BY THE COMPTROLLER GENERAL

RELEASED

Report To The Subcommittee On
Census And Population
Committee On Post Office And
Civil Service, House Of Representatives
OF THE UNITED STATES

The Census Bureau Needs To Plan Now For A More Automated 1990 Decennial Census

Decennial census data annually affect the distribution of billions of dollars and influence business and government decisionmaking. The Census Bureau is taking 2 to 3 years to publish the bulk of the 1980 census data. Users want the data sooner to avoid the potentially harmful effects of using outdated data.

Processing 88 million questionnaires and 3 billion items of data is an enormous task, costing $271 million for the 1980 census. Although tabulation of data is highly automated, the processes for receiving and checking the preliminary data include many manual operations.

Advances in automatic data processing technology now offer the opportunity to automate many of the manual operations and to possibly reduce future census processing time and costs. Six to 7 years are needed to prepare for a census. The Census Bureau's initial planning efforts for the 1990 census are not well organized. Better planning is needed or time will overtake the opportunity for changing the census operations from those of the 1980 census.

The Commerce Department agreed with the report and has taken or plans to take positive actions.

GAO/GGD-83-10
JANUARY 11, 1983

Contents

ABBREVIATIONS

ADP	Automatic Data Processing
FOSDIC	Film Optical Sensing Device for Input to Computers
GAO	General Accounting Office

COMPTROLLER GENERAL OF THE UNITED STATES
WASHINGTON D.C. 20548

B-208738

The Honorable Robert Garcia, Chairman
Subcommittee on Census and Population
Committee on Post Office and Civil
Service

The Honorable James A. Courter
Ranking Minority Member
Subcommittee on Census and Population
Committee on Post Office and Civil
Service
House of Representatives

In response to your July 16, 1981, request, this report identifies the reasons for the time that will be needed by the Bureau of the Census to publish the results of the 1980 census and suggests ways to possibly reduce the time and costs to process the data from the next decennial census. Because of the 6- to 7-year lead time for planning the census, early decisions are needed to select and implement improvements in processing data for the 1990 census.

As arranged with your subcommittee, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days from the date of the report. At that time we will send copies to interested parties and make copies available to others upon request.

We are available to discuss our findings and to provide any further assistance you might need on the matters discussed in the report.

Charles A. Bowsher

Comptroller General
of the United States

COMPTROLLER GENERAL'S
REPORT TO THE SUBCOMMITTEE
ON CENSUS AND POPULATION
COMMITTEE ON POST OFFICE
AND CIVIL SERVICE
HOUSE OF REPRESENTATIVES

THE CENSUS BUREAU NEEDS TO PLAN NOW FOR A MORE AUTOMATED 1990 DECENNIAL CENSUS

D I G E S T

Data from the Decennial Census of Population and Housing, collected by the Census Bureau, annually affect the allocation of billions of Federal and State dollars and influence business decisions of governments and the private sector. Much of the census data collected as of April 1, 1980, was not published for over 2 years. Even after 3 years some data will not have been published.

At the subcommittee's request, GAO reviewed the 1980 census data processing procedures to identify the reasons for the time taken by the Bureau to publish the data and to determine whether changes in the procedures and equipment could reduce the time and cost. GAO did not evaluate the need for all the information collected or the harm to users caused by the time it took to publish the data.

The Bureau collected 88 million questionnaires, containing 3 billion items of data about the Nation's 226.5 million persons and their housing. Processing that data was an enormous task. Because of the large volume of data, the desire for accuracy, and a great reliance on manual procedures, the Bureau expected to take 2-1/2 years to process the data. For a variety of valid reasons, including operational problems and a concern over sufficient funds, the Bureau will need an additional year.

Advances in computer technology now provide the Bureau with the opportunity to automate much of the manual processing and to possibly lower future census processing time and costs. Planning a census requires a 6- to 7-year lead time. Although the Bureau has expressed interest in increased automation,

its initial planning efforts for the 1990 census need better coordination and development, and to provide for the time needed to acquire and test new equipment. The Bureau needs to properly organize its efforts and commit resources or time will overtake the opportunity for increased automation. The Commerce Department agreed with GAO and has taken or plans to take positive actions.

TIME CONSUMING AND COSTLY 1980 CENSUS PROCEDURES

Based on the Bureau's current schedule, the data processing for the 1980 census will require about 3-1/2 years and $271 million. About $106 million has been spent for clerical activities performed by about 55,000 temporary clerks at 409 temporary field offices to generally account for the receipt and completeness of the 88 million questionnaires. About $115 million has been spent for three special processing centers, including 6,300 temporary clerks, to prepare the questionnaires for machine processing and help resolve data problems. Bureau headquarters will spend about $50 million tabulating, reviewing, and publishing the data. (See pp. 6 and 7.)

Once the questionnaire data are transferred to computer tape, the Bureau's computers are used, although detailed manual checks are still made for accuracy. Getting the data to the point where computers can be used, however, is an antiquated process with many expensive manual procedures. (See pp. 6 to 15.)

The 88 million questionnaires were checked-in, edited, and accounted for manually. Logging in and editing the questionnaires cost $48.4 million. Editing required about 37,000 clerks to check each questionnaire for complete and consistent entries. Manually accounting for the questionnaires required 11,000 clerks and cost $9.6 million. Preparing preliminary population and housing counts by hand required 3,400 clerks and cost $6.9 million. (See pp. 8 to 11.)

Handwritten responses to some questions were required because of detailed data

classifications desired by data users. The responses had to be manually coded before the computer could read them. Coding and quality checks took 9 months to complete, required 3,000 clerks, and cost $27.2 million. (See p. 12.)

The time for performing clerical operations contributed to the 1-year slip in the schedule for publishing census data. The major problems were: housing units were underestimated by 2 million; funding was uncertain; data errors needed correction; and boundaries had to be updated. (See p. 15.)

POTENTIAL FOR MORE AUTOMATION

When the Bureau planned the 1980 census, reliable equipment was not available to increase the degree of automation. The Bureau relied mainly on thousands of temporary employees who used detailed manual procedures to ensure the quality of census results.

Since the early 1970s, when planning started for the 1980 census, the automatic data processing (ADP) industry has made major technological advances. For example, microcomputers and related equipment can now do many of the functions, at a fraction of the cost, formerly performed by large-scale computers. They offer much potential to improve the labor intensive census operations, such as questionnaire control and editing. To maximize the use of ADP equipment the Bureau may need to design the census questionnaire so that manual coding is minimized or eliminated. The Bureau is currently using some of this new equipment in other programs to perform the same tasks as manually done in the 1980 census. (See p. 22.)

PLANNING FOR A MORE AUTOMATED CENSUS

The advantages of automating manual census procedures are clear. Planning for the 1990 census is needed now to consider the costs and benefits of a more automated census that will meet the Bureau's needs. The Bureau

took 6 to 7 years to plan the 1980 census. At least that amount of time will be needed for the next census if new ADP equipment is to be considered. Four to 5 years may be needed to have automated equipment available for use after its need is identified. (See p. 26.)

RECOMMENDATION

The Secretary of Commerce should require the Director, Bureau of the Census, to develop a 1990 census plan that includes decision points to evaluate the acquisition, testing, and installation of ADP equipment that are based on past times for planning the 1980 census and acquiring new ADP equipment. The plan should provide for (1) an analysis of alternative data processing systems, (2) the possibility of redesigning the census questionnaire to eliminate or reduce manual coding, (3) an estimate of the time needed to release 1990 census data based on improved data processing, (4) clearly defining responsibilities of Bureau units working on ADP modernization, (5) a budget for implementing the census plan with initial funding requested for fiscal year 1984, and (6) internal periodic progress reports. (See p. 31.)

AGENCY COMMENTS

Commerce agreed with the recommendation. It believes the general recommendation of more automation is good, and it will work toward achieving this goal. Commerce identified activities underway or planned for the 1990 census which are consistent with the objectives of the recommendation, such as developing a master schedule of the major completion dates, testing automated coding of questionnaire items, identifying functions most suitable for automation, and reviewing the Bureau's organization and focusing its fiscal year 1984 budget on the need for automation. Although it had reservations about some of the details, Commerce supports the general thrust of the report. (See pp. 19 and 32.)

CHAPTER 1

INTRODUCTION

The results of the 1980 Decennial Census of Population and Housing, taken as of April 1, 1980, by the Census Bureau, have lasting importance over the next decade for users at all levels of government and the private sector. A primary management objective of the Bureau was to publish the results quickly so that the data collected would be as current as possible and useful to government and other users. After more than two years, much of the data from the 1980 census, exclusive of the population count, has not been published. The Bureau's publication plans show that some data are not scheduled to be published until more than 3 years after being collected. Census information quickly becomes outdated in our ever changing society with its population movement. Using outdated data can lead to improper decisions by government policymakers and businesses in the private sector.

The availability of current census data also is important for programs using census data in their allocation formulas. For example, in the case of a program designed to help meet the special educational needs of children from low income families, the Secretary, Department of Education, concluded in May 1982, that satisfactory 1980 census data would not be available early enough to use such data as the basis of allocations for the 1982-83 school year. The Secretary went on to explain that the Department would continue to use 1970 census data to distribute about $2.4 billion of Federal assistance to State and local education agencies. This decision resulted in a suit filed by 10 States and the parents of three affected children. [1] They contended that the Secretary's decision, if allowed to be carried out, would deprive them unjustly of about $105 million for the school year 1982-83. The Congress provided a legislative remedy for the dispute. [2] It provided $148 million in supplemental funds for fiscal year 1982 so that each county would receive the larger amount of assistance as determined on the basis of either 1970 or 1980 census data, when available. In October 1982, the Secretary distributed the additional $148 million to about 800 counties that benefited from the remedy.

---

[1] Ambach et al v. T.H. Bell, Secretary of Education, Civ. No. 82-1460 (filed May 27, 1982, D.D.C. 1982.) The case became moot and is no longer in litigation in view of Public Law 97-257. (See note 2.)

[2] Public Law 97-257, Chapter IX, 96 Stat. 818, 845 (Supplemental Appropriations Act 1982, Sept. 10, 1982.)

Because of the $1.1 billion cost of the 1980 census, the importance of the data and its timely publication, the House Subcommittee on Census and Population, Committee on Post Office and Civil Service, asked us to examine the reasons for the length of time taken to publish the data and to determine if changes in the Census Bureau's processing procedures and equipment could reduce the time and costs.

AN OVERVIEW OF THE 1980 CENSUS

The decennial population count is required by the Constitution. The census results are extremely important to the Nation during the following decade. The count is used to apportion seats among the States in the House of Representatives, and provide the data to determine congressional districts and potential redistricting of State legislatures. However, the census is much more than just a population count. It responds to a wide range of social and economic information needs in both the public and private sectors. For example, census data help identify where markets are, how many people need schools, jobs, housing, and transportation. Population and other social and economic data, such as personal income data, are used to distribute billions of Federal and State dollars through formula grant programs.

