U.S. DEPARTMENT OF COMMERCE
Office of Inspector General

**PUBLIC
RELEASE**

BUREAU OF THE CENSUS

*A Better Strategy Is Needed for
Managing the Nation's Master Address File*

Inspection Report No. OSE-12065/September 2000

Office of Systems Evaluation

Table of Contents

EXECUTIVE SUMMARY

The 2000 Decennial Census enumerates the U.S. population and housing as of April 1, 2000. This evaluation focuses on the Master Address File (MAF), which supplies addresses used to support operations responsible for mailing out questionnaires, enumerating nonresponding households, and controlling the collection and tabulation of population data. The quality of MAF addresses directly affects the accuracy, completeness, and cost of the decennial. In the 1990 decennial, one-third of persons missed were not counted because address data for their housing units was missing from the address file. Also, the General Accounting Office reported in the 1990 decennial that the bureau spent $317 million on operations to identify 4.8 million nonexistent housing units and 8.6 million vacant housing units and to remove the former from the address file.

To address questions raised regarding how well programs to create the MAF for Census 2000 were working and what its quality would be at the start of the decennial, the bureau decided in 1997 to reengineer the MAF-building strategy. The objectives of our evaluation were to (1) determine if steps taken to improve the MAF before delivery of the address file worked as planned, (2) assess steps taken to identify and correct MAF data quality problems as the decennial progressed, (3) determine whether the software development approach ensured high quality data, and (4) evaluate whether the data quality standard for the MAF provides a meaningful benchmark for decision-makers. This evaluation focuses on city-style addresses, which comprise over 80 percent of the nation's residential addresses.

The bureau estimated through demographic methods that as of July 1998, there were 112.5 million housing units nationwide. The addresses contained in the MAF represent the nation's housing units and, together with each address's geographic location found in the Topologically Integrated Geographic Encoding and Referencing (TIGER®) mapping system, provide an essential tool for collecting responses and counting people where they are located. Unless found to be nonexistent or duplicate, an address equates to a housing unit in the decennial census. For the decennial, the bureau created the decennial MAF (DMAF), which contains all MAF addresses meeting decennial eligibility requirements such as being a residential address that links, or *geocodes*, to a unique geographic location in TIGER.

To build and maintain the address file, the bureau has implemented operations designed to decrease *undercoverage* (missed housing units) and *overcoverage* (duplicate, nonexistent, and other erroneous addresses). The bureau's reengineered strategy for city-style addresses allowed more time for local governments to review MAF address lists and submit corrections. A key component of the reengineering was 100 percent block canvassing. Designed to verify address data provided by local governments and the Postal Service, bureau employees were to canvass 100 percent of the blocks in an assigned area to confirm existing MAF addresses and add new ones.

The bureau did not have sufficient time available to ensure high quality address data. Reengineering of the MAF had to be completed in the two years before the addresses were needed for labeling questionnaires and did not leave enough time to accommodate difficulties in receiving addresses from local governments in time to be verified by block canvassing. We found that incomplete address lists were used in block canvassing, reducing its effectiveness in improving address quality. Moreover, the bureau did not have time to resolve questions about the accuracy of over 5 million addresses and decided to include them and an unknown number of duplicate addresses in the decennial until more information became available. (See page 13.)

The bureau has taken steps to improve address quality and has potentially identified 10.2 million nonexistent or duplicate addresses, approximately 4.3 million before nonresponse followup and an additional 5.9 million during nonresponse followup, while identifying and retaining 9.7 million vacant housing units. The official number of vacant and nonexistent housing units will be determined after a subsequent field operation, which will provide information to confirm vacant and nonexistent addresses, convert them to occupied, or remove them from the census. However, the policy for determining the addresses eligible for census operations has not been well-defined, and, at the time of our field work, the decision about which addresses to include in the final decennial results had not been made. Rather than being presented explicitly in the bureau's decision memorandum series that documents decennial policy, address eligibility rules are implicit in software specifications, which often are not finalized until data processing for the operation is imminent or underway. Finally, the bureau could increase use of information already on the MAF to identify missing housing units and potential errors. (See page 18.)

MAF addresses need to be linked to unique geographic locations (geocoded) to ensure that the bureau can count persons in their correct locations and that users of the data can accurately redraw congressional, state, and municipal legislative district lines. Because MAF and TIGER were developed separately and are not integrated, consistency between them cannot be easily maintained. TIGER data can be modified without ensuring that both databases have accurate and consistent information, causing some decennial addresses to no longer link to TIGER. In addition, some decennial addresses received a geocode from block canvassing but do not link to TIGER. As of April 2000, 4.5 million decennial addresses did not have a current link to TIGER. Although these housing units will still be in the census, they risk being inaccurately located. The bureau needs to take steps to ensure that decennial addresses are geocoded accurately. In addition, in developing MAF and TIGER software, the bureau does not follow rigorous software engineering practices and therefore cannot ensure that all results are accurate. Software engineering standards should be used in the planned modernization of the systems that support the MAF geocoding process and overall data quality. (See page 21.)

The bureau should report success in meeting housing unit accuracy and completeness goals. The bureau has created a housing unit coverage performance standard and methods to evaluate if it is met. The goal for Census 2000 is to miss not more than 2.5 percent of existing housing units

and include not more than 1.5 percent in error, for a net undercoverage rate of 1 percent. The Accuracy and Coverage Evaluation will measure housing unit coverage. However, the bureau has not clearly stated how it will report the standard and its success at meeting it. The Annual Performance Plan and Program Performance Report mandated by the Government Performance and Results Act of 1993, in which agencies report performance goals, measures, and accomplishments to the President and Congress, are appropriate vehicles for reporting on this important data quality standard, including its separate overcoverage and undercoverage components. (See page 24.)

