United States General Accounting Office

GAO

Briefing Report to Congressional Requesters

May 1986

DECENNIAL CENSUS

Status of Plans to Computerize Questionnaire Data

GAO/GGD-86-76BR

GAO

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

General Government Division

B-222850

May 5, 1986

The Honorable William D. Ford
Chairman, Committee on Post Office
and Civil Service

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

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

Your March 18, 1985, letter requested that we assist the Subcommittee in its oversight and evaluation of the Census Bureau's 1990 Decennial Census planning efforts and stated that controlling the cost and improving the quality of the census will be high priority items for the Subcommittee. You asked that we place special emphasis on the Bureau's automation activities so that cost savings and timely reporting of census results could be effected.

According to the Bureau, one of the most promising ways to improve the census is to convert the data obtained from the questionnaires to a computer-readable format (data capture) at an earlier stage than in the 1980 census. This approach is important if data products are to be released earlier. Earlier data capture will also help to improve data accuracy by allowing for computerized editing and more time for review. Computer records of questionnaires could serve as backups to the originals in case they are prematurely destroyed, a situation that occurred in the 1980 census. Also, we believe it could help to hold down costs by substituting automation for a number of manual, labor-intensive processing efforts performed in 1980. However, optimizing the advantages of automation requires much advance detailed planning and early decisions.

This briefing report focuses on the Bureau's planning efforts with an emphasis on its activities on data capture and its decision to exclude optical mark recognition (OMR) technology as an option for the 1990 census. The Bureau made an early commitment to increased automation, but detailed planning started late and progressed at a slow pace. Recent planning activities have been accelerated, but we doubt that this will make up for lost time. As a result, the Bureau's planning activities may have jeopardized its ability to achieve the maximum benefits from computer technology and reduced its opportunity to hold down costs for the 1990 census.

We are particularly concerned with the Bureau's planning efforts for the possible use of OMR technology. We believe the Bureau's decision to discontinue consideration of OMR technology was influenced by its late start in detailed planning, reluctance to revise the questionnaire form, and a slow procurement process. Whether OMR equipment could have been adapted for use during the 1990 census may never be known. Because of its actions, the Bureau may have excluded a potentially useful option without fully exploring it. Because the OMR data capture option has been eliminated from consideration, the Bureau is left with two remaining options: its unique film-to-computer-tape technology used in the past three censuses and data keying.

Because of our concerns with the Bureau's planning efforts to date we question whether the Bureau

--will be in a position to make informed judgments on data capture in a timely manner,

--will effectively automate data capture using the remaining two options, and

--can decentralize its processing offices to reap important benefits from processing the 1990 census questionnaires during the data collection phase while maintaining a manageable and cost-effective census.

Detailed information on the Bureau's efforts to automate the 1990 census are included in the appendix to this report. We are also issuing an accompanying briefing report that focuses on the 1990 short form questionnaire, which we believe should be shorter and simpler than the 1980 short form.

Our report on automation was prepared on the basis of

discussions with officials of the Census Bureau, the National Bureau of Standards, the Department of Commerce, and vendors of data capture equipment; our observations of the data processing activities in the 1985 test census; our review of Census Bureau's OMR evaluations and procurement documents, including correspondence; and the National Bureau of Standards' assessment of the Bureau's decision on OMR. In addition, we reviewed position papers prepared by the Census Bureau and National Computer Systems (NCS) on the termination of the OMR option and the Bureau's decennial census planning documents. We also used information developed from prior work on the 1980 census.

We obtained oral comments from Census Bureau officials. In general, these officials believed that the Bureau has a planning process to guide its decisions although they recognized that it could be improved, and stated that they were reluctant to take risks with the OMR when they already had their film-to-computer-tape technology. Bureau officials also offered a number of technical and wording comments which we considered in preparing the final report. Our work was done in accordance with generally accepted government audit standards. It covered the period up through March 1986 and thus does not reflect Bureau decisions subsequent to that date.