To conduct the census, the Bureau distributed two types of questionnaires, a short form and a long form. [3/] About 81 percent of the Nation's housing units received a short form questionnaire that contained 19 basic population and housing questions. The other 19 percent received a long form questionnaire which contained all of the questions on the short form, as well as 20 additional questions about the housing unit and 26 additional questions for each household member. Questions asked of all households are called 100 percent questions, and those asked of a sample of households are called sample questions.

Taking the census was an enormous task. The Bureau established 409 temporary field offices to collect and check for completeness the 88 million questionnaires containing about 3 billion items of information about the Nation's 226.5 million persons. The Bureau employed about 275,000 temporary employees

---

[3/] For most areas of the country, a sample of one in every six households received the long form. In communities with estimated populations below 2,500, one in every two households received the long form in order to obtain more reliable statistics for revenue sharing and other purposes.

at these field offices at the peak employment period. About 220,000 of the employees collected questionnaires from nonrespondents, and the remaining 55,000 performed clerical work to account for the receipt and completeness of the questionnaires. The Bureau also operated three specialized processing centers to prepare the questionnaires so that the data could be transferred onto computer files (data capture). About 6,300 temporary employees were on the centers' payroll at the peak work level. At Bureau headquarters, permanent staff analyzed and reviewed the data collected and made corrections when necessary so that the data released would be as accurate as possible.

Census accuracy not only entails accounting for the Nation's population and housing, but also specifically identifying where they are. The Bureau collected data in administratively created workload units called enumeration districts. The districts contained an average of about 270 housing units to coincide with the workload levels expected of the census takers (enumerators). The enumeration district boundaries are defined so the data collected can be tabulated by political and statistical jurisdictions. In tabulating census results the Bureau attempts to publish data at low geographic levels to provide the greatest benefits to the data users without disclosing individual data. The tabulation level in many cases is at the block level. Census blocks average about 70 people.

For the 1980 census, the Bureau, as required by law, provided the State population counts to the President within 9 months after census day (April 1, 1980). This period of time was 1 month longer than allowed for the 1970 census. The Bureau also satisfied the other legislated deadline of providing to the States, within 1 year after census day, the population counts for their recognized political areas such as counties and cities. [4]/ A deadline for this data was first established by law for the 1980 census. Aside from those two deadlines, the dates for the release of census results are not specified by statute. The Bureau publishes a schedule of release dates for the information at the various geographic levels to inform the users when to expect the data. The Bureau is about a year behind its precensus schedule and as much as a year behind its publication of comparable data from the 1970 census.

To compensate for the delay in data publication the Bureau released some preliminary data. The Bureau published provisional data on a number of social and economic characteristics

---

[4]/For the 1980 census the Census Bureau is publishing data for about 3 million statistical tabulation units, an increase of 1 million above the 1970 census.

of the population at the State level and some other large geographic areas based on a sample of the data collected. The Bureau provided data that was partially edited for errors to the Office of Revenue Sharing to meet data needs for distributing $4.6 billion in fiscal year 1983 general revenue shares, and to some other Federal agencies for their formula fund allocations.

OBJECTIVES, SCOPE, AND METHODOLOGY

Our objectives as established by the requesting congressional subcommittee were to identify the reasons for the time taken by the Census Bureau to publish the results of the 1980 census, and to determine if changes in procedures and equipment used by the Bureau would reduce the time and cost for processing the census data. Data processing activities, from receipt of the questionnaires at the temporary district offices until the data are published, account for about a fourth of the cost of the census. Our review was directed to

- --determining the Bureau's procedures, time, and cost for processing the 1980 data collected from receipt at a temporary district office until the data are published, with particular emphasis on the sample data;
- --exploring whether alternative equipment, procedures, and questionnaire design would expedite and reduce the cost of data processing; and
- --reviewing the Bureau's data processing planning efforts for the 1990 census.

Our review included an examination of 1980 census planning documents, procedural manuals, and operational reports. We reviewed the format used for the 1980 census questionnaires and observed operations at the Jeffersonville, Indiana, questionnaire processing center. We analyzed Bureau monthly expense statements and field operations cost and progress reports covering the 1980 census; interviewed Bureau personnel including officials responsible for data processing, field operations, and geographic activities for the 1980 census, and management officials responsible for planning the 1990 census. We also interviewed some automatic data processing (ADP) equipment manufacturers to determine potential applications for the capture and processing of census type data.

We did not address the issues of the need for all the information collected from respondents or the harm to users caused by the long time to release the data. Both issues are

important to consider in assessing the cost and benefits of shortening the time for release or publishing census data. Our review was intended to provide insight into problems that occurred in processing the 1980 census data which would be useful in planning the 1990 census, identify some alternatives to overcome those problems, and alert census policymakers that commitments and decisions are needed early in the decade to implement any major changes to data processing procedures.

In a prior report "A $4 Billion Census in 1990? Timely Decisions on Alternatives to 1980 Procedures Can Save Millions" (GGD 82-13, Feb. 22, 1982), we reviewed Bureau procedures for collecting and processing address data, following up on nonrespondents to census questionnaires, and improving the completeness of the census counts. That report generally discusses 1980 census operations for collecting the data and complements the subject of this report. GAO and the Department of Commerce's Office of Inspector General have issued several other reports on various aspects of the 1980 census. GAO believes all these reports should be considered in planning the 1990 census. A list of these reports is included as Appendix II.

Our work was performed in accordance with generally accepted Government auditing standards.

CHAPTER 2

WHY IT WILL TAKE 3-1/2 YEARS TO PROCESS 1980 CENSUS DATA

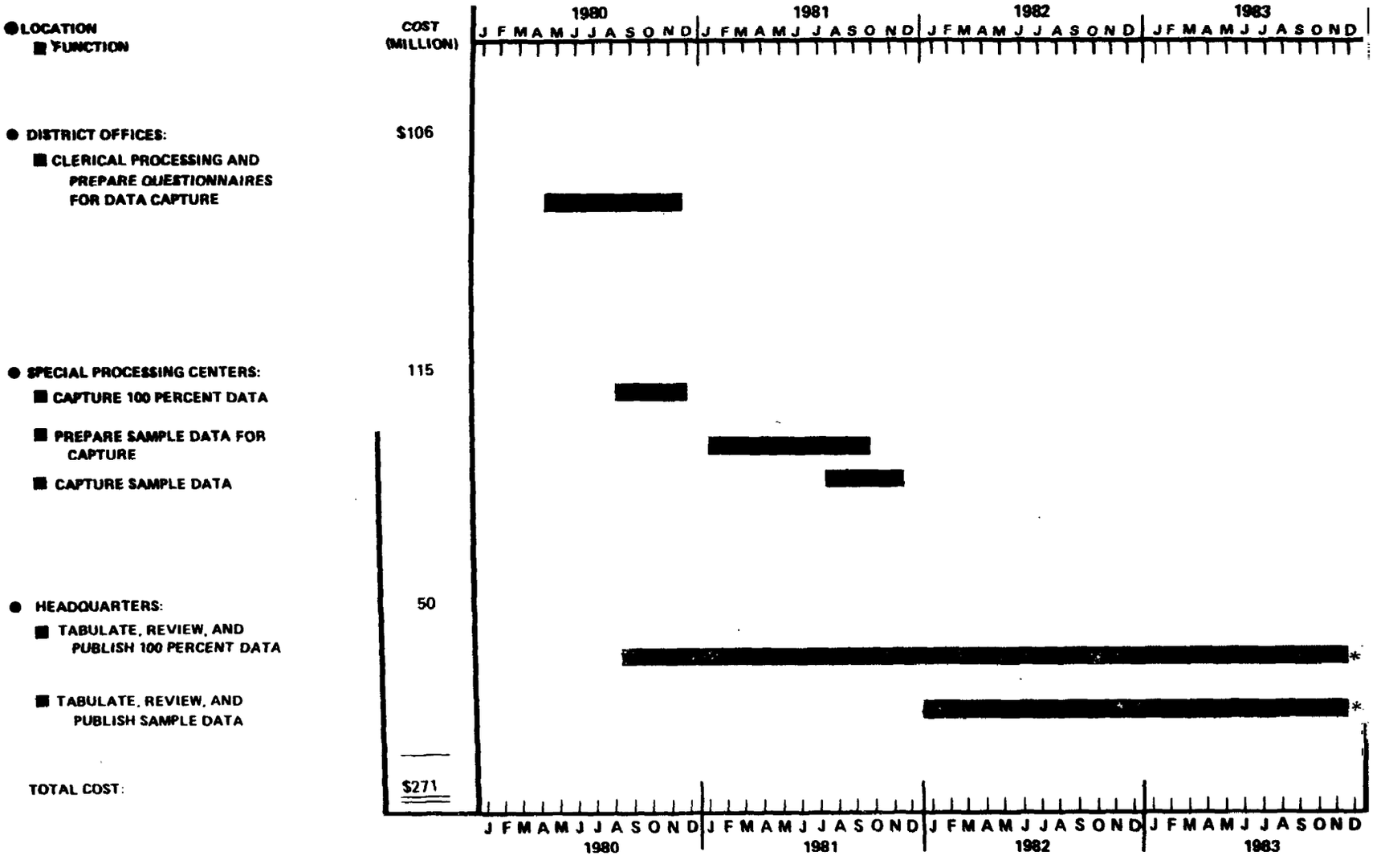
In the Census Bureau's quest to publish complete, accurate, and detailed results, it used a series of detailed procedures to process the data collected from the Nation's 88 million households. Many of the procedures, such as multiple levels of review at different stages of processing, were designed to ensure the quality of the data to be tabulated and published. Because of the high volume of data, the desire for complete accuracy at small geographic areas, and a heavy reliance on manual procedures, data processing required a substantial amount of time and money.

The Bureau recognized in its time estimates that a long period would be needed to complete 1980 census processing. It anticipated that processing would require about 2-1/2 years and be completed by the end of 1982. The Bureau encountered several unexpected problems, including insufficient funds and a high volume of data errors requiring corrections, which contributed to delays. By the time all the reports containing census results are published, the data processing will have taken 3-1/2 years, cost about $271 million, and required the services of thousands of employees.

PROCESSING CENSUS DATA INCLUDES MANY MANUAL PROCEDURES

For processing the 1980 census data, the Bureau's operations included numerous detailed manual and expensive procedures conducted at 409 district offices located throughout the country, three special processing centers in California, Indiana, and Louisiana, and the Bureau's headquarters in Suitland, Maryland. At the district offices, temporary employees performed clerical activities to assist followup operations for obtaining questionnaires from nonrespondents and to prepare questionnaires for automated data capture. At the processing centers, clerks finished preparing questionnaires for data capture; and technicians, using special equipment, captured the data. After the data were captured, Bureau headquarters staff began tabulating, reviewing, and publishing census results. Data processing activities started after April 1, 1980, when questionnaires were received at the district offices, and will conclude with the publication of census reports. The chart on page 7 shows the time and cost of these activities.