For Census 2000, we recommend that the bureau (1) issue a decision memorandum that explains the eligibility policy for the addresses to be included in the final decennial count, (2) ensure that any further TIGER changes made during the decennial are verified with the MAF so that no additional decennial addresses lose their link to TIGER, and (3) report evaluation results measuring housing unit coverage, including its separate overcoverage and undercoverage components. We make additional recommendations for future censuses and surveys designed to improve the accuracy and completeness of the MAF and promote a rigorous software engineering approach to the modernization of MAF and TIGER. Finally, we recommend that the bureau provide housing unit coverage standards and report on its progress toward meeting them in future Government Performance and Results Act reporting. (See page 28.)

In its response to our draft report, the bureau stated that, with one exception, it concurs with or had already acted upon our 11 recommendations. The exception regards a portion of Recommendation 10, implementing an annual national or small area MAF coverage measurement—which the bureau stated would not be cost effective or practical to accomplish annually. We believe other methods, such as comparing tallies of MAF addresses to demographic estimates as employed by Population Division, may serve as an alternative to more costly measurements. Although the bureau concurred with Recommendation 5, to issue a memorandum that explains the address eligibility policy for the final delivery of addresses to be included in Census 2000, we believe that the bureau needs to augment this "general policy" with specific criteria for determining which addresses are eligible for the decennial. Also, in concurring with Recommendation 6, to use information in the MAF as a management tool to increase the completeness and accuracy of the address file, the bureau described many innovations but omitted additional techniques we believe should be used—such as applying address history data to locate geographic areas where addresses are likely to be missing and to pinpoint addresses likely to be geocoded in error. Finally, the bureau provided comments on several aspects of our observations. Based on these comments, we have clarified the appropriate areas of the report. The bureau's response is included in its entirety as Appendix B to this report.

INTRODUCTION

The 2000 Decennial Census enumerates the U.S. population and housing as of April 1, 2000. Reflecting a long tradition, Census 2000 will be the 22nd decennial enumeration in an unbroken chain that our nation has undertaken. The decennial provides information that describes the nation's population within small geographic areas and is used to apportion the U.S. House of Representatives. Decennial data is used to redraw congressional, state, and municipal legislative district lines and provide the basis for determining the distribution of $200 billion of federal funds. The decennial is the only data-gathering operation in the United States that is mandated by the Constitution.

An accurate and credible decennial depends on the Census Bureau's implementing a complex set of data collection operations and data processing systems that must work together to meet data quality standards. This evaluation focuses on the quality of addresses supplied to the decennial from the Master Address File (MAF). The MAF supplies the addresses used to support the operations responsible for mailing and hand-delivering questionnaires, enumerating nonresponding households, and controlling the collection and tabulation of Census 2000 data. The MAF is often referred to as the heart of Census 2000 because it is intended to supply a complete list of living quarters used to identify all households that will be receiving a questionnaire for mail return. Decennial addresses provided by the MAF become the contents of the decennial master address file (DMAF). Responses received from the public are processed and merged with the DMAF to create the file that is the basis of decennial results. Decennial addresses must be precisely located based on geographic location information found in the bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER®) system.[1] The bureau must create a complete list of addresses that correctly identifies the block[2] where the address resides and excludes nonresidential, nonexistent, and duplicate addresses.

The quality of addresses chosen for the decennial is the cornerstone of the decennial's accuracy and completeness and an important determinant of its cost effectiveness. The bureau found that in the 1990 decennial, nearly one person in every three who were missed were not counted because their housing unit was missing from the address file.[3] Also, address data in error increases costs of operations for following up on people who did not respond to the

---

[1]The remainder of this report refers to this system as TIGER.

[2]A *block* is the smallest entity for which the bureau collects decennial information. A block is bounded by physical features and county boundaries.

[3]*Report to Congress – The Plan for Census 2000, Bureau of Census*, August 1997. Stated in percentages: "Based on the1990 PES [Post-Enumeration Survey] results, 69.5 percent of the coverage error came from enumerated housing units and the remaining 30.5 percent came from housing units that were not enumerated at all."

questionnaire delivered to them. Also, the General Accounting Office (GAO) reported in the 1990 decennial that the bureau spent $317 million on operations to identify 4.8 million nonexistent housing units and 8.6 million vacant housing units and to remove the former from the address file. For this decennial, as of June 14, 2000, the bureau has identified 10.2 million nonexistent households, while retaining 9.7 million vacant housing units. Similar to the 1990 decennial, the bureau has implemented a second followup operation to verify nonexistent and vacant units. Costs to verify nonexistent households are not yet known.

PURPOSE AND SCOPE

The overall objectives of this evaluation were to determine the extent to which the bureau had compiled complete and correct housing unit data by the time the data was needed for addressing census questionnaires and to determine the extent to which further corrective actions were needed during the decennial. Our specific evaluation objectives were to (1) determine if steps taken to improve the MAF before delivery of the address file worked as planned, (2) assess how well the bureau identified and addressed MAF data quality problems as the decennial progressed, (3) determine whether the software development approach ensured high quality data, and (4) evaluate whether the data quality standard for the MAF provides a meaningful benchmark for decision-makers.

We chose these objectives because during the 1998 Dress Rehearsal, housing unit data quality did not meet all goals, and the bureau has undertaken several large operations to improve MAF quality, resulting in the need to assess the success of these operations and placing new demands on software processing. We limited our evaluation to bureau-identified households in urban areas that receive mail at their address through the U.S. Postal Service.