As arranged with your office, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days after its issue date. At that time we will send copies to the Senate Subcommittee on Energy, Nuclear Proliferation, and Government Processes; other appropriate congressional committees; the Secretary of Commerce; and the Director, Office of Management and Budget. Copies will also be made available to other interested parties upon request.

If you have any questions about this report, please call me on 275-8387.

Gene L. Dodaro
Associate Director

APPENDIX

NEED FOR INCREASED AUTOMATION FOR 1990 CENSUS

- CENSUS WORKLOAD AND COSTS HAVE INCREASED
- GAO IDENTIFIED NEED TO AUTOMATE MANUAL OPERATIONS
- BUREAU IS COMMITTED TO INCREASED AUTOMATION

NEED FOR INCREASED AUTOMATION FOR 1990 CENSUS

A decennial census is a mammoth task. For the 1980 census, the Census Bureau processed 88 million questionnaires containing 3 billion items of data about the Nation's 226.5 million persons and their housing. The Bureau employed a temporary work force of over 280,000 persons and spent about $1.1 billion. The cost, exclusive of inflation and workload, was twice the cost of the prior census. Even with this effort, much of the data, exclusive of the population count, was published 2 to 3 years after Census Day. Additionally, about 50 communities and groups challenged the accuracy of census results. With an expected increase in the number of households for 1990--an estimated 106 million questionnaires to be processed--the decennial workload will be even greater and cost estimates to complete the next census have been as high as $4 billion.

In a report, The Census Bureau Needs to Plan Now for a More Automated 1990 Decennial Census (GAO/GGD-83-10, Jan. 11, 1983), we described the various data processing activities performed in the 1980 census and discussed the possibilities of automating a number of labor-intensive operations. For example, about 37,000 clerks had to check returned questionnaires for complete and consistent entries before the questionnaires were sent to processing centers to be converted to automated files. This function could be performed automatically if the data were initially converted from the questionnaires to computer files.

We made several recommendations geared toward increasing productivity and reducing the cost of the census through automation. These recommendations included

- --analyzing alternative data processing systems for decennial census operations,
- --possibly redesigning the census questionnaire to eliminate or reduce responses requiring manual coding, and
- --assessing progress made on a plan for automation on a periodic basis.

The Department of Commerce agreed with the thrust of our report. It specified that approaches would be examined to automate 1990 census activities to attain greater cost efficiency

APPENDIX

and more timely production of the data and to reduce the reliance on a temporary clerical staff.

The Bureau testified before your Subcommittee on April 18, 1985, that one of the most promising ways to improve the census, and one of its biggest challenges, was to convert the response data on questionnaires to a computer-readable format (data capture) earlier in the census process. The Bureau believed this approach was essential if it were to release data products sooner. Earlier data capture would help improve accuracy by allowing computerized editing and by providing more time to review the count data before they are provided to the President and the states by the legally mandated dates, and it would help to hold down costs. The Bureau has also pointed out that early data conversion would provide a computerized back-up file to the paper questionnaires in the event of premature destruction of the original documents as happened during the 1980 census.

THE BUREAU'S LATE START IN DETAILED PLANNING MAY LIMIT SUCCESS OF DATA CAPTURE AUTOMATION FOR 1990 CENSUS

- GAO STRESSES NEED FOR EARLY PLANNING TO OBTAIN OPTIMUM BENEFITS FROM AUTOMATION BECAUSE OF THE LONG LEAD TIME REQUIRED TO OBTAIN EQUIPMENT

- NO MASTER PLAN FOR 1990 CENSUS WAS DEVELOPED UNTIL FEBRUARY 1985

- DECISION CONFERENCE IN OCTOBER 1985--GAPS IDENTIFIED IN INFORMATION

- SUPPLEMENTAL PLANS BEING DEVELOPED, BUT THEY WILL NOT BENEFIT FROM TEST EXPERIENCE

THE BUREAU'S LATE START IN DETAILED PLANNING MAY LIMIT SUCCESS OF DATA CAPTURE AUTOMATION FOR 1990 CENSUS

In our January 1983 report, we stress the need for advance planning and early decisions to obtain optimum benefits from automation because of the long lead time required to obtain needed equipment. Based upon historical experience in the Bureau and the Department of Commerce, 4 to 5 years would be needed to make automated equipment available after its need was identified. This time period includes identifying the type of equipment, developing specifications, requesting and evaluating bids, awarding contracts, installing equipment and testing. We mentioned in the report that although the Bureau had expressed interest in increased automation, its initial planning efforts for the 1990 census were not well coordinated. We said in the report that if the Bureau did not plan adequately and make decisions early, time would not permit the Bureau to take advantage of automation opportunities.