TIME AND COST BY LOCATION AND FUNCTION TO PROCESS 1980 CENSUS DATA



*The Bureau estimates that the last reports will be released about the end of 1983

District office data processing--
many temporary employees performed
many clerical tasks

At the district offices thousands of temporary employees performed numerous data processing procedures as well as completing the data collection activities. The primary mission of the district offices was to assure that a questionnaire was returned for every housing unit and that these questionnaires met the Bureau's standards for completeness. Because the questionnaires were returned and processed as received, the processing and data collection procedures were carried out concurrently. In its quest for a complete and accurate census, the Bureau performed a series of expensive processing and data collection procedures. In our report on planning for the 1990 census [1/] we identified possible alternatives to reduce the costly data collection procedures. For the district office data processing activities, the Bureau employed about 55,000 temporary staff at a cost of $106 million, $39 million more than budgeted. Generally, the district office processing required 5-1/2 months to complete.

After the questionnaires were returned, temporary staff manually performed about two dozen procedures associated with handling and processing questionnaires. While there was some overlap, each procedure could generally be classified as (1) receiving, controlling, or batching questionnaires; (2) checking for questionnaire completeness; (3) preparing for or performing a special purpose activity; or (4) recording and compiling counts. These procedures were both time consuming and cumbersome. Initially, clerks manually sorted the returned questionnaires by enumeration district and serial number. Then, the clerks recorded the receipt of the questionnaires along with the population counts in a control list (address register) for each enumeration district by serial number. The address register contained the addresses of all known housing units in each enumeration district that were mailed questionnaires. The check-in operation was used to identify nonresponse and to ensure that a questionnaire had been received from each housing unit listed in the address registers. At the same time, clerks examined responses to question number H-4 which asked the number of living quarters in the building in which the respondent lived. The question was designed as a census coverage check to determine if the respondent in a building with less than 10 units reported more housing units

---

1/"A $4 Billion Census in 1990? Timely Decisions on Alternatives to 1980 Procedures Can Save Millions" (GGD-82-13, Feb. 22, 1982).

than were recorded in the address register. The check-in and H-4 edit operations cost $18.9 million, $11.6 million more than budgeted.

The results of these two operations determined the Bureau's phase one followup workload. The Bureau's policy for the 1980 census was to send census enumerators to each housing unit where a questionnaire was to be returned in the mail and was not received by April 14, 1980. According to the results of the check-in operation, about 23 million questionnaires were not returned. In addition, if the response to the H-4 question showed more housing units than questionnaires mailed, a census enumerator was supposed to visit the building to check whether every housing unit had been enumerated.

A major procedure at the district offices was editing the returned questionnaires. The Bureau spent $29.5 million to determine if the questionnaires had been completed in an acceptable manner. At the peak of the editing workload, the Bureau employed about 37,000 clerks for that operation. During the edit operation, clerks reviewed each questionnaire for incomplete or inconsistent entries using a 35-page instruction manual and a template. The template was a cardboard overlay for the questionnaire that contained additional editing instructions. On the basis of prescribed editing standards, the clerks corrected inconsistent responses and determined whether the quality of the questionnaires was acceptable. Of the 64 million questionnaires that were returned by mail, 13 percent of the short forms and 36 percent of the long forms did not meet the Bureau's standards for completeness.

Similar to most other major field operations, quality control checks were conducted to ensure the accuracy of the edit work performed. In total, the questionnaires may have been processed through six clerical phases of the edit operations: work unit preparation, pre-edit, edit, verification, re-edit, and edit control.

The editing operation and phase two of the followup operation were intended to improve the quality of the data collected. Phase two consisted of several activities including: a field check to verify the status of all addresses reported in phase one as vacant or nonexistent, followup to obtain missing questionnaire information and nonresponses not resolved in phase one, and special procedures intended to improve the population count. The Bureau first attempted to complete the questionnaires that failed edit through telephone followup. If the questionnaires could not be completed by telephone, census enumerators visited the household to complete the questionnaires.

After the census enumerators had finished their personal visits to followup on nonresponse cases, the Bureau reconciled the questionnaires and address registers to ensure the accuracy of its records. The objectives of the reconciliation were to resolve any duplicate questionnaires and to ensure that a questionnaire existed for each entry in the address register, that each entry in the address register conformed to information on the corresponding questionnaire, and that there was an entry in the address register for each questionnaire. Also, the two copies of each address register were compared line by line to ensure that their contents were identical. The registers had been prepared in duplicate so that at all times one copy was available for field use and one for district office use. The operation cost $9.6 million. The Bureau employed 11,000 clerks to perform the reconciliation at the peak workload for this operation.

The Bureau also required several special procedures that were intended to improve the coverage of the population. For example, for selected district offices that covered urban areas thought to have large concentrations of minorities, the Bureau obtained lists of names and addresses from independent sources to use as a check of the accuracy and completeness of census coverage. The lists were obtained from such sources as State files of licensed drivers and the Immigration and Naturalization Service's files of registered aliens. Clerks compared the names and the addresses from the lists to census questionnaires to determine if the persons listed had been enumerated in the census. Unmatched names represented potential misses in census counts and were included as part of the Bureau's phase two follow-up. If these persons had been missed, the census enumerators added them to the census. The Bureau spent about $3 million and employed about 2,700 clerks in the district offices to perform this special purpose processing procedure. In total, the special purpose processing operations cost $12.7 million.

One of the processing procedures performed at the district office was the compilation of population and housing counts. The primary purposes of this procedure were to produce preliminary counts for local officials to check for accuracy and for quality control purposes in subsequent operations at the processing centers. Clerks totaled the population and housing counts recorded in the address registers. The totals were entered at the bottom of each page in the address registers and then posted to tally sheets. The tally sheets were totaled and the counts put on the cover of the address registers. The tally sheets were then sent to the district's regional office for review and the contents were subsequently recorded on a computer file at the Bureau's headquarters. The totals for

long form questionnaires and began when most of the processing of the 100-percent questions was complete. The 100-percent questions were given priority because they contained the data that would be used to provide the President and the States with the population counts required by law.

The 100-percent questions and sample questions were processed in the same manner except handwritten responses to many sample questions had to be converted into machine-readable form before the data could be captured. According to the Bureau, detailed data classifications, such as income, occupation, place of work and ancestry, were needed for a number of sample questions. So the respondents were asked to write in the exact answer instead of choosing from a list of possible responses. Coding handwritten responses into machine-readable form was the most time consuming and expensive procedure at the processing centers. The Bureau employed about 3,000 clerks and spent $27.2 million to code the written responses. The coding operation took about 9 months to complete, primarily from January 1981 to October 1981.

Each long form questionnaire included 32 questions that required coding, 25 questions for each individual listed and 7 for the housing unit. The coding operations were conducted by three separate sections of clerks: one section worked solely on the place-of-work and travel-time-to-work questions, a second on the industry and occupation questions, and the third on all other general items such as income. Using various reference materials, clerks determined the codes for the handwritten responses and then filled in the appropriate circles in the designated code space on the long form questionnaires. The reference material ranged from a single coding guide to stacks of books such as coding guides, telephone directories, and national ZIP code directories.

Because the production atmosphere of a large clerical operation such as coding could result in errors, a quality control operation was instituted to check the clerks' work. At times, this involved the comparison of three separate codings of the same data, and when the quality was unacceptable, correcting the coding errors. Depending on the problems encountered, a group of questionnaires could stay in the coding sections more than 6 weeks.

The Bureau used a 2-step system to transfer the data from the questionnaires onto magnetic computer tape. The questionnaires were grouped by enumeration districts for processing. Initially, the questionnaires were microfilmed by specialized high speed cameras and then the microfilm was scanned by the Bureau's film optical sensing device for input to computers

each enumeration district were reviewed by headquarters staff and if the totals were acceptable on the basis of precensus estimates, the district offices were authorized to close. Over 3,400 clerks were involved in compiling the counts at a cost of $6.9 million.

The final task at the district office was boxing questionnaires by enumeration district and shipping them to the processing centers. While some of the district offices closed in August 1980, the majority of them did not close until mid-September, and the last office closed in December 1980. 2/

Processing centers--more manual handling

The processing centers' basic mission was to transfer the data from the 88 million questionnaires onto computer tapes. To carry out that mission, temporary staff were required to perform numerous manual operations to prepare the data before they could be transferred onto the tapes. The Bureau took about 1-1/2 years and employed about 6,300 temporary staff members at peak levels to complete the operations. Overall, the Bureau reported that it spent $115 million to prepare for and complete the processing center activities.

As the boxes of questionnaires arrived at the processing centers, they were checked in to ensure that all boxes sent by the district offices were received and given machine-readable bar-code labels. The bar-code labels were used to maintain control over the flow and location of questionnaire boxes. During processing, the labels were electronically scanned as the boxes were checked in and out of each processing operation so it would be possible to tell where the boxes were at any time. The questionnaire boxes and address registers were stored in a library until they were checked out to the various processing operations. As an additional control, the questionnaires were always checked back into the library from one major operation before being routed to another. The check-in and control operations cost $6.3 million.

The processing of the census questionnaires occurred in two separate but overlapping phases. The first phase involved the 100-percent questions found on the short form questionnaires and on the initial pages of the long form questionnaires. The second phase involved the sample questions on the

---

2/At that office a fire occurred in late October which required a complete recount of the district.

(FOSDIC) machines which converted responses appearing on microfilmed questionnaires as darkened circles to a computer-readable form on computer tape. Handwritten responses could not be read by FOSDIC. The questionnaires were specially designed so that once microfilmed, any coded marks could be detected by a FOSDIC machine. In addition, microfilming allows the Bureau to satisfy the mandated requirement of preparing authenticated transcripts or copies of all questionnaires.

According to Bureau officials, the data capture system was effective. The Bureau processed the census questionnaires at the system's designed speed. The FOSDIC machines electronically transmitted the 100-percent data to the Bureau's computer in Suitland, Maryland, between August 1980 and late December 1980. The sample data was transmitted between July 1981 and December 1981. The data capture operation, including equipment expenses, cost about $14.3 million.

To ensure that all questionnaires were processed completely and accurately, the Bureau's headquarters compared individual enumeration district population and housing counts obtained from the preliminary counts prepared by the district offices. If a significant difference was found, clerks at the processing centers were directed to review the questionnaires and address registers to resolve the discrepancies. When necessary, modifications were made, and the questionnaires in the rejected enumeration district were remicrofilmed and the microfilm scanned by FOSDIC. Several other checks were made which could have caused an enumeration district to be rejected, including such things as discrepancies in geographic area designations. These comparison and reconciliation operations, performed between August 1980 and December 1981, cost about $4.5 million.

Headquarters processing--
application of statistical
editing and professional review

At the Bureau's headquarters the data were further checked for accuracy and completeness using statistical editing procedures and time-consuming professional review. The Bureau estimates that these additional checks, including publication costs, will total about $50 million.

Headquarters data processing started considerably later than census day. Although Bureau headquarters began receiving 100-percent data in August 1980, processing generally did not start until January 1981 when that data was completely captured and accepted. Processing of the sample data did not begin until 1 year later. Tabulations for a State could not be prepared until all the State's data were processed.