To gain a high-level understanding of decennial plans and strategies, we reviewed the *Report to Congress–The Plan for Census 2000,* originally issued July 1997 and revised August 1997. We also reviewed the *Census 2000 Operational Plan Using Traditional Census-Taking Methods,* dated January 1999 and subsequent updates. In addition, we reviewed concerns about the time needed by the bureau to obtain and verify address data provided by local governments raised in our prior report on the Local Update of Census Addresses (LUCA) Program.[4]

To accomplish our first objective, we reviewed the *Census 2000 Address List Reengineering, Case for Change*, dated September 24, 1997, to identify the plan, schedule, and goals for MAF accuracy and completeness. We also reviewed dress rehearsal MAF evaluation reports and supporting documentation to obtain an understanding of issues pertaining to the quality of the

---

[4]*Additional Steps Needed to Improve Local Update of Census Addresses for the 2000 Decennial Census*, IPE-10756, September 1998.

MAF at the time of the dress rehearsal. To understand operations to improve the MAF, we reviewed plans for key improvement operations and bureau documentation about these operations. We also reviewed pertinent Commerce Inspector General and GAO reports.

To determine whether steps taken to improve the MAF before delivery of the address file worked as planned, we reviewed the study conducted by the bureau's Population Division, *Results from the County Level Demographic Benchmark Analysis of the Decennial Master Address File.* The purpose of this study was to examine the accuracy of the housing unit counts based on the DMAF addresses used to deliver questionnaires when compared to the division's independent estimates of the numbers of housing units in each county. To learn more about operations used to build the MAF and TIGER, we interviewed decennial officials, including the Assistant Division Chief, Geographic Operations, and her staff and the Head of the Geographic Planning and Budget Team. We also interviewed quality assurance and evaluation officials in the Decennial Statistical Studies Division (DSSD), Geography Division, and Planning, Research, and Evaluation Division (PRED). Many of those whom we interviewed were also members of the Address List Development Operations Planning Group, which is responsible for resolving decennial address issues as part of its overall charter to design all address list development activities and communicate the operational requirements.

To accomplish our second objective, we evaluated the decision-making process used to identify addresses for inclusion in the decennial. Our analysis of the bureau's address data was initiated by setting up several test cases as a result of identifying potential errors on TIGER maps and analyzing corresponding addresses. For a 3.4 x 3.7-square-mile area of Prince Georges County, Maryland, we compared the TIGER map to a commercial map, visited areas with discrepancies between the two, and noted addresses on streets in question. We then queried the MAF to see if these addresses were in the MAF and marked for inclusion in the decennial. We also reviewed a small sample of addresses that were included in the decennial even though they were listed as nonresidential. After physically verifying errors in data selected for the decennial, we traced these errors to criteria in the specifications used for selecting data for inclusion in operations to prepare for and start the decennial. This information makes it possible to locate potential errors and determine whether they have broader significance.[5] We then reviewed the bureau's plans for resolving these addresses and spoke with officials in the Geography Division and DSSD responsible for this work.

To accomplish our third objective, we interviewed the Assistant Division Chief, Geographic Application Systems, and members of his staff and the Assistant Division Chief, Geoprocessing Systems, and members of his staff. We also interviewed the Branch Chief of the Geographic

---

[5] This field work was confined to a small, local geographic area and was intended only to identify types of errors that can occur, not to provide any statistical quantification of the problem.

Products Quality Assurance Team and other quality assurance staff. We reviewed software specifications and quality assurance practices followed in implementing the software that processes and produces address data for decennial operations, including operations to improve the MAF. We interviewed officials responsible for improving and maintaining the housing unit address data and for providing extracts of this data to support the decennial and other bureau programs, such as the American Community Survey.[6] In the Decennial Systems and Contracts Management Office (DSCMO), we interviewed the program manager responsible for processing the data provided from the MAF for decennial operations and the data obtained from those operations.

To accomplish our fourth objective, we learned about the bureau's efforts to gather, measure, and improve its housing unit data by reviewing memorandums and reports from the 1990 decennial and tests, operations, and studies designed to provide the bureau with information needed to prepare for Census 2000. This documentation included 1998 Dress Rehearsal evaluations of the MAF and reports from the DSSD MAF Quality Improvement Program. We also interviewed bureau officials from the Decennial Management Division, DSSD, and PRED regarding the status of the data quality standard for the MAF and the policy for reporting on progress toward achieving it. We also reviewed the Government Performance and Results Act of 1993 to determine the requirements for reporting performance.

This inspection has been conducted in accordance with the Inspector General Act of 1978, as amended, and the *Quality Standards for Inspections*, March 1993, issued by the President's Council on Integrity and Efficiency. Our field work was conducted from August 1999 through June 2000 at the bureau's headquarters in Suitland and Upper Marlboro, Maryland, and selected areas in Prince Georges County.

BACKGROUND

In conducting the decennial, the bureau attempts to deliver a questionnaire to every household in the country. To accomplish this task, the bureau needs to know the address for each housing unit. The MAF and TIGER are two databases that together contain the nation's addresses and their geographic locations. The MAF database contains both residential and nonresidential (for example, business and religious organizations) addresses. To be valid for the decennial, an address must be for a housing unit where people reside—and, therefore, is typically residential.

---

[6]The American Community Survey is designed to provide the data communities need every year instead of every 10 years. It is an ongoing survey that the bureau plans will replace the long form in the 2010 Census. See the survey web site at http://www.census.gov/acs/www/ for more information.