In the Department of Commerce's response to our report dated October 26, 1982, it stated that a master schedule of the major completion dates for 1990 was being developed. Before the development of a master schedule, the Bureau established committees and held conferences to consider 1990 census approaches. Although Bureau officials said a document had been prepared as early as March 1984 which included dates and events, as late as December 1984 the lack of such a schedule was cited by the Secretary of Commerce as a material internal control weakness in his annual report required by the Federal Managers Financial Integrity Act. A master schedule for accomplishing the 1990 decennial census was not prepared until February 1985.

The schedule was incorporated as a part of the Bureau's management information system (MIS). The MIS established decision points and milestones for various 1990 census planning issues and activities. The decision date established for the data capture technology was September 1986. The MIS plan, however, is not a detailed plan. It does not include detailed information on what criteria will be used or how

--required decisions will be made;

--information will be gathered to make these decisions; and

--additional requirements will be identified, defined, resolved, and completed.

In October 1985, the Bureau held a decision conference. The major issue discussed was the configuration (number, composition, and location) of the data processing offices. As an initial result of this conference, the Bureau decided to (1) process the 1990 census questionnaire during the data collection phase (called "concurrent processing") and (2) consider using 12 to 24 large processing offices using high speed equipment for the more urbanized areas and 50 combined collection/processing offices using data keying for the more rural areas.

However, the Bureau recognized that additional information had to be developed before these initial decisions could be implemented. The Bureau decided to prepare action plans to obtain information needed to answer some basic questions, such as the problems and possibilities associated with decentralizing a modified version of the data capture system used in the 1980 census. The first portion of these action plans was scheduled for completion in March 1986.

The data processing configurations that were initially decided upon in the October conference are being tested in the 1986 pretests. However, the action plans are scheduled to be completed before receipt of evaluative information from these tests.

GAO'S OBSERVATIONS

Our basic concern is that the detailed planning needed to optimize the benefits of automation was not started early enough. As a result, decisions reached on identifying the number and type of data capture equipment may be controlled by the time remaining in the census planning period rather than by what type of equipment is best suited for use in the census.

The Bureau's evaluation of OMR technology, as discussed in the next section, provides a good illustration of how the lack of detailed advance planning and early decisions have limited the Bureau's options for data capture.

RESERVATIONS ABOUT CENSUS BUREAU EVALUATIONS OF OPTICAL MARK READER (OMR) TECHNOLOGY

- THE BUREAU DID NOT REDESIGN THE QUESTIONNAIRE OR REDUCE ITS CONTENT TO ALLOW FOR PROCESSING BY COMMERCIALLY AVAILABLE EQUIPMENT
- EQUIPMENT WAS TESTED WITHOUT ELIMINATING KNOWN CONSTRAINTS THAT WOULD LIMIT ITS USEFULNESS
- NOT ALL EQUIPMENT FEATURES WERE CONSIDERED
- PROCUREMENT EFFORTS PROCEEDED SLOWLY
- RESEARCH WAS TERMINATED ON MODIFIED OMR WITHOUT TESTING

RESERVATIONS ABOUT CENSUS BUREAU EVALUATIONS OF OPTICAL MARK READER (OMR) TECHNOLOGY

The Bureau initially considered three types of technologies for capturing questionnaire data for the 1990 census: (1) the film optical sensing device for input to computer and automated camera technology, known as the FACT system, a technology used in the 1960, 70, and 80 censuses, which is unique to the Census Bureau; (2) direct data entry which entails keying information onto a computer disk; and (3) optical mark reader (OMR) technology which involves the use of scanning equipment to pick up and record data from a precoded format. The OMR was a new data capture technology considered by the Bureau for the 1990 census.