The Bureau, using ADP equipment, processed the data for each person and housing unit through various statistical editing and weighting routines on the basis of very detailed specification programs. A series of computer editing procedures were used to: ensure that the data recorded from a questionnaire reflected actual responses and not just stray marks, eliminate inconsistent data, and provide data missing on the questionnaires. The latter editing procedure was used to make the statistics produced more accurate in their description of the population and housing, and more useful than if "not reported" categories were added to each tabulation. After editing, the sample data were weighted to produce estimates of the figures that would have resulted if all the nation's households had responded to the sample questions. When these processes were completed, edited data about all individuals and housing units and their geographic location were stored on computer tapes.

To ensure that the edit specification programs produced reasonable results, the Bureau's professional staff made time-consuming reviews of the data summaries prepared from the computer tapes. The reviews by as many as 27 subject matter specialists required from 2 hours to 3 weeks. The review time varied depending on the number of problems encountered, the time allowed by management for the reviews, and the geographic level of the tabulated data. The subject matter specialists reviewed the data for three test States in greater detail than the data for the other States. The Bureau believed problems associated with programming errors would probably be identified and corrected during the review of the data for the test States.

Subject matter specialists reviewed the data for consistency and reasonableness. This involved comparing the data to precensus estimates and other information such as results from other Bureau surveys. If the reviewer found inconsistent and unreasonable data, research was done to determine if a problem existed. After the cause of the problem was determined, the Bureau made the necessary corrections. In some instances the Bureau had to change the edit programs and re-run the data through the revised programs. For example, one problem noticed was that the wage income in a particular geographic area was low. After researching the problem, the Bureau found that some respondents, who either failed to answer the wage income question or reported they did not have any wage income, responded to another question that they had worked. Because the respondents had worked, the Bureau assumed that the respondents should have reported wage income. The Bureau changed the edit specification programs to cover this type of situation. Instead of counting these respondents as having zero wage income,

--Correcting population and housing count errors delayed completion of headquarters' processing operations.

--Redefining geographical areas interrupted processing operations.

Low production and greater workload delayed the district office processing operation

According to Bureau officials, lower than expected production and greater workload contributed to the delay in completing the district office processing operation. Clerical production in checking-in and editing questionnaires was lower than the Bureau had planned. In addition, the workloads at the district offices were greater than planned because the Bureau underestimated the number of housing units to be counted by about 2 million.

Failure to complete on time the initial processing operation, questionnaire check-in, affected subsequent district office processing operations. The questionnaires which were processed as received were sent to the next processing operation after they were checked in. Because the check-in was behind schedule, other operations, such as followup for nonrespondents and editing, did not receive the full planned workload of questionnaires on schedule. This slippage was further aggravated by lower-than-expected editing production. Check-in and editing of questionnaires at the district offices cost about $48.4 million, which was $15.1 million over budget. The Bureau also relaxed the editing tolerances to reduce the workload.

Overall, because of low production, greater workload and other operational problems the district offices, originally planned to be closed in about 4 months, generally remained open for 5-1/2 months. The failure to close the district offices on time affected subsequent operations at the processing centers.

Uncertainty of sufficient funding slowed processing center production and reduced the quality of some census results

Because of the uncertainty of sufficient funding for 1980 census activities, the Bureau reduced the coding staff and the number of narrative responses to be coded at the processing centers. The first action contributed to delays in the completion of sample coding operations and the second reduced the quality of census results for the sample questions affected.

a wage income was allocated on the basis of the number of weeks worked and the respondent's occupation.

The Bureau completed its editing and reviews of the individual record files for the 100-percent data and the sample data in January 1982 and June 1982, respectively. Similar reviews as discussed above are now taking place for the final tabulations by a variety of geographic areas such as blocks, towns, and counties.

The last phase of the data processing is the publication of data in two forms, computer tapes, and printed reports. The major portion of the 1980 census results are provided in a series of five summary tape files for data users with access to computers. The summary tape files, which are released on a State by State basis, provide data with much greater subject and geographic detail than was feasible or desirable in printed reports. The printed reports contain only a small portion of the tabulations but serve a broader audience of data users.

The summary tape files are publicly available before the release of printed reports. The first two summary tape files contain population and housing data collected on a 100-percent basis. The other three contain data collected on a sample basis, and generally include 100-percent items for purposes of cross-classification. The first summary tape file was released in late August 1981. The Bureau originally planned to complete the publication of the reports by September 1982. However, according to the latest schedule of tentative release dates, the Bureau will not complete publication until the end of 1983.

The Bureau will spend about $13.9 million for publishing census data. This includes costs for computer programming, review of data tables, and printing.

PROBLEMS DELAY DATA PROCESSING OPERATIONS 1 YEAR

Data processing operations, originally planned to be completed in 2-1/2 years, are currently scheduled to take 3-1/2 years. Problems contributing to the delay included:

--Low production coupled with greater workloads extended the data processing operation at the district offices. The late closing of district offices created a ripple effect on further processing operations.

--Uncertainty of sufficient funding slowed processing center production and reduced the quality of census results.

To cut costs, the Bureau reduced the coding staff by approximately one half in March 1981. The Bureau operated at this reduced staffing level for about 2 to 3 months. In June 1981 when the Bureau received a supplemental appropriation, it stepped up the coding operation by increasing staff. However, many of the coding clerks who had been released were no longer available for rehire. The Bureau had to again recruit and train persons to perform the coding operations. Overall, the Bureau estimated that the coding operation was extended about 3 months by the temporary reduction in staff. According to Bureau officials, a lack of space precluded the Bureau from hiring more coding staff than it originally employed to make up some of the lost time.

The Bureau also reduced the coding workload to save money. The Bureau decided to code the responses to the place-of-work and travel-time-to-work questions on only half the questionnaires. The Bureau chose to use this smaller sample for these questions because the information was considered less critical than the data from the other two categories of questions that required coding. For example, the information from these questions is not used in formulas to distribute funds. According to the Bureau's former Associate Director for Demographic Fields, the use of a smaller sample for these questions will reduce reliability but will still permit the presentation of useful place of work data for small geographic areas. The Bureau did not revise the sample size for these questions after it received a supplemental appropriation because increasing the sample size would have created difficult logistical and administrative problems.

Correcting population and housing count errors delays completion of headquarters' processing operations

The Bureau's efforts to correct population and housing counts affected by geographic or processing errors contributed to delays in completing the headquarters' processing operations. As required by law, the Bureau provided the President with State population counts by January 1, 1981, and each State with the geographically detailed population counts it needed for redistricting purposes by April 1, 1981. However, these counts included a number of errors. Some of the errors were discovered by the Bureau while others were brought to the Bureau's attention by government officials who received the April 1, 1981, population counts. In May 1981, the Bureau decided to insert as many corrections as possible in the computer tapes containing the 100-percent data rather than use errata sheets in printed data reports as it did in 1970. The corrections were made between June 1981 and mid-September 1981.

Because the Bureau believed that a significant number of errors may have been made, it decided to identify and correct as many of the errors as possible before any further 100-percent data processing. Processing of the 100-percent data was delayed for about 3-1/2 months until the modification process was completed. The modification process involved conducting research to determine if the suspected errors were actually errors, instituting changes that would correct the errors and possibly re-microfilming the questionnaires affected by the changes. The Bureau changed 40 State population counts with a net increase in the total national population of about 41,000. Although the change in the national total was only 41,000, numerous changes were required at the county and other political subdivision levels. For example, the changes affected the data reported for 397 counties, or about 13 percent of all counties.

Redefining geographical areas interrupted processing operations

The necessity of redefining geographic areas interrupted the flow of data processing activities and caused delays. Although the Bureau publishes census data corresponding to geographic boundaries in effect on January 1, 1980, the slow mapmaking process required the Bureau to prepare some maps reflecting January 1, 1978, or January 1, 1979, boundaries. The Bureau used the maps to define enumeration districts for the census. To maintain flexibility in aggregating collected data into tabulation data, the Bureau designed enumeration districts so that they did not extend into multiple statistical or political areas. Subsequent to the preparation of the maps, local communities changed their boundaries through activities such as annexations. These boundary changes and errors discovered in review of census data required the Bureau to redefine many enumeration districts so that the census data could be published according to the actual boundaries that existed on January 1, 1980. As a result the Bureau made about 37,000 changes, mainly because of boundary changes, to the enumeration districts.

Changing enumeration districts involved numerous time-consuming activities at the district offices, the processing centers, and Bureau headquarters. District office clerks redrew the boundaries on the maps, divided address registers so that they would coincide with the new enumeration districts, changed the enumeration district numbers on the questionnaires and the storage boxes, and separated the questionnaires into new boxes. At closing, the district offices forwarded the boxes of questionnaires to the processing centers along with a list of the enumeration districts for which changes had not been completed. The processing centers could not directly

check-in some of the boxes of questionnaires, because the enumeration district numbers did not agree with the numbers on a control list that reflected the revised districts. Processing center clerks diverted these boxes from the normal check-in flow. These boxes could not be checked-in until the necessary changes were made including correcting the numbers on the questionnaire boxes.

Before Bureau headquarters accepted the data from the processing centers, it checked the data to ensure that they were assigned to a valid enumeration district. Headquarters personnel notified the processing centers of all invalid enumeration districts. To correct these situations, the processing center clerks changed the enumeration district numbers on the questionnaires and the boxes and reprocessed the questionnaires. Also, headquarters' personnel changed some enumeration districts after the data were accepted. The Bureau created a separate computer file to change these enumeration districts.

CONCLUSIONS

Processing the large volume of data collected in the 1980 census included many manual procedures. The processing took a long time to complete, was expensive, and subject to the human error factor that can be expected in a large operation with numerous detailed manual procedures. If problems or delays had not occurred, the Bureau still would have needed 2-1/2 years to complete publication of census data. That did not happen. Problems occurred that delayed processing about 1 year. If the Bureau completes the release of the data reports according to its latest schedule, processing the data collected in the 1980 census will require about 3-1/2 years.

To reduce the time and cost of processing information for the next census, the Bureau will need to address the laborious manual handling of questionnaires. The following chapter discusses opportunities to increase the use of automated equipment.

AGENCY COMMENTS AND OUR EVALUATION

The Department of Commerce made several comments to clarify or amplify the importance of accurate census data, the definition of data processing we used, and reasons for the time it will take to process the 1980 census data. (See app. III.)

Commerce commented that accurate and complete census results are needed at low geographic levels to ensure the appropriateness of political representation and to determine eligibility for Federal programs. Commerce said that the importance of this accuracy has not been properly emphasized in

the report, with the implication that the various checks made during the census to achieve the desired accuracy might be unnecessary.

We agree with Commerce that completeness and accuracy of census results is needed. Our report does not intend to question the need for checks and controls to achieve the desired accuracy. The report is intended to point out the large-scale manual performance of many of the quality assurance steps that possibly could be automated to achieve more efficient data collection and processing and more timely reporting of census data.

Commerce believes that the report does not make the proper distinction between collection of data and the processing of data for the purpose of tabulation. Commerce also contends that our report defines processing as virtually all activities conducted between Census Day and release of the data products. Commerce points to confusion in the report about processing by citing two references. Commerce quoted our statements that " . . . data processing activities started after April 1 when questionnaires were received at the district office, and conclude with the publication of the census results." (p. 6); and " . . . processing generally did not start until January 1981 . . . " (p. 13).