The bureau estimated, as of July 1998, that there were 112.5 million residential housing units through use of independent demographic analysis. A year later, approximately 120 million residential addresses in the MAF were identified for use in the decennial. The number of MAF residential addresses was higher than the 1998 estimates, in part, because the MAF contained duplicate and nonexistent addresses. Residential addresses are divided into two basic address styles, city-style and non-city-style. A city-style address consists of a house number and street name—101 Main Street, for example, with optional apartment number and direction such as "North." A non-city-style address may have a delivery route number or box number. Residential addresses comprise approximately 80 percent city-style addresses and 20 percent non-city-style addresses.

In areas where city-style addresses predominate, the TIGER database contains streets with associated address ranges; in areas where non-city-style addresses predominate, TIGER contains roads and map spots to indicate the location of housing units. In TIGER, the entire country is divided into discrete geographic areas called blocks. Each of the country's 3,142 counties has its own set of uniquely numbered blocks. To be valid for the decennial, a MAF address must be linked to exactly one block in TIGER. The process of linking or assigning addresses to blocks is called *geocoding*.

A third database, the Decennial Master Address File, incorporates address data from the MAF with control data for tracking questionnaire responses submitted by the public. DMAF data links the public's responses to each housing unit to calculate housing unit participation in the decennial. If an address is not in the DMAF at the start of the decennial, it must be added during decennial operations to be included in the decennial results. These new addresses will also have to be added to the MAF and geocoded. Similarly, decennial operations identify addresses that should not be in the decennial either because they are duplicates, because no such housing unit exists, or for other reasons. Such addresses are not physically deleted from the MAF or DMAF but are flagged so that they are not included in further decennial operations or in the decennial results. The DMAF is also used to provide address lists for operations such as Nonresponse Followup (NRFU) in which enumerators attempt to find nonrespondents and Coverage Improvement Followup (CIFU), in which enumerators, among other things, attempt to find and enumerate households not included in NRFU.

MAF, TIGER, and DMAF are separate databases that contain address data. They are each updated as new address data is provided through decennial operations. Because data from all three is used to implement and manage decennial operations, consistency between them is essential to producing accurate results for Census 2000.

A critical milestone for use of housing unit data in the decennial occurred in July and August 1999, when the initial DMAF was created with address data extracted from the MAF. The July

1999 and supplemental August 1999 MAF extract for the DMAF provided the initial universe of housing units for the decennial. This address data was used to label questionnaires for eventual mail and hand delivery to the nation's housing units in time for Census Day— April 1, 2000.

Strategies to Build the MAF

In 1992, the Congress mandated a study of the fundamental requirements for the nation's decennial census by the National Academy of Sciences' National Research Council. The council recommended that the bureau develop cooperative arrangements with local/tribal governments to improve its address data. Section 9 of Title 13 authorizes the bureau to protect the confidentiality of the persons from whom it collects data, and in general prohibits the sharing of any individual's data collected by the bureau. To allow the bureau to implement the Council's recommendation, the Congress passed the Census Address List Improvement Act of 1994, which amended Title 13 to require the bureau to solicit address list feedback from local/tribal governments. This act also mandated that the bureau use the U.S. Postal Service's Delivery Sequence File, a nationwide list of individual mail delivery points.

The Initial Plan

In September 1997, the bureau began to build the MAF for the entire nation by combining the 1990 decennial address list with the Postal Service file for counties with city-style addresses. After merging the two files, all addresses were submitted to automated processing that geocoded them to TIGER. Some addresses could not be geocoded by automated means because TIGER data did not include the house number within the address range for that street or the street was missing. These addresses were sent to an operation called the MAF Geocoding Office Resolution, which researched the correct information for the locale and updated TIGER so the address could be geocoded. To incorporate address changes or additions reflected by the Postal Service data, the bureau has updated the MAF periodically with later versions of the Postal Service file. The geocoding process occurs as part of each such update.

To obtain local and tribal government input, the bureau created the Program for Address List Supplementation. Through this program, the bureau invited local governments to submit lists of city-style addresses to obtain any that still might be missing or listed incorrectly. The bureau planned field operations to verify discrepancies between existing MAF data and that provided by the program. The bureau also planned targeted canvassing to identify housing units whose addresses were missing from the MAF. Finally, a Local Update of Census Addresses (LUCA) program to be started about a year before creating the initial DMAF would provide an opportunity for participating local/tribal governments to review the address lists and provide updates and corrections.

The Reengineered Plan

Even before the merging operation that created the initial MAF, the bureau realized that its current plan would result in an address list with rates of overcoverage (erroneous addresses) and undercoverage (missed housing units) that would burden decennial operations. The bureau formed a team with members from various divisions to develop an approach to achieving the most accurate list possible, using local and tribal government involvement. Time to implement the plan was short—within a little over two years, the bureau had to supply addresses to the printers to label questionnaires. The bureau documented its new plan in September 1997.[7]

For city-style addresses, the new approach allowed more time for local governments to review MAF address lists and submit additions, deletions, or other corrections in a program now called LUCA 98. This new program would assist governments that do not maintain address data as mailing lists. To verify address data provided by local governments and account for deficiencies in the Postal Service file, the bureau changed its plan from targeted canvassing to 100 percent block canvassing by bureau employees, an operation considered critical to achieving a database with uniformly high quality. For the block canvassing operation, listers were to canvass 100 percent of their assigned area and conduct brief interviews at approximately every third housing unit, every multi-unit structure, and all added housing units.[8] Block canvassing was called a "dependent listing" because listers used lists of addresses that the bureau generated from the MAF for their specific assignment area. The listers were to compare each address with those on the list, mark correct ones as verified, and record all additions, deletions, and corrections. In addition, the listers were to update TIGER maps. The bureau believed that block canvassing was the only method that could identify and correct all types of accuracy problems. A separate reconciliation operation was planned for LUCA 98 results that differed from bureau results.