The Bureau, however, limited its research of the OMR to one commercially available off-the-shelf version even though it knew that the off-the-shelf OMR was not feasible to use as the primary data capture equipment for the census unless changes in the Bureau's requirements were made. This was pointed out in our January 1983 report. For example, commercially available OMR equipment is designed to process a form that is 8-1/2 by 11 inches, whereas the census form is 11 by 28 inches. Also, the Bureau for the past several censuses has used a multipage booklet form for its long form questionnaire sent to a sample of the Nation's households. The OMR is not designed to process a multipaged form unless the pages are separated.

The OMR also was designed for use in a controlled environment, such as grading student test answer sheets, where the students are provided with #2 pencils and the answer sheets are not folded, and conducting censuses in foreign countries where specially trained enumerators complete the forms. The U.S. decennial is performed in an uncontrolled environment where people complete the forms in their own residences and use any available writing instrument. In addition, the questionnaire is folded in order to accommodate the mail-out/mail-back concept used by the Bureau.

In July 1984, the Bureau leased an OMR for in-house research and development testing. The Bureau's report on its testing stated that the primary purpose of its evaluation was to determine possible hardware/software operational problem areas and to evaluate whether identified problems could be corrected. The

report cited problems experienced with the OMR and the vendor. Several of these problems related to known limitations, such as paper size, controlled operating environment requirements, and required marking instruments. Other cited problems related to the amount of debugging required after the vendor certified the machine as ready to operate, paper feed jams, and vendor responsiveness. It should be noted that as early as October 1984 the vendor recognized some of these problems and began preparing plans to modify its equipment.

In spite of the known limitations of the OMR and problems experienced during in-house evaluations, the Bureau decided to use it in the 1985 test census without modifying the equipment or reducing the content of the questionnaires to shorten the form. Bureau officials advised us that the test could assist in identifying the magnitude of the limitations and problems.

The OMR was used in the 1985 test census to capture data provided by Tampa, Florida, residents. The OMR only processes an 8-1/2- by 11-inch form; however, the Bureau decided that all of the 1980 questions from a much larger form be placed on a single two-sided 8-1/2- by 11-inch sheet. As a result, the questionnaire for the Tampa 1985 pretest was physically reduced in size but still contained all of the 1980 short form questions.

This reduction in form size provided much smaller spaces for the questions and responses than that provided in 1980. As an apparent result, respondents had difficulties with the form. This problem was confirmed by comparing the nonresponse rates in the Jersey City pretest which used the 1980 forms to the Tampa nonresponse rates. For example, many Tampa respondents did not answer the question on sex (12 percent compared to 2 percent in the Jersey City test). In addition, an even greater number of respondents did not answer the question on century of birth (17 percent compared to 4 percent in the Jersey City test).

Despite several breakdowns requiring minor repairs, the OMR operated well mechanically during the test census. According to a Bureau report, none of the breakdowns required major maintenance. In addition, Bureau personnel found it easy to operate. On several occasions, personnel unfamiliar with the OMR were taught to operate the equipment with less than an hour of training.

Bureau evaluations of the OMR during the test census were not planned to identify its full potential

The Bureau stated that the 1985 test census was to provide information on effects of OMR limitations on census processing requirements. The Bureau, however, did not plan the necessary evaluations required to fully test the OMR's limitations or potential. For example, the OMR's built-in edit capability was not evaluated because the Bureau was concerned that the edit might have slowed down processing of the 1985 test questionnaire. According to the vendor, the OMR has the ability to perform an edit of the data it captures and to notify the operator, on the computer screen, of any problems encountered either with double marks or with document alignment during the scan.

Another area not evaluated was the number of instances where respondents did not use #2 pencils and the success of the OMR in reading other marking instruments. The type of instruments used by respondents to complete census questionnaires is likely to affect any automated equipment used for data capture. For the 1985 test #2 pencils were enclosed in all questionnaires sent to Tampa residents.