The objectives, scope, and methodology section of our report beginning on page 4 defines "data processing" to generally include each step the Bureau performed from the time a questionnaire was received at a district office until the data are published. It is not relevant to the message of the report whether a step in that process is called collection or processing. In reference to the comment that our report on pages 6 and 13 contains conflicting dates for the start of data processing, it should be noted that the dates refer to the beginning of data processing at district offices and headquarters.

Commerce commented that the report did not acknowledge the scope of the activities included in the processing system. Commerce cited as an example the correction of geographic boundaries (discussed on p. 18) which it stated was planned as part of the processing system. The section of the report referred to is intended to describe problems that delayed the 1980 census data processing operations. Whether the correction of geographic boundaries was planned for or not, the correction did hinder the processing operations.

Commerce commented that although our report acknowledges that the Bureau encountered some difficulties and resultant delays due to staff turnover, larger than expected workload, and litigation, our report may not have accurately portrayed

the effect of these difficulties. Commerce further explained the problem caused by litigation, which is not discussed in our report. Although litigation was a nagging problem for the Bureau, information was not available to quantify its effect on the processing of census data.

CHAPTER 3

PLANS ARE NEEDED NOW TO IMPROVE THE PROCESSING OF CENSUS DATA

ADP equipment that could reduce the time and costs of conducting the 1990 census is available. The Census Bureau and the Department of Commerce recognize that data processing technology must be examined as a way of improving on the conduct of the 1990 census. In response to our report "A $4 Billion Census in 1990? Timely Decisions on Alternatives Can Save Millions" (GGD 82-13, Feb. 22, 1982), the Commerce Department said that it plans to examine the possibility of much greater automation of some field office operations. The Bureau needs 6 to 7 years to prepare for a decennial census. If the Bureau is to make major changes in its data processing operations for the 1990 census, it needs to make early commitments and develop a plan that includes the time for acquiring and testing new ADP equipment. Early Bureau planning for the 1990 census did not do this.

AVAILABLE ADP EQUIPMENT HAS POTENTIAL FOR DECENNIAL CENSUS APPLICATION

ADP equipment that could address the 1980 processing problems by reducing the great emphasis on manual data handling is now available. Great strides have been made in ADP technology in the past decade. The Bureau has recognized the benefit of such equipment by using it in nondecennial census applications. The use of ADP equipment in decennial census field offices appears most advantageous. If the Bureau could automate those offices, data could be captured a year earlier than for the last census and most of the manual handling of questionnaires that occurred could be reduced. Cost reductions are also possible. To maximize the benefits of automation for the 1990 census, the questionnaire used to collect data must be redesigned to be compatible with available equipment.

Advancements in technology

Since the early 1970s, when planning started for the 1980 census, there have been tremendous advances in computer technology. Computer circuitry has been miniaturized to the extent that today's computers have more capacity than previously thought possible. Today's small size computers (microcomputers) can dependably store and handle almost as much data as

some large computers of the early 1970s. Existing microcomputers do not require a controlled environment. Generally, they can operate in any office environment. The price of computers has been significantly lowered from hundreds of thousands of dollars to thousands of dollars and even less. Today's microcomputers have a full range of peripheral equipment which was previously limited to large scale computers.

Optical scanners are a type of peripheral equipment available for use with microcomputers. As will be discussed later they can play an important role in future census processing. Optical scanners are computer assisted devices that use light for examining patterns such as printed marks or characters on a page. For census purposes these devices are used to transfer data from questionnaires onto computer files (data capture). In addition to the standard mark and character scanners, there are specialized devices. Bar-code reading devices, which can identify data imprinted on machine readable labels like ones on grocery products, are a type of optical scanning equipment. FOSDIC, the equipment designed for the Census Bureau and used in the past three decennial censuses, is another specialized type of optical scanning equipment.

As more sophisticated equipment has become available, the Bureau has recognized that it is feasible to employ ADP systems to perform tasks similar to the manual procedures carried out in the 1980 census district offices. For the 1982 economic census, the Bureau has acquired equipment to automate most of the processing of about 3.5 million questionnaires. For that census, the Bureau plans to use its computer file of all known business establishments as an address register control list for the census. As questionnaires are received from respondents, they will be processed through a bar-code reader/sorter to record the return of the questionnaire in the control file, and batch the questionnaires by type of business. The Bureau plans to program a computer to identify nonrespondents, prepare a series of addressed reminder cards and letters, and edit questionnaires for incomplete or incorrect responses. Payroll and employment data missing from questionnaires returned by smaller firms will automatically be inserted based on data available from other sources. All other failed-edit questionnaires will be identified by the computer for human followup.

For the Puerto Rican portion of the economic census, the Bureau plans to use microcomputers and peripheral equipment to do much of the processing in the field office. The address register, inventory control of questionnaires, and edit routines will be computerized.

As part of its administrative activities the Bureau is using an optical scanner for tabulating the time personnel work

on various assignments. As timesheets with handwritten entries are read into a small computer, characters that the scanner cannot read are visually displayed. For these problem characters an operator keys them into the computer file. The small computer tabulates the data, and the results are automatically printed.

Advantages of automating decennial census procedures

Decentralized automated data processing in the 1990 census has the potential to reduce the manual handling of questionnaires, and the time and cost required to process data in the 1980 census. Dependable optical scanners, microcomputers, and other ADP devices now available make it possible for the Bureau to capture data contained on questionnaires as they are received from respondents, and to subsequently automate much of the costly and labor intensive operations of the 1980 district offices. To achieve maximum benefit from automation, the census questionnaire must be redesigned to be compatible with commercially available scanners.

Early data capture is the key to improved census data processing. If the Bureau determines that a decentralized collection process will be used for the next census to facilitate followup of nonresponse and incomplete response, the potential exists for collected data to be captured and processed in the field offices before they are closed. Tabulations and reviews of the collected data at headquarters could commence as data are captured by field offices. With more timely capture and review of data, the results seemingly could be published earlier than for the 1980 census. Also, specialized processing centers possibly could be eliminated. For the 1980 census the Bureau spent $115 million for processing center operations and did not capture sample data until 12 months after the last district office closed. The use of optical scanners in field offices, when coupled with mechanical sorters, could allow for automation of the $19 million operation for checking-in the 1980 census questionnaires.

The optical scanning equipment used to capture data for the last census is not suited for use in a decentralized data capture system. FOSDIC is efficient and effective equipment with a proven history of success. There are, however, drawbacks for its use in field offices and therefore the Bureau decided against decentralized processing for the 1980 census. Skilled technicians are needed to operate and maintain FOSDIC. The equipment requires specialized cameras, and a controlled environment, and is therefore too expensive to deploy in numerous field offices. Because it is not a standard production item, there are no backup machines available from manufacturers

in the event of a breakdown. The computer that controls the equipment is now obsolete and repair parts will not be available in the near future.

Modifying the equipment for use in the next census so that spare parts would be available is expensive. About 260 more modern commercially available scanners could be leased for what the Bureau spent to update FOSDIC for use in the 1980 census. Commercially available scanners can be operated by nontechnical persons with minimal clerical skills.

To use commercially available optical scanners, the Bureau will have to redesign its questionnaire. That questionnaire was designed specifically for use with the FOSDIC system of data capture. The questionnaire page size was larger than commercially available optical scanners can accommodate. Existing optical scanners read single pages. The pages of the census form were joined and could not be easily separated.

To maximize the potential economy and efficiency of utilizing optical scanners for early data capture, the Bureau will have to revise the questions asked in the last census that required manual coding before the data could be captured for automated processing. If entire questionnaires, as completed by respondents are computerized by optical scanners as they arrive, the Bureau could do all processing sooner than if questionnaires contain responses that have to be manually completed before capture. During the 1980 census, clerical checks and manual coding to prepare questions for capture and to verify questionnaire accuracy cost $57 million. These edits were not completed until a year after most district offices closed.

After data are captured, microcomputer systems are capable of performing processing and control activities without human intervention except for reviewing the captured data. Most of the clerical tasks of the 1980 census could be automated including: control over questionnaires, preparing population and housing counts, maintaining address registers, preparing work assignments, and searching files. In particular, the $29.5 million edit operation could be automated. The new small computers are also capable of refining data using statistical editing techniques. Using computer terminals and printers, persons with minimum typing skills could revise and correct data stored in the computer, produce printed copies of address registers, work assignments, counts, or anything else in the computer files.

Possible automation configurations suggest savings in cost to process census data. Two rough, unofficial estimates prepared independently by Bureau staff members indicate that field offices might be automated for a leased cost in the range of $21 million to $25 million during the census field processing

period. While underlying assumptions, such as workload, number of field offices, and equipment differed in the two estimates, both suggest that all processing except the tabulations and reviews at headquarters could be completed by the time field offices closed. The $25 million cost of equipment including microcomputers and optical scanners is considerably less than the $106 million spent for district office processing operations in 1980. The $81 million difference allows funds for staff and possible savings. Both estimates envision the need for far fewer clerks than employed in 1980. Also, these estimates suggest that the specialized processing centers, which cost about $115 million for the 1980 census, could be eliminated.

While these two estimates illustrate the potential attractiveness of an automated decentralized processing operation, they are based on assumptions that may not materialize, and neither may be the best approach for the next census. Instead of the traditional several hundred field offices, the most economical and efficient field office arrangement could be something else, such as 50 State offices with data collecting suboffices. Any number of organizational structures are possible; for example, individual field offices could be equipped differently depending on assigned functions.

ADVANCE PLANNING WITH EARLY DECISIONS NEEDED TO ACQUIRE NEW ADP EQUIPMENT

To design a census that would make much greater use of automation requires an extensive planning effort including the designation of realistic decision points. The Bureau devoted much time and resources in planning the 1980 census. That census incorporated data processing procedures similar to those used in the 1970 census. To change processing procedures will probably require a more intense and longer planning effort.

Formal planning for the 1980 census began in 1973. The Bureau had requested and received funding for fiscal year 1974, which began in July 1973, to start planning the census. For the 1990 census the Bureau also expects to obtain initial funds for planning from its fiscal year 1984 (October 1983 through September 1984) appropriation. The Bureau believed as early as the summer of 1973 that it would continue to use the FOSDIC data capture system in the 1980 census. However, the Bureau in its early preparatory work for the 1980 census experimented, although unsuccessfully, with using increased automation at district offices. The Bureau concluded that ADP technology had not developed to that level where it could be relied on to change from 1970 census processing procedures. The situation is different today as has been discussed.

Selecting, acquiring, and getting a major data processing system ready for use in a decennial census is a long term project requiring a definite set of milestones. In September 1974, the Bureau believed it was too late to get a completely new ADP installation and therefore decided to use its existing computer configuration to process the 1980 census. For that census the configuration included the equipment needed after the data was captured from the questionnaires. The Bureau also believed that considering the long lead-time involved, the fiscal year 1976 budget, submitted to the Office of Management and Budget in September 1974, was the last opportunity for getting funds for the equipment which could have a substantial impact on the processing of the 1980 census.

The Bureau and the Department of Commerce historically have taken 4 to 5 years to have automated equipment available for use after its need was identified. This time period includes the identification of the type of equipment, developing specifications, requesting and evaluating bids, contract award, equipment delivery and installation, and testing the computer programs designed for the equipment. Based on experience, the Bureau would need to start its acquisition process no later than spring of 1986 for the next census.

However, before that time much needs to be done. The Bureau would have to test and evaluate the use and reliability of various types and combinations of equipment. It needs to determine the equipment's compatibility with the type of questionnaire expected to be used in the next census. It also needs to determine the most efficient and economical equipment on the basis of its proposed organizational structure. The appropriate time to evaluate equipment performance under quasi-realistic conditions is during a test census.

For the Bureau to derive the most benefit from its first pretest it needs to make some preliminary decisions and make appropriate plans for the test. The Bureau told us that it plans to examine the possibility of much greater automation prior to the design of the 1990 census pretests. To its credit, the Bureau is planning to have its first pretest of census procedures for the 1990 census in 1985, 1 year earlier than its first test for the 1980 census. This earlier test will allow more time for acquiring equipment and refining the planned procedures for the census. Prior to that pretest, tentative decisions need to be made on the basic elements of the census including the amount of data to be collected and how and where the data will be processed. The data processing technology must be compatible with the other basic elements of the census such as questionnaire design and field organization

structure. To prepare for this test the Bureau needs to identify one or more possible configurations that will be used in the test. The Bureau needs about a year to prepare for its scheduled April 1985 pretest. Time is required to develop procedures, prepare operating manuals, design and print questionnaires and identify the types of staff needed to conduct the test.

INITIAL PLANNING FOR THE 1990 CENSUS NEEDS BETTER DIRECTION

Early Bureau planning efforts did not stress technological change, or provide for the time needed to acquire and test new equipment. Also, some of the Bureau's planning for the 1990 census was not coordinated with other Bureau technological planning efforts.

Early Bureau planning activities for the 1990 census which have not been specifically financed by appropriated funds have emphasized philosophical concepts of census taking, referred to as basic census methodology. Basic census methodology according to the Bureau's Assistant Director for Demographic Censuses includes: determining what constitutes enumeration for a decennial census, and whether the enumeration can be accomplished using statistical methods such as sampling. The Bureau recognizes that the use of some statistical methods in counting the population for apportionment purposes are being challenged in court. The Bureau held a conference in July 1982 to discuss different statistical methods of enumerating the population.

In addition to defining enumeration for census purposes the Bureau has tentatively identified several other methodological issues. For example, the Bureau is considering whether to conduct a 2-stage census for 1990. In a 2-stage census, the questionnaire to be provided to all households would contain just a few basic questions. A supplementary questionnaire with more detailed questions would be provided later to a sample of households. The Bureau plans on making final choices of basic census methodologies by mid-1986. On the basis of the time needed to plan for and acquire ADP equipment, the Bureau would need to decide upon its general questionnaire design long before mid-1986.

In our draft report we mentioned that the Bureau had set the fourth quarter of fiscal year 1988 as its milestone for completing its evaluation for alternative processing techniques and methodologies for the 1990 census. That milestone would allow the Bureau only about 1-1/2 years before census day to complete its acquisition of ADP equipment--a time period considerably less than experience indicates would be needed.

In commenting on our draft report, the Department of Commerce advised that the Bureau has publicly stated that the methodology for processing will be determined in 1986. Subsequent to the receipt of Commerce's comments, the Bureau's Assistant Director for Demographic Censuses advised that the 1986 date was discussed at a Bureau's advisory committee meeting. The Assistant Director also advised that the 1988 date used in our draft report, which was derived from the Bureau's management-by-objective report, was meant to include the finalization of data processing software specifications. The report is used in Commerce's management planning system which was established to improve the Department's performance in accomplishing major policy, programmatic, and management initiatives. The Assistant Director further commented that the report in the future will be revised to officially reflect the 1986 decision on the extent of automation and to clarify the work to be completed by the 1988 date.

The Bureau currently has two long term planning efforts that affect the 1990 decennial census--an activity directed towards replacing its ADP equipment, and an approach for improving geographic support. Neither appear to be well coordinated with the planning for the decennial census. Centralized control is needed to integrate these activities.

The stated objective of the first activity is to implement a fully integrated automated data processing and telecommunications system which will meet the Bureau's mission needs through the mid-1990s. Because of the cost of the anticipated acquisition, about $98 million, the Department of Commerce required that the guidelines defined in Office of Management and Budget Circular A-109 be applied. Under those guidelines there are a number of steps that must be taken to ensure that the new system is responsive to the agency's needs at the lowest life-cycle cost. Because of the various formalized steps in the A-109 process, Bureau officials have expressed their concern about the timeliness of the acquisition for the census. The Bureau's progress is currently about 2 years behind schedule.

As part of the Bureau's future systems design, the A-109 process is directed towards acquiring computers, and ADP equipment for data capture and geographic support. According to the Bureau's head of the A-109 effort, the acquisition of the ADP equipment for data capture and geographic support is on an accelerated time schedule to satisfy the requirement of the 1990 census. However, he considers all equipment to be part of the future design of the Bureau's integrated system.

The Bureau's Assistant Director for Demographic Censuses stated that he is not certain how the Bureau's A-109 planning fits in with the decennial census planning. He advised us that

the Bureau is moving forward on its work on data capture and geographic support for the next census. The Chief of the Bureau's Technical Services Division, responsible for data processing equipment research, advised us in May 1982 that little or no work was taking place for the data capture effort for the decennial census. He also doubted whether the Bureau's A-109 effort could be completed in time to meet decennial census needs for 1990.

In October 1981, the Bureau's Geography Division, part of Field Operations, on its own initiative prepared a detailed technical plan for using computer-assisted procedures. That plan was directed at eliminating many of the problems that occurred in prior censuses including inadequate and late maps and outdated geographic files. However, the plan relies on computer system support that is part of the A-109 acquisition process. As of mid-June 1982, the Bureau had no definite plans to evaluate the Geography Division's proposal. However, the Associate Director for Field Operations established a task force to assess geographic problems of the 1980 census and to develop recommendations for the next census. He established the task force because of concern about the practicality of the Geography Division's plan and its high estimated cost of between $150 million and $160 million. In September 1982, we were advised by Bureau officials that subsequent to the completion of our field work in June 1982, the Bureau approved the plan and has included initial funding for it in the fiscal year 1984 budget request.

Currently, some of the funds appropriated for the Bureau's A-109 effort are being used to buy additional manually operated data capture equipment to place map information developed from the 1980 census onto computer files. The effective use of these files is dependent on the availability of adequate computer capacity presumably governed by the A-109 acquisition.

CONCLUSIONS

In the period since the Bureau planned for the 1980 census, great strides have been made in the design and development of ADP equipment. This advancement provides the Bureau an opportunity to substitute automatic data processing for the labor intensive procedures employed in the 1980 census. The Bureau could derive several significant benefits from a successful substitution. In an ideal situation, wherein all data from questionnaires would be machine readable, the time and cost to release data could be substantially reduced. In addition the Bureau could reduce the human error that frequently is associated with operations which are manual. The problems of recruiting, selecting, and training the temporary workforce could be reduced.

Much needs to be done to take advantage of the possibilities of a more automated census. An early start with a plan that incorporates realistic decision points is essential. In deciding to have the first 1990 pretest 1 year earlier than for the prior census, the Bureau has taken a step in the right direction. Historically, 6 to 7 years have been needed to plan for a decennial census. At least that amount of time will be needed if new ADP equipment is to be considered. Possible data processing changes should be incorporated into the 1985 test so that evaluations could be made early enough in the event that the Bureau decides to acquire new ADP equipment for the 1990 census. Before the test, planning is necessary to decide upon the type of equipment and configurations, questionnaire format, workload, and the organizational structure for taking the census.

At the completion of our field work in June 1982, early Bureau planning efforts were not well coordinated, did not emphasize the need to assess options for evaluating alternatives to the many manual data processing procedures used in the 1980 census, and did not provide for the time needed to acquire and test new ADP equipment. The Bureau plans to obtain initial planning funds for the 1990 census beginning with its fiscal year 1984 budget request. Funds for early planning on automation should be included in that budget request.

RECOMMENDATION

We recommend that the Secretary of Commerce require the Director, Bureau of the Census, to develop a 1990 census plan that includes decision points for evaluating the acquisition, testing, and installation of ADP equipment that are based on past times for planning the 1980 census and acquiring new ADP equipment. The plan should provide for:

(1) An analysis of alternative data processing systems that meet census needs. It should identify the total cost to perform the task including acquisition, maintenance, and personnel.

(2) The possibility of redesigning the 1980 census questionnaire to eliminate or reduce responses requiring manual coding.

(3) An estimate of the expected time to release 1990 census data based on data processing improvements.

(4) Clearly defining the responsibilities of the organizational units working on Census Bureau ADP modernization and identify how their activities will be integrated with 1990 census planning.

(5) A budget for implementing the plan with initial funding requested in the Census Bureau's fiscal year 1984 budget submission to Congress.

(6) Internal periodic reports to assess the progress of the plan and identify any revisions needed.

AGENCY COMMENTS AND OUR EVALUATION

The Department of Commerce agreed with our recommendation and supports the general thrust of the report. (See app. III.) Commerce pointed out that it has started several planning activities to help accomplish the report's recommendation. These include: meetings to discuss and identify those functions most suitable for automation in 1990; considering the use of outside expertise to assist the Bureau in its planning; developing a master schedule of the major completion dates for the next census; reviewing the Bureau's organizational structure to achieve a more effective distribution of functions; recruiting staff to work on the development of processing techniques; and developing plans to test automated coding of questionnaire items.

In addition, Commerce stated that the fiscal year 1984 budget submitted to the Office of Management and Budget focuses on the need for automating the 1990 Census. Commerce also provided a recently prepared paper, included in appendix III, containing the Bureau's general plans for the 1990 Census. Commerce cautioned that the development and implementation of automated systems may not be as easy as it first appears, and it cited several potential problems, such as the need for an automated system that is cost-effective and dependable. However, Commerce noted that our recommendation for more automation is a good one and it will work towards that end.

The planning efforts identified by Commerce are a good start to achieving the objectives of a more automated 1990 census. Proper execution and follow through of these early planning efforts could lead to the accomplishment of the intent of our recommendation. We agree with Commerce that automating the census process is a difficult job with many potential pitfalls. We believe that the difficulty of developing an automated system highlights the need for early planning and adequate testing of a possible changed processing system.

Commerce made several specific comments about some of the details of chapter 3 which are discussed below.

Commerce contends that our report states no effort is being made to evaluate the Bureau's 1990 geographic proposal for using computer-assisted procedures while in fact a committee has been appointed for that purpose. Our report states on page 30 that as of mid-June 1982, the Bureau had no definite plans to evaluate the geographic proposal. The report also states that in September 1982, subsequent to the completion of our field work in June 1982, the Bureau approved the geography plan and included initial funding for it in the fiscal year 1984 budget request. The committee referred to in the Commerce comments was established to evaluate the implementation of the plan.

Commerce commented that our report states that the Bureau's activities for ADP modernization conducted using Office of Management and Budget Circular A-109 are not coordinated, while in fact the Bureau set up a Future Systems Design Staff to handle such coordination.

Our report on page 29 discusses the Bureau's A-109 ADP modernization effort and its approach for improving geographic support, and states that "Neither appear to be well coordinated with the planning for the decennial census." The conclusion was reached because of the uncertainty that exists at the Bureau as discussed on pages 29 and 30 about how the A-109 planning fits in with the decennial census planning, especially automated data capture, and because of doubts about whether the A-109 effort could be completed in time to meet decennial census needs for 1990. The Bureau officials mentioned on pages 29 and 30 who stated those concerns have met with the Future System Design Staff, which is led by the Bureau's head of the A-109 effort. As further evidence that Bureau ADP efforts are not well synchronized, the Bureau's planned reorganization of its organizational and functional responsibilities, mentioned in Commerce's comments, includes its ADP functions; and currently Commerce is reevaluating the Bureau's ADP modernization strategy that followed the A-109 guidelines.

Commerce commented that the statement in the report on page 28 that the fourth quarter of fiscal year 1988 is the milestone for completion of other processing techniques and methodologies is not correct. This comment is discussed on page 29.

Commerce also commented that a master schedule of the major completion dates for 1990 is now being developed. Our report was intended to alert the policymakers for the 1990 census of the need for early planning action with realistic decision points to take advantage of the possibilities of a more automated census. We believe the master schedule is a positive step in the desired direction.

Commerce stated that the figure of $25 million cited in our report on page 25 for automating certain office activities is inappropriate because it was not reviewed or provided by either the Bureau's Acting Associate Director for Information Technology or its Assistant Director for Demographic Censuses, the individuals responsible for assessing automation for the next census. It is not our intent to imply that the rough estimates obtained from Bureau officials constitute the position of the Bureau. In response to Commerce's concerns, the estimates as discussed on page 25 have been specifically labeled "unofficial." It should be recognized that these rough estimates were prepared by those Bureau staff who would be a party to preparing such cost estimates. Also, at the time of our field work, the estimates were the best information available to illustrate that presently developed ADP equipment not available to the 1980 census might be cost effective for the 1990 census.

APPENDIX I

NINETY-SIXTH CONGRESS

ROBERT GARCIA, N.Y., CHAIRMAN

U.S. House of Representatives
COMMITTEE ON POST OFFICE AND CIVIL SERVICE
SUBCOMMITTEE ON CENSUS AND POPULATION
601 HOUSE OFFICE BUILDING ANNEX 1
Washington, D.C. 20515

July 16, 1981

The Honorable Milton J. Socolar
Acting Comptroller General of
the United States
General Accounting Office
441 G Street, N.W.
Washington, D. C. 20548

Dear Mr. Socolar:

It has come to the Subcommittee's attention that much of the data results of the 1980 decennial census will not become available for two to four years after the respondents have returned their questionnaires. This data is obtained from a large sample of the respondents who are required to fill out the long questionnaires. The information covers important population characteristics such as income, employment, education, and housing.

In this era of high speed data processing, it seems highly questionable why the Census Bureau should take so long to process, tabulate and publish data so important to the country.

In view of the extremely high cost of the decennial census, its importance to the country and the long delay in obtaining much of the results, we believe that it would be appropriate for the General Accounting Office to examine the reasons for the long time period needed to process, tabulate and publish the data. Specifically, we request GAO to:

1. determine the time required by the Census Bureau to process, tabulate and publish the information for the sample questions,

2. determine the reasons for the long delays,

3. make inquiries about the technologies such as optical reading that may be available in lieu of the Census Bureau's processes,

4. determine if questionnaire redesign or streamlining would expedite processing, and

5. explore whether cost savings would be achieved through changes in Census Bureau methods, technology, or equipment.

APPENDIX I

The Honorable Milton J. Socolar
Page Two
July 16, 1981

We suggest that your respresentative should contact officials of companies involved with processing statistical information.

Sincerely,

James Courter, Ranking
Minority Member

Robert Garcia, Chairman

GAO AND DEPARTMENT OF COMMERCE'S OFFICE OF INSPECTOR GENERAL REVIEWS OF THE 1980 CENSUS

GAO reports include "Programs To Reduce the Decennial Census Undercount" (GGD-76-72, May 5, 1976); "Problems in Developing the 1980 Census Mail List" (GGD-80-50, Mar. 31, 1980); "Problems in Test Censuses Cause Concern for 1980 Census" (GGD-80-62, Jun. 3, 1980); "Procedures to Adjust 1980 Census Counts Have Limitations" (GGD-81-28, Dec. 24, 1980); "An Assessment of the 1980 Census Results in 10 Urban Areas" (GGD-81-29, Dec. 24, 1980); and "A $4 Billion Census in 1990? Timely Decisions on Alternatives to 1980 Procedures Can Save Millions" (GGD-82-13, Feb. 22, 1982). We also issued two other reports on census pretests and planning, budgeting, and accounting for the 1980 census: GGD-78-2, Oct. 11, 1977 and GGD-79-7, Nov. 9, 1978.

Office of Inspector General reports on the 1980 census covered such areas as security provided over confidential census information, preparatory work for the census, space management and leasing agreements, and a review of allegations of mismanagement of the Bedford-Stuyvesant area recount.

APPENDIX III

UNITED STATES DEPARTMENT OF COMMERCE
The Inspector General
Washington, D.C. 20230

November 9, 1982

Mr. Henry Eschwege
Director, Community and Economic
Development Division
U.S. General Accounting Office
Washington, D. C. 20548

Dear Mr. Eschwege:

This is in reply to your letter of September 7, 1982, requesting comments on the draft report entitled "The Census Bureau Needs to Plan Now for a More Automated 1990 Decennial Census."

We have reviewed the enclosed comments of the Under Secretary for Economic Affairs and believe they are responsive to the matters discussed in the report.

Sincerely,

Sherman M. Funk
Inspector General

Enclosure

APPENDIX III

UNITED STATES DEPARTMENT OF COMMERCE
The Under Secretary for Economic Affairs
Washington, D.C. 20230

OCT 26 1982

Mr. Henry Eschwege
Director, Community and Economic
Development Division
General Accounting Office
Washington, DC 20548

Dear Mr. Eschwege:

Thank you for the invitation to comment on the draft report of the General Accounting Office (GAO) entitled, "The Census Bureau Needs to Plan Now for a More Automated 1990 Decennial Census."

We agree with the recommendations presented, and we support the general thrust of the report. The recommendations are in fact consistent with our current activities and plans for designing the 1990 census. Several planning activities for the 1990 census are already underway. These planning activities will examine other approaches for 1990, including the automation of the 1990 census activities to attain greater cost-efficiency, more timely production of the data, and reduction of reliance on a temporary clerical staff.

Current activities include the following:

(1) During the past 15 months, Bureau representatives of the participating divisions have been meeting periodically to discuss and identify those functions most suitable for automation in 1990. Greater emphasis is being given to the geographic portion of the census due to the necessary long lead time to automate an integrated production system in time for use in the 1990 census.

(2) A master schedule of the major completion dates for 1990 is now being developed.

(3) The Bureau has been reviewing its organizational and functional responsibilities, and a reorganization is planned to achieve a more effective distribution of functions.

(4) Greater attention is also being given to the use of non-Bureau expertise to assist the Bureau in the planning of the 1990 activities, including the use of automation and application of state-of-the-art technology to this massive undertaking.

APPENDIX III

(5) The FY 1984 budget submission to the Office of Management and Budget clearly focuses on the need to develop and use automation for the 1990 census.

(6) A technical staff is being recruited to work within the Bureau on the development of processing techniques for the 1990 census.

(7) The testing of automated coding of questionnaire items is a priority item in testing plans being developed.

A paper was recently prepared containing the Bureau's general plans for the 1990 census; a copy is enclosed for your information. We believe these activities, together with others scheduled over the next 2 years, will accomplish the recommendations offered in this report.

While we agree with your recommendation to automate the census process to the extent possible, there are some potential problems with introduction of automated systems that must be considered. An automated system for 1990 must meet the complex and multiple requirements of the decennial census but must also permit the eventual disposition of any hardware purchased in conjunction with the 1990 census in a cost-effective manner. Any system adapted for use in the 1990 census must meet very stringent requirements regarding maintenance, training of personnel, and accuracy. The time constraints under which the Bureau must conduct and process a decennial census do not permit expensive, time-consuming repair or maintenance activities, complex training, or dependency on a system which might be vulnerable to security infringements. Further, the automated system must allow quick and efficient recovery in the event electrical/mechanical difficulties are encountered. The Bureau cannot afford the problems which would result from shifting from an automated to a manual procedure midstream in the enumeration and/or data capture period. I point out these concerns to indicate that the development and implementation of automated systems may not be easy as first appears. Still, the general recommendation of more automation is a good one and one we will work toward.

We do have some specific comments about the details of the report which are noted below:

(1) The Bureau is required to produce a multitude of data for small areas. The Constitution requires that a decennial census be conducted for the purpose of apportioning the U.S. House of Representatives. Public Law 94-171 further mandates the Bureau to issue population counts at the block level for the purpose of redistricting within states. Additionally, a number of legislative programs were approved by Congress during the 1970's which stipulate that decennial data must be used for the purpose of determining eligibility for Federal programs. These critical uses of census data require that the Bureau ensure the data are accurate and complete, down to the smallest geographic level, since any inaccuracies could ultimately affect the eligibility of a governmental unit for funding or the drawing of precinct boundaries. The importance of this accuracy has not been properly emphasized in the report. Considerable

APPENDIX III

comment in the report was made regarding the editing and processing activities undertaken at the district office and processing centers, with the implication that such checks might be unnecessary. Checks are made to obtain the accuracy necessary to avoid political and monetary adversities of error.

(2) There is a distinction between the collection of data and the processing of data for the purpose of tabulation. While these two operations overlap, they really are two distinct operations. In this report, processing is defined as virtually all activities conducted between Census Day and release of the data products. As defined on page 6, "...data processing activities started after April 1 when questionnaires were received at the district office, and conclude with the publication of the census results." While some of the activities performed in a district office are "processing," we feel this is not a proper time reference. In fact, in the report there is some confusion. On page 13 a statement is made "...processing generally did not start until January 1981...." Additionally, the report's description of the processing system does not acknowledge the scope of the activities completed under the system and suggests, for example, that the correction of geographic boundaries (discussed on page 18) was an interruption of the processing system rather than a component of it. It should be noted that the need to reflect political boundaries as of January 1, 1980, required the Bureau to conduct corrective activities during 1980, and these activities were planned as part of the processing system.

(3) The discussion of the Bureau's 1990 activities overlooks or misrepresents the Bureau's position on several issues:

(a) In discussing the 1990 geographic system, the report states no effort is being made to evaluate the proposal. In fact, a 1990 planning committee on geography has been established with the purpose of reviewing the Bureau's geographic plan relative to its application to the 1990 census program.

(b) In discussing the A-109 activities, the report states the Bureau has not coordinated its efforts. In fact, the Bureau set up a Future Systems Design Staff to handle such coordination. As discussed earlier, meetings have been held over the last 15 months with members of the participating divisions and the Future Systems Design Staff to develop requirements for an automated system for 1990.

(c) It should be noted that the Bureau has publically stated that the methodology for processing will be determined in 1986. The report, however, states that the fourth quarter of FY 1988 is the milestone for completion of other processing techniques and methodologies. This is not correct. The 1986 date is the milestone under which the Bureau is planning to determine process for the acquisition of equipment, thus allowing 3 to 4 years for this activity rather than the 1 1/2 years stated in the report.

APPENDIX III

Mr. Henry Eschwege 4

(4) The report does acknowledge the Bureau encountered some difficulties and resultant delays due to staff turnover, larger than expected work load, and litigation. The effect of these difficulties may not have been accurately portrayed. Processing would have proceeded more expeditiously had the Bureau not been enjoined by the Court from shipping census questionnaires from district offices to the processing centers. At one point, these injunctions in fact postponed the shipment for almost 30 places and cities. These delays were costly not only in a monetary sense but also logistically.

(5) Many people are involved in an operation as large as a census. While we recognize the difficulty of the GAO staff in identifying the responsible managers, we are concerned that failure to contact some key census personnel may lead to some inappropriate statements. For example, the $25 million figure cited (p.25) for automating certain office activities is inappropriate. This figure was not reviewed or provided by either the Acting Associate Director for Information Technology or the Assistant Director for Demographic Censuses, who are the key individuals responsible for assessing automation for the next census.

I would like to thank the GAO for the points mentioned in this report which will assist the Bureau in planning the 1990 census.

Sincerely,

ROBERT G. DEDERICK
Under Secretary for Economic Affairs

Enclosure

GAO note: Page references in the appendix have been changed to agree with the page numbers in the final report.

APPENDIX III

1990 Planning

The planning of the 1990 census program officially begins with the budget authorization of fiscal year 1984, which begins October 1983. There are already a number of activities presently underway which will provide a foundation for the planning period. These include the 1980 evaluation and experimental program results which are now being processed and analyzed; the establishment of a number of 1990 planning task forces mandated to review and/or inventory available options for various issues of the 1990 program, such as basic design methodology, redistricting, coverage research, and recommendations of the data users; and the planning of a mail list compilation test to be conducted in spring 1984.

The Census Bureau plans to spend time studying alternative approaches for taking a census, delaying the selection of the basic 1990 census methodology until about mid-1986. This kind of approach is essentially different from that used in advance of the 1980 census. The development of the 1980 census was based on a decision early in the decade that the census would be mail out/mail back enumeration. The primary focus of the intercensal testing was the evaluation of coverage improvement techniques, the application of automated transmittal of the population and housing counts, and content testing. It is the intention of the Bureau to build the foundation of the 1990 program without making any preliminary choices as to the enumeration and processing methodology to be used. The Bureau intends to inventory possible approaches and test them, as appropriate, prior to the selection of a technique (or techniques) for use in the 1990 census. The 1990 program will expand the use of outside participation initiated in the 1980 cycle to investigate and develop new techniques for the census. As mentioned earlier, the Bureau has already begun to inventory suggestions for the 1990 census. In addition, the Bureau will continue to seek and encourage input from advisory committees, Federal agency councils, professional organizations, etc., as well as some less traditional sources. For example, the Bureau is currently trying to arrange for the 1990 census to be the topic of discussion for an honors class at a university that has taken similar topics in the past.

As a basis for all decisions, however, we must consider the purpose of the census. In the strictest sense, the purpose is to satisfy the constitutional and legislative requirements. As part of the 1990 planning process, the Bureau will examine its approach to meeting these requirements. In particular, what constitutes "enumeration" in order to satisfy constitutional and legal requirements? Should the census continue to attempt to identify each specific person? Can sampling be used to obtain the basic count? Is it possible to use independent counts rather than conduct a direct enumeration? Decisions concerning "what constitutes enumeration?" are fundamental in deciding on a basic methodology for the 1990 Census. As a first step in this process, the Bureau conducted a workshop of national experts in July 1982. The Bureau plans to conduct similar sessions (additional outside conferences-academic, congressional briefings) legal, etc., over the next few years to gather information about this issue and to make a final choice.

APPENDIX III

A number of specific areas will be closely examined during the planning period.

1. The automation of the census process.

2. The examination of alternative collection approaches, in particular, a drop-off/mail back census or a two-stage census.

3. Development of a control list in a different manner from that used for the 1980 census (the Bureau was specifically criticized by GAO for the method used to obtain the mail list for the 1980 census).

4. Preliminary looks at a "paperless" census (example: drop off/phone back, voice recognition interviews, etc.).

5. Coordination of census operations with the automated geographic system including the investigation of the use of satellite technology to specify the geographic location of each housing unit.

6. Investigation of the use of non-Bureau resources (and funds) for input to census activities.

7. Investigation of the use of new statistical techniques in taking or evaluating the census.

Automation

The Bureau is hopeful of automating many of the functions of a census traditionally done clerically. In particular, the Bureau is requesting funds to develop an automated geographic system.

As far as collection and processing are concerned, there are many possibilities for automation. The Bureau plans to test automated check-in (bar codes, etc.) and sorting. There is a long delay between the collection of data and actual processing. The Bureau plans to test capturing the data in the computer at the collection phase. There are many advantages of such early capture. First, there is the ability to process and publish earlier. Second, many of the operations traditionally done clerically (for example, edit) could be done by computer.

The Bureau intends to examine the application of other automation ideas as well; for example, computer assisted interviewing and automated coding of handwritten questionnaire entries. In fact, work on these concepts has been underway for a period of time.

Alternative Collection Methods

Though the 1980 mail out/mail back approach was successful, the Bureau plans to address other techniques. Two main alternatives are drop-off/mail back and a two-stage census.

Delivery of the questionnaire by a census employee may have some advantages over mailing by the post office. There can be cost and time savings on address list production and correction. There is the potential advantage of early contact by a census employee which could improve respondent cooperation. One potential problem is the risk of coverage loss.

APPENDIX III

A two-stage census would allow for quick collection of basic census data. As a separate and sequential operation, data now collected in the long form could be obtained. The major advantage to this approach is speeding up the collection and processing of basic count and characteristics data. One major problem is the difficulty of locating and obtaining cooperation from the appropriate respondent to collect the second-stage information.

Control List

The control list (mail list) in 1980 was either obtained by commercial purchase or was generated by employees hired by the Bureau (prelist). There are alternatives -- allow the post office to prepare the list or update the 1980 list. The cost and accuracy of these techniques will be evaluated in a test in 1984.

Paperless Census

The Bureau also plans to begin experimenting with a "paperless" census approach. There are many alternatives -- more use of phone, direct phone-to-computer, use of home computers, etc. Though it may not be possible to perfect such systems for the next census, they will be carefully examined and considered for testing as an experimental project during the 1990 census.

Coordination with Automated Geographic System

The three sets of geographic products produced for a census (maps, a file to code addresses to specific geography, directory of geographic, and political boundary hierarchy) have been largely clerical operations conducted independently. Nonmatches and errors between the three sets have resulted in costly errors and necessitated unplanned correction activities. The Bureau is planning to develop a computerized system to integrate all these activities, thereby making them consistent. The discussion of the automated geographic system will be handled separately. The possible application of this system to other aspects of the census program will have to be examined. One particular aspect is worth mentioning. In rural areas (areas without house numbers and street names) the Bureau must find a way of associating specific units to specific areas of geography. Clerical use of maps has been the traditional method. The Bureau plans to test the LORAN C technique of obtaining longitude/latitude coordinates.

Local Input

The Bureau will examine potential methods to encourage local participation and support of the census process. Possibilities include the use of local lists for address purposes and local resources for producing specialized files such as election precinct equivalency files if required in 1990 for producing counts needed by the states for redistricting purposes in accordance with 13 USC 141(c).

New Statistical Techniques

The Bureau will examine alternative approaches to the measurement of undercount. In order to avoid some of the litigation and public relations problems of the 1980 census, the conceptual and statistical foundation for identifying undercount and implementing correction techniques must be understood in advance of the 1990 census.

APPENDIX III

To accomplish the investigation, testing, and evaluation of these issues, the Bureau proposes some specific projects during the planning phase which are expected to continue for a period of several years.

1. Special tests to investigate specific issues, such as the Mail List Compilation Test to evaluate alternative techniques for development of an address file.

2. The conduct of the first 1990 pretest in 1985, which will examine an alternative collection methodology such as two-stage or list/leave enumeration, the application of automated check-in of census questionnaires and integration of the new geographic system with the census process. It should be noted that the Bureau plans to conduct the same number of pretests as for the 1980 program, but they will be scheduled on an annual basis rather than at 6-month intervals to provide sufficient time for analysis of each test prior to the preparation of the next one; therefore, the 1990 pretest program is an advancement of 1 year over the first 1980 pretest.

3. Contracting with outside experts to examine the proposed techniques and to develop alternative methodologies and systems for use in the 1990 census; for example, data capture and training.

4. The development of automated systems for application in the enumeration and processing activities including automated coding of handwritten questionnaire entries, computer-assisted interviewing, and data capture techniques; for example, in-house development and file matching.

5. Investigation and testing of coverage measurement and application for undercount adjustment. The Bureau proposes to examine new technologies for demographic estimation and the application of statistical techniques such as those now used by Statistics Canada. An understanding of how and what to measure in determining census coverage is critical to both the conduct of the census and the use of data products derived from the census.

6. Encourage the participation of the primary data users (Federal, state, and local agencies) to:

   (a) Inventory their program needs which require specific information from the 1990 census.

   (b) Obtain their recommendation on the design and procedures for the 1990 census.

   (c) Explore joint ventures which will address shared problems, such as address list development, local review, and the provision of data by non-census geography such as election precincts.

APPENDIX III

7. The investigation and development of alternative sampling techniques such as nested samples and sampling controls required for a two-stage census.

Each of these issues requires the investigation, testing, and evaluation of complex statistical and/or technological application to the census, each of which must be understood within itself before it can be evaluated in combination with the other components of the census program. All of these issues are interrelated and interact with one another. For example, the enumeration methodology affects the application of automation to accomplish specific activities, the use of satellite technology, and the use of statistical technology for taking or evaluating the census. Likewise, each of those can determine the requirements for the actual enumeration methodology.

Although various standards will be used to test the effectiveness of alternatives, the selected techniques must support three objectives:

1. Cost effectiveness

2. Timely release of data products

3. Continued high coverage of the Nation's population and housing inventory with attempts to reduce the undercount differential for persons by race and/or Spanish origin.

In conclusion, the planning of the 1990 census program will actively research, and test the application of alternative statistical techniques and automation of the census operations to accomplish the mission of the decennial census.

(275161)

23984

AN EQUAL OPPORTUNITY EMPLOYER

UNITED STATES
GENERAL ACCOUNTING OFFICE
WASHINGTON, D.C. 20548

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE,$300

POSTAGE AND FEES PAID
U. S. GENERAL ACCOUNTING OFFICE

THIRD CLASS