To ensure the completeness of city-style addresses, the bureau also planned a Postal Service validation, which would consist of the postal carriers placing pre-addressed cards in their mail sorting cases to identify either undeliverable or missing addresses. The bureau believed that the Postal Service validation would update the MAF with new residential construction occurring in 1999 and early 2000 that was missed by other operations. The check was to be conducted as close to Census Day as possible.

---

[7]*Census 2000 Address List Reengineering, Case for Change*, September 24, 1997.

[8]*Census 2000 Block Canvassing Program Master Plan*, Bureau of Census, March 1999.

Changes to the Reengineered Plan

The bureau subsequently modified some aspects of the reengineered plan. It replaced the Postal Service validation with a letter carrier review and correction of addresses in the summer of 1999 and again in January 2000.[9] The bureau made this decision after months of evaluation and discussions with the Postal Service. The new plan allowed earlier incorporation of new addresses and reflected actions taken by the Postal Service to increase the currency and accuracy of the Postal Service file. The bureau also added a program to allow local/tribal governments to identify newly constructed housing units starting in January 2000. This program responded to concerns raised by local/tribal governments that housing units constructed between January 2000 and Census Day would not be included the decennial.

Bureau Tests and Evaluations Support the Decision to Revise the MAF-building Approach

Bureau evaluations and tests conducted before and after the decision to alter the strategy supported the need to improve MAF accuracy and completeness. The reengineering plan cited data quality goals for the MAF that called for overcoverage of 1.5 percent and undercoverage of 2.5 percent, for a net undercoverage of 1 percent.[10] However, according to this plan, the bureau's experience with using the combined Postal Service file and 1990 address list during the 1995 Census Test revealed that the bureau was not meeting its goals. According to the reengineering plan, the 1995 test of two cities showed undercoverage ranging from 3.9 to 6.7 percent and overcoverage ranging from 6.0 to 9.0 percent. Further, the data indicated that coverage problems were worse in multi-unit structures, where undercoverage ranged from 4.0 to 7.9 percent and overcoverage ranged from 5.7 to 9.4 percent, depending on the size of the structure.

Two other evaluations conducted by PRED and DSSD during 1997 and 1998 also found coverage problems.[11] The 1997 MAF Quality Improvement Program Pilot Study, which

---

[9]In addition, the bureau used updated Postal Service data to update the MAF in February 2000.

[10]The reengineering plan cited no more than 1 percent duplicates as part of the overcoverage measurement goal. We estimated the total overcoverage measurement goal, including duplicates and other nonexistent housing units, as 1.5 percent based on the bureau's goal for missing housing units of 2.5 percent and net housing unit undercoverage measurement of 1 percent.

[11] *1997 Master Address File Quality Improvement Program Pilot Study*, Bureau of the Census, PRED, April 1999 and *1998 Master Address File Quality Improvement Program*, Bureau of the Census, PRED, June 1999. The purpose of the 1997 pilot study was to test the operational feasibility of using the Integrated Coverage Measurement (ICM) methodology to measure the accuracy and completeness of the initial MAF. The pilot concluded that with a few modifications, the ICM operational methodology worked for the 1997 Quality Improvement Program.

evaluated six counties with high rates of geocoded addresses, found undercoverage rates ranging from 5.1 percent to 26.9 percent and overcoverage rates ranging from 6.7 percent to 18.2 percent.[12]

The 1998 MAF Quality Improvement Program Study also found large error rates and concluded that they confirmed the need for significant improvement in MAF-building operations before the decennial.[13] The 1998 study included estimates of geocoding errors to account for housing units missing from a block because they were assigned to the wrong block. At the national level, the study found a 9-percent undercoverage rate and a 13-percent overcoverage rate. The study also found that 6 percent of addresses were geocoded to the wrong block, and 6 percent of existing housing units with addresses in the MAF were not geocoded. At a regional level, undercoverage ranged from 5 to 16 percent and overcoverage ranged from 8.5 to 16 percent. From 2.5 to 11 percent of addresses were geocoded to the wrong block, and from 2 to 12 percent of existing housing units with addresses in the MAF were not geocoded. In the six counties that were evaluated, undercoverage estimates range from 3 to 7 percent and overcoverage estimates range from 7 to 36 percent. From 2 to 7.5 percent of addresses were geocoded to the wrong block, and from 0.1 to 1.5 percent of existing housing units with addresses in the MAF were not geocoded.

The bureau also evaluated MAF housing unit undercoverage and overcoverage experienced during the 1998 Dress Rehearsal.[14] Undercoverage estimates at the dress rehearsal sites ranged from 2.9 to 24.6 percent, and overcoverage estimates ranged from 2.9 to 14.8 percent.

Another type of dress rehearsal evaluation of the DMAF was done by the Population Division to examine the consistency of housing totals. This evaluation compared the numbers of addresses in the DMAF for a specific county to an independent demographic benchmark calculated by using the results of the 1990 decennial, then adding the number of new housing units and subtracting the number demolished. DMAF tallies differed widely from demographic benchmarks in South Carolina where the DMAF did not retain housing units with mailing addresses that were post office boxes and did not adequately obtain address data for newly constructed housing units. In areas not experiencing these problems, DMAF address totals were

---

[12]Because this study only used residential geocoded addresses, it did not measure the extent to which coverage errors were caused by coding errors. These types of errors include (1) geocoding errors, an address coded to the wrong block, which erroneously decreases housing units on one block while increasing them on others; (2) ungeocodables, addresses on the MAF but not geocoded; and (3) nonresidential coding errors, addresses that are incorrectly coded nonresidential. This study also did not specifically look at coverage in multi-unit structures.