The Bureau did not evaluate the number of instances where the pencil was not used, and it also did not evaluate whether the OMR read questionnaires not completed with #2 pencils. Although the Bureau did not collect the data, we observed that the OMR did in fact read some questionnaires that were completed using various inks. Similarly, it was not clear whether errors recorded were due to machine errors or improper completion by respondents. Following the test census some Bureau analysts concluded that the OMR was very accurate in capturing data, while others believed that too many errors were attributed to the respondent rather than the OMR.

The Bureau also did not test the OMR's capability to process the multipaged booklet used as the census long form questionnaire. The booklet could not be tested in its bound format but could have been separated for processing and then collated and stapled. The vendor told us that the OMR software could be adapted to read multipaged booklets one page at a time and retain the forms in proper sequence. Vendor officials noted that other off-the-shelf OMRs have been retrofitted with a stapler to assure that no pages are lost.

A Bureau official observed the processing of a multipage form at one of NCS's offices but decided that it was too risky for the pretest because numbers would have to be printed on each page of the questionnaire and if a page should happen to get out of sequence all forms processed thereafter would be out of sequence. While this may be a valid concern, the ability of the OMR to process multipage census forms would not be known unless it was tested. It seems reasonable that this potential would have been tested since the Bureau says its workload to process the multipaged long form questionnaire sent to about 19 percent of the households is equal to that for processing the short forms sent to the balance of the households.

Vendor officials told us that their involvement with the in-house Bureau testing and the 1985 test gave them a better understanding of Bureau requirements. This experience led the vendor to submit a second proposal to the Bureau for modifying the existing OMR model.

Modified OMR Procurement Slow and Eventually Terminated

In January 1985, National Computer Systems (NCS), the OMR vendor, submitted an unsolicited proposal to the Bureau outlining a plan to develop a modified OMR that would better meet the 1990 decennial's data capture requirements. The potential delivery date for this equipment was January 1987. In March 1985, the vendor amended this proposal to reflect delivery of a modified OMR by February 1986 which could capture data from an 11- by 17-inch form for use in the Bureau's planned 1986 pretest and a prototype of a fully modified model by February 1987. NCS submitted another proposal in April 1985 to provide the modified OMR with the capacity to process the larger size form by late January 1986.

The Bureau placed an announcement in the June 20, 1985, Commerce Business Daily (CBD) notifying interested parties that the Bureau intended to purchase two modified OMRs from NCS. Two machines would be delivered by January 1986 for the 1986 test census, and one would have several additional enhancements completed by September 1986. Minimum specifications were used to describe the expected capabilities for each machine. Although NCS proposed a fixed price for developing the required equipment, the Bureau announced its intentions to award the contract on a cost-reimbursable basis. This type of contract usually encourages more vendor interest.

The notification stated that this procurement was specifically for a prototype which would be used to determine the equipment's potential for use in the 1990 census. Any future procurements would be competitively awarded. The Bureau planned to award the contract for the prototypes to NCS on a sole source basis because of the necessary modifications and delivery requirements for the two machines. Interested vendors had 30 days to submit written notification if they could provide the same services/equipment described in the notice, including detailed data on their capability to respond to the announcement.

Because seven vendors responded to the announcement, the Bureau decided that competition was necessary. On August 7, 1985, NCS was notified that a sole source contract was no longer possible and that the procurement would be competitive due to the number of vendors that were interested in providing the required equipment.

The responses from the interested vendors varied in length and content. None of the responses included a specific proposal on when or how the announced requirements could be met, and only one vendor claimed to have access to existing OMR equipment that could be modified to meet test census requirements. This vendor stated that it had two solutions, both involving the modification of equipment that was made by two other manufacturers. Several of the other vendors requested additional information to prepare a proposal. Based on these responses the Bureau planned to release a request for proposal (RFP) in mid- to late August.

The RFP was never issued; instead, a Request for Information (RFI) was sent on September 12, 1985, to each of the vendors responding to the CBD notice. The RFI requested comments from the prospective vendors within 15 days on several issues, including their ability to provide the equipment within required delivery schedules and their willingness to proceed under a firm fixed-price contract.

The RFI specifications were developed to reflect the potential 1990 decennial environment and not the 1986 test census environment. In addition, the delivery date was changed to early July 1986. The cover letter to the RFI stated that the Bureau intended to issue an RFP within 2 weeks after comments were received and expected to receive proposals within 30 days following receipt of comments.

NCS was one of two respondents to the RFI and the only vendor to provide comments on the draft specifications. NCS replied that it was able and willing to meet the requirements outlined, with a few minor exceptions. In November 1985, about 10 months after its initial proposal, NCS was told that the Bureau had decided to terminate efforts to obtain a modified OMR.

According to the Bureau, its decision not to continue efforts to test and consider the OMR as a primary data methodology for the 1990 census was based on

--costs to develop the proposed system;

--the short time frame remaining to accomplish the modifications;

--the risk of the system not succeeding, particularly if it was operated in a decentralized mode where it may be difficult to stabilize the environment;

--results of tests previously performed; and

--archival concerns in providing copies of the data.

However, we have noted that some of these same concerns may exist for one or both of the other two data capture methodologies still being considered for the 1990 census. A discussion of concerns about the other technologies is included on pages 18 to 21.

National Bureau of Standards' Assessment of Bureau Decision

On December 20, 1985, the Department of Commerce requested that the National Bureau of Standards (NBS) make a technical review of the Bureau's decision to cease considering the OMR technology. This request was precipitated by NCS' complaint, submitted to the Department, that the Bureau's decision was unwarranted. NBS' report was requested by January 20, 1986.

NBS concluded that commercially available OMR scanners do not meet the Bureau's needs for the 1990 census but noted that no testing by the Bureau was necessary to reach this conclusion. NBS also noted that there is now insufficient time to make the necessary modifications so that commercially produced equipment could be used in 1990. NBS also suggested that in order to use existing commercially available scanners, Bureau requirements, such as form size, would have to be revised.

GAO'S OBSERVATIONS

The cancellation of the OMR evaluation ruled out one data capture option, the only new technology being considered. The Bureau's efforts to explore the usefulness of OMR technology were undermined by a testing and procurement process which started late in the 1990 census planning cycle and consumed a considerable amount of time. Testing commercially available equipment was prudent only if the Bureau was willing to consider a change in the size and composition of its questionnaire form. Absent this willingness, the Bureau should have initiated a research and development arrangement early in the decade to develop a modified OMR to meet its requirements.

APPENDIX

STATUS OF REMAINING DATA CAPTURE OPTIONS-- UNANSWERED QUESTIONS

- WILL THE CENSUS BUREAU BE IN A POSITION TO MAKE INFORMED DECISIONS ON DATA CAPTURE BY SEPTEMBER 1986?

- WILL THE BUREAU EFFECTIVELY AUTOMATE DATA CAPTURE USING THE MODIFIED FACT SYSTEM OR DATA KEYING?

- CAN THE BUREAU DECENTRALIZE ITS PROCESSING OFFICES FOR 1990 WHILE MAINTAINING A MANAGEABLE AND COST-EFFECTIVE CENSUS?

STATUS OF REMAINING DATA CAPTURE OPTIONS--UNANSWERED QUESTIONS

After the Bureau's decision to eliminate the OMR technology from any further consideration for the 1990 census, the two remaining options are FACT and data keying. The Bureau must evaluate these technologies in view of its desire to benefit from concurrent processing, which has the objective of converting the data on the questionnaires into computer-readable format earlier in the census process than in past censuses. Currently, the Bureau is considering having many more processing offices or combination district/processing offices in 1990 (12 to 173) than in 1980 (3) to expedite the concurrent processing. Thus, the extent of decentralization is an essential issue that must be resolved by September 1986 when the Bureau plans to select the data capture technology and the processing office arrangement. These decisions will have an important bearing on the data processing costs and on the Bureau's ability to effectively manage and control the data processing operations in the 1990 census.

FACT System

For the 1980 census the Bureau captured the data at three processing offices using the FACT system. This system incorporates three distinct processes: first, the questionnaires are microfilmed, then the film is developed, and lastly, the film is read by scanning devices. In the 1980 census all three FACT processes were performed at each of the three processing offices. For 1990 the Bureau is currently considering filming the questionnaires at possibly 48 locations. However, because of concerns about the Bureau's ability to have the film developed at so many locations, it is also considering having this function and the scanning performed in a centralized mode. The FACT system is being used in the 1986 pretest but all three processes will be performed in one office. Therefore, the Bureau's option of using a decentralized microfilming approach and centralized film development and scanning is not scheduled to be tested before the September 1986 decision date.

In 1980, questionnaires were reviewed and edited before being sent to the cameras for filming. Thus, problems with the forms, such as improper markings, could have been corrected before filming. In the 1986 pretest, however, the questionnaires will not go through an editing process before filming.

In the prior census, the questionnaires were boxed and shipped to the processing offices from the collection offices after the work was completed. In many instances the boxes remained for a period of time in the processing office before filming. According to a Bureau official who helped supervise the 1980 data processing operations, these conditions in 1980 could have had the beneficial effect of smoothing out the forms and removing excess humidity from them. This is not anticipated for the 1990 census and will not be done in the 1986 test.

If the Bureau does film the questionnaires in a number of offices, another consideration is its ability to assemble a sufficient number of cameras in the time remaining as well as hire a sufficient number of persons to operate the equipment. For 48 offices, the Bureau will need 96 cameras and employees to operate the equipment in each office. It has about 30 cameras remaining from the 1980 census.

Another complicating factor is that the Bureau anticipates that it will have to modify its existing cameras and scanners to fully satisfy the demands for 1990. However, the Bureau has not identified the modifications needed, the costs to make these modifications, nor whether all the work can be accomplished in-house in the time remaining. Some of these needed modifications will be determined from the results of the 1986 tests.

DATA KEYING

Data entry keying is the slowest most error-prone and least automated of the three types of technologies considered for 1990. It is also the most expensive. When the Bureau and NBS developed the forerunner of the current FACT system in the 1950s, they recognized that keying was too slow for the massive amount of data collected in a decennial census. The decentralization of the keying process currently being considered by the Bureau raises many important issues. Will the Bureau be able to

--hire and supervise the thousands of data keyers needed?

--assure the quality of the captured data?

--afford the costs of paying the keyers and purchasing the needed pieces of equipment and still achieve its goal of not increasing the unit household costs beyond those incurred for the 1980 census?

Another unanswered question about keying concerns is how the Bureau will satisfy archival requirements. The Bureau has not yet determined whether keyed questionnaires will have to be filmed or if a data tape will be acceptable for archival requirements.

The Bureau believes that keying is the most flexible of the data capture methodologies under consideration because it will allow the Bureau to capture handwritten data such as names. Name capture is an integral part of the Bureau's evolving approach to determining the census count's completeness and could possibly be used for count adjustment purposes.

On the other hand, if names are keyed and incorporated into a computer file, it may create a negative public perception on the privacy of census data. In recent census history, the Bureau has publicized that names have not been included on computer files; only the marks on the forms have been recorded using the FACT system.

Unfortunately, final evaluations of the 1986 test, which in part incorporate data keying which would be useful in weighing the advantages and disadvantages of that technology, will not be completed prior to the Bureau's decision date on technology in September 1986.

GAO'S OBSERVATIONS

If the Bureau does not address the unanswered questions on the remaining two data capture technologies and related data processing arrangements before its announced decision date of September 1986, it may have problems achieving a census that is cost efficient and manageable and that provides timely and accurate data.

(016000)

Requests for copies of GAO reports should be sent to:

U.S. General Accounting Office
Post Office Box 6015
Gaithersburg, Maryland 20877

Telephone 202-275-6241

The first five copies of each report are free. Additional copies are $2.00 each.

There is a 25% discount on orders for 100 or more copies mailed to a single address.

Orders must be prepaid by cash or by check or money order made out to the Superintendent of Documents.

34418

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

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100