems and procedures are in place to meet all risk management, system testing, and quality-assurance requirements. To divert staff from ongoing Census 2000 operations at this late stage would jeopardize the Census Bureau's ability to produce Census 2000 data on time and require changes in the existing schedule for the dissemination of the Census 2000 counts.

Moreover, past censuses indicate that census data exhibit unique characteristics resulting from questionnaire responses. Given that it is impossible to control this incoming data, Census Bureau staff believes that system testing is effective, and in fact essential, once the processing of census data is underway. Sufficient time is built into the Census 2000 schedule for assessing and revising software applications, along with validating any revisions as necessary, during operations to process data. Again, though, these procedures would be jeopardized if GAO's short-term recommendations were implemented at this time.

We agree that decennial census operations could benefit from all of the GAO's recommendations at earlier stages in the development of headquarters' processing systems. During Census 2000, there was insufficient funding for early operational requirements definition and associated software development and testing based on known system requirements.

Pending the availability of adequate resources early in the decennial census cycle for 2010, the Census Bureau would welcome the opportunity to work with GAO to enhance procedures for system testing and review prior to the implementation of decennial census operations. We look forward to GAO's support for garnering these resources in the years ahead.

Sincerely,

Kenneth Prewitt
Director

Ordering Information

The first copy of each GAO report is free. Additional copies of reports are $2 each. A check or money order should be made out to the Superintendent of Documents. VISA and MasterCard credit cards are accepted, also.

Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:
U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013

***Orders by visiting*:**
Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders by phone:
(202) 512-6000
fax: (202) 512-6061
TDD (202) 512-2537

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

Orders by Internet:
For information on how to access GAO reports on the Internet, send an e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web home page at:

http://www.gao.gov

To Report Fraud, Waste, or Abuse in Federal Programs

Contact one:

- **Web site: http://www.gao.gov/fraudnet/fraudnet.htm**
- **e-mail: fraudnet@gao.gov**
- **1-800-424-5454 (automated answering system)**

United States
General Accounting Office
Washington, D.C. 20548-0001

Official Business
Penalty for Private Use $300

Address Correction Requested

Bulk Rate
Postage & Fees Paid
GAO
Permit No. GI00