[13]*1998 Master Address File Quality Improvement Program*

[14]*Results of the Housing Unit Matching Phase of the Integrated Coverage Measurement*, Bureau of Census, DSSD ICM Dress Rehearsal Results Memorandum Series Number HU-1, September, 1998.

broadly consistent with the independent demographic benchmarks. These results indicate that benchmarks provide a useful tool for evaluating the MAF and DMAF.[15]

Decennial Operations Rely on and Improve the MAF, DMAF, and TIGER

During the decennial, response collection operations use address data and can add to the accuracy and completeness of the MAF, TIGER, and the DMAF.   In Nonresponse Followup—a decennial operation that occurs after the initial period when responses are returned from the public—temporary field staff, called enumerators, visit housing units with addresses for which the bureau has not recorded a response. The bureau provides enumerators with lists of information from the DMAF for both respondents and nonrespondents. Enumerators use the lists to find nonrespondents and to help them identify whether respondents may have completed a form for the wrong address and if so, to locate the true nonrespondents. Enumerators then attempt to interview a household member to obtain resident and housing unit information.

The Geography Division is responsible for the data processing that identifies the addresses eligible for NRFU, and DSCMO is responsible for the data processing that generates the NRFU universe of nonrespondents and the address lists containing the data to produce enumerator work lists. This processing occurred during March and April 2000. In addition to finding nonrespondents, enumerators also identify vacant housing units as *vacant* and nonexistent addresses as *deletes.* To the extent that nonexistent addresses are identified and flagged on the MAF and the DMAF, overcoverage is decreased.

Although enumerators focus on obtaining missing responses rather than looking for missing units, when they do discover the latter, they record pertinent data in the address register and attempt to complete a questionnaire for the household. If the address recorded on the completed questionnaire can be geocoded and is not already in the MAF, it is added. To the extent that NRFU adds valid addresses to the MAF, undercoverage is decreased.

During Coverage Improvement Followup—an operation that followed NRFU—the bureau enumerates new housing units found during update/leave, housing units associated with lost or blank questionnaires, partially completed Be Counted and Telephone Questionnaire Assistance questionnaires, and new addresses or incomplete responses that were obtained too late to be in NRFU.[16] CIFU also verifies some housing units classified as vacant or nonexistent in earlier decennial operations. Similar to NRFU, Geography identified addresses eligible for CIFU, and DSCMO produced the CIFU universe. This processing occurred during June 2000. Similar to

---

[15]*Census 2000 Dress Rehearsal Evaluation Summary*, Bureau of Census, PRED, August 1999.

[16]*Coverage Improvement Followup Program Master Plan*, Bureau of Census, October 1999.
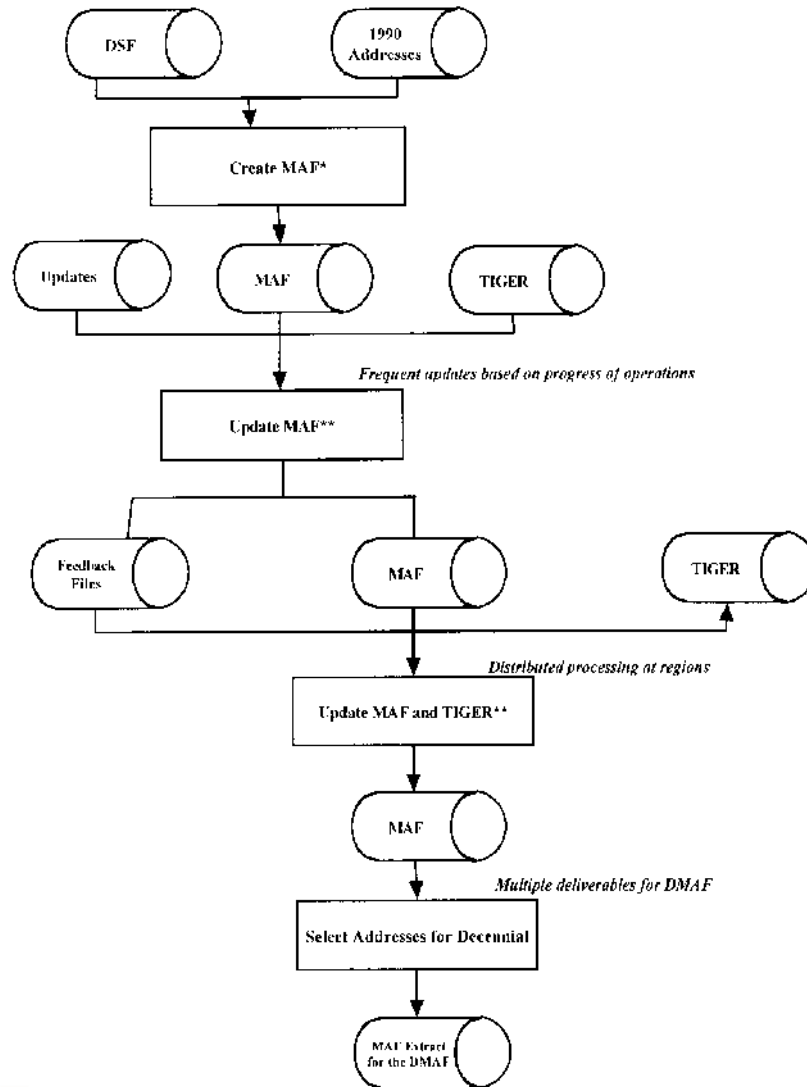
NRFU, enumerators identify vacant housing units as *vacant* and nonexistent housing units as *deletes.* To the extent that nonexistent addresses are identified and flagged on the MAF and the DMAF, overcoverage is decreased. If a missing housing unit is found, the enumerator will add the address and attempt to enumerate the household. To the extent that CIFU adds valid addresses to the MAF, undercoverage is decreased.

Two operations, the Be Counted Program and Telephone Questionnaire Assistance, although not designed specifically to identify housing units, can result in new addresses. If a housing unit or households within a housing unit have not received a questionnaire from the bureau, they can obtain an unaddressed Be Counted questionnaire. Similarly, a household can call a toll-free number to submit a response regardless of whether they received a questionnaire in the mail. In both programs, if the address provided with the response can be geocoded and is not already in the MAF, it will be added. Although the intent of these operations is to count people within housing units who are not included in the return for that unit, to the extent that valid addresses are added to the MAF, housing unit undercoverage is decreased.

Computer Systems Process Information for Decennial Operations

The bureau has developed software to extract information from MAF and TIGER to update the DMAF and to process data obtained from decennial operations to update MAF and TIGER for further decennial operations. Software was used to merge the Postal Service file with the 1990 decennial address list to produce the initial MAF; reconcile TIGER with the MAF; produce address lists for LUCA 98, block canvassing, and LUCA 98 Field Verification; merge data collected from LUCA 98, block canvassing, and Postal Service file updates with MAF and TIGER; and produce MAF extracts for creating and updating the DMAF. MAF extracts are planned for many DMAF updates during the decennial. Final housing unit selection from the MAF is delivered to headquarters processing for merging with the response data. The results of that processing are submitted for the selection of housing units to be included in decennial results. Figure 1 diagrams these components. Critical system subcomponents include the matching and merging software and the geocoding software. The matching and merging software updates existing address records or adds new address records with data obtained from decennial operations. The geocoding software links an address in the MAF with a unique block (geographic location) in TIGER.

Figure 1. System Components for MAF and TIGER®

DSF

1990 Addresses

Create MAF*

Updates

MAF

TIGER

Frequent updates based on progress of operations

Update MAF**

Feedback Files

MAF

TIGER

Distributed processing at regions

Update MAF and TIGER**

MAF

Multiple deliverables for DMAF

Select Addresses for Decennial

MAF Extract for the DMAF

Source: Derived from U.S. Census Bureau flowcharts

* Employs the matching and merging software
** Employs the matching, merging and geocoding software

OBSERVATIONS AND CONCLUSIONS

I. Sufficient Time Not Available to Ensure High Quality Address Data

Although the bureau made a strong case for reengineering its approach to building the MAF and spent close to $100 million for the necessary operations, making that decision just two years before the addresses were needed did not leave enough time to carry out the new plan. A bureau study of address quality in the initial DMAF raised concerns about high levels of housing unit overcoverage and undercoverage but was not designed to report on the cause. We found that LUCA 98 schedule slips and bureau policy regarding which addresses were eligible to be included in block canvassing reduced the effectiveness of this major operation designed to improve address quality. In addition, the bureau did not have sufficient time to resolve conflicting LUCA 98 and block canvassing addresses before including these and other unresolved addresses on the initial DMAF for questionnaire addressing. Housing unit overcoverage, caused by including over 5 million unresolved addresses on the DMAF, increased the burden on decennial operations to resolve nonexistent and duplicate addresses.

The bureau created the MAF by updating the 1990 address file with Postal Service and local/tribal address information and plans to update and use the MAF resulting from the 2000 decennial as the basis for future surveys and censuses, including the 2010 decennial. Using its experiences from the 2000 decennial as a guide, the bureau needs to develop an improved approach for updating the MAF that includes sufficient time to conduct MAF-building operations designed to detect and resolve overcoverage and undercoverage and that incorporates a clear definition of addresses eligible for these operations.

A.   High Levels of Data Quality Not Achieved for Initial DMAF

The Population Division compared its 1998 estimates of housing unit coverage to numbers of addresses in the initial DMAF to find potential undercoverage and overcoverage problems. That comparison showed that a year after it had estimated a total of 112.5 million housing units nationwide, the initial DMAF contained 120.2 million units—a difference of 7.7 million or 7 percent. Knowing that the national average can vary considerably at a county level and wanting to provide a tool for flagging counties with potential overcoverage and undercoverage, analysts compared the benchmarks with the total contained in the initial DMAF for each county. The

methodology separated counties by type of enumeration area[17] and tallied the number of addresses for each county. The analysis found that of 148 counties with only city-style addresses, 57 percent had measurable overcoverage or undercoverage.[18] Similarly, of 1,499 counties with a combination of city-style and non-city style addresses, 67 percent had measurable overcoverage or undercoverage.[19] As noted previously, housing unit coverage will improve during the decennial—bureau operations will add housing units and eliminate nonexistent and duplicate housing units. In addition, housing unit estimates may not adequately reflect fluctuations in population for some counties. However, such a high percentage of counties with coverage discrepancies indicates significant data quality problems at the start of the decennial.

For the purpose of the Population Division's analysis, counties were considered to have undercoverage if the DMAF count was below the 1998 estimate. In these counties, the housing unit count is expected to fall even more after the normal process of removing duplicate and nonexistent addresses. Unless update operations add housing units to make up the difference, there will be a shortfall. The study identifies counties where the DMAF is 0 to 5 percent higher than the 1998 estimate as also having the potential to fall below the 1998 estimate after duplicate and nonexistent housing units are removed during the decennial.

For the same analysis, counties were considered to have overcoverage if the DMAF count was above the 1998 estimate by 20 percent. In these counties, the bureau expects a larger than normal

---

[17]Types of Enumeration Areas (TEAs) include TEA 1- block canvassing and Mailout/Mailback; TEA 2 - Address Listing and Update/Leave; TEA 3 - List/Enumerate; TEA 4 - Remote Alaska; TEA 5 - "Rural" Update/ Enumerate; TEA 6 - Military; TEA 7 - "Urban" Update/Leave; TEA 8 - "Urban" Update/Enumerate; and TEA 9 - Additions to Address Listing Universe of Blocks. For further discussion of operations for each TEA, see *Census 2000 Operational Plan Using Traditional Census-taking Methods*, Bureau of the Census, January 1999.

[18]Nationwide, there are 3,142 counties. In addition to the 148 city-style and 1499 city-style and non-city-style counties cited above, the study included 818 counties that had all non-city-style addresses and 460 counties that had a combination of city-style, non-city-style, and other TEAs. The report also did not cover another 217 counties, primarily because these counties' addresses were obtained through other than MAF-generated activities. We included the bureau's analysis of counties with only TEA 1 and combination of TEAs 1and 2. We did not include the results for TEA 1 in combination with other TEAs. For example, the results from Cook County, Illinois and Philadelphia County, Pennsylvania were not included because these counties are a combination of TEAs 1 and 7.

[19]The study results are documented in Count Review Memorandum Series 99-01, Subject: *Results from the County Level Demographic Benchmark Analysis of the Decennial Master Address File–Part A: Differences 5 Percent or below for Selected Types of Enumeration Areas, January 10, 2000,* and Count Review Memorandum Series 99-02, Subject: *Results from the County Level Demographic Benchmark Analysis of the Decennial Master Address File–Part B: Differences in Excess of 10 Percent for Selected Types of Enumeration Areas,* February 10, 2000, issued by the bureau's Population Analysis and Evaluation Staff, Population Division.

number of deleted addresses resulting from removing duplicate and nonexistent addresses. The study also identifies counties where the DMAF is 10 to 20 percent higher that the 1998 estimate as counties with a similar potential for a higher number of deleted addresses resulting from removing duplicate and nonexistent housing units. Table 1 shows the results of the analysis, including the total number of counties in each category and overcoverage and undercoverage combined.

Table 1: Demographic Evaluation of Initial DMAF Undercoverage and Overcoverage by Counties Having City-Style and Combination City-Style and Non-City-Style Addresses

Type of Address and Number of Counties	Counties with Overcoverage		Counties with Undercoverage		Total Undercoverage and Overcoverage
	Exceeds 1998 by 20% or higher	Exceeds 1998 by 10 to 20%	Under 1998 Estimate	Exceeds 1998 estimate by 0-5%	
City-Style—148 counties	10 (7%)	27 (18%)	6 (4%)	41 (28%)	84 (57%)
City-Style and Non-City-Style—1,499 counties	37 (3%)	264 (18%)	163 (11%)	543 (36%)	1007 (67%)*

*Percentage total difference due to rounding.

The results for the counties cited above raise questions about the success of reengineering in meeting overcoverage and undercoverage goals. Reengineering in some counties appears to have resulted in undercoverage rates comparable to those experienced at the time of the dress rehearsal. Given that the 100 percent block canvassing component of the reengineering was intended to improve the MAF beyond levels experienced during the dress rehearsal, we are recommending that the bureau explore what caused overcoverage and undercoverage and use the results as lessons learned when planning future MAF improvement operations.

*B.   LUCA 98 Schedule Slips and Bureau Policy Reduced the Effectiveness of
         Block Canvassing*

The bureau designed block canvassing so that it would include all LUCA 98 addresses and verify both LUCA 98 and Postal Service address data. Prior bureau evaluations indicate that erroneous addresses are more likely to be corrected by field operations if they are included on the address

lists used during the operation.[20]   However, not all LUCA 98 or MAF addresses were included in block canvassing. Thus, block canvassing enumerators did not have complete housing unit address lists, making it difficult to find missing addresses and areas not covered effectively by the Postal Service. As bureau officials have explained to us, although block canvassing is designed to be a 100-percent verification of assigned areas, listers often rely too heavily on address lists and associated TIGER maps and do not find areas where addresses are not listed or where streets are missing on the maps. Not including all MAF addresses in block canvassing resulted in unverified addresses being included on the initial DMAF.

We found that block canvassing lists were incomplete for two reasons: (1) the LUCA 98 operation took longer than the bureau estimated, resulting in approximately 98.6 percent of LUCA 98 addresses that were not available in time to be included in block canvassing, and (2) block canvassing address lists did not include some MAF addresses from the Postal Service file, as well as other addresses that were not geocoded.[21] Specifically, MAF addresses added by the November 1997 Postal Service file but not included on the September 1998 Postal Service file were omitted from block canvassing address lists. Bureau officials told us that there was not a clear policy on which addresses should be included for block canvassing, and at that time, a decision was made not to use these addresses, which the Postal Service no longer considered valid. However, this decision was not consistent with the later decision to include these same addresses in the initial DMAF. Addresses with missing geocodes were not included on block canvassing lists because bureau officials believed that addresses without block numbers would be difficult to find. However, since other identifying information—such as the street address, city, state, and zip code—was available, we believe that these addresses could also have been verified.

We are recommending that in establishing its strategy for updating the MAF for future surveys and censuses, the bureau ensure that sufficient time is planned for MAF improvement operations. We are further recommending that this strategy include developing a consistent policy for address eligibility for these improvement operations and an approach for verifying addresses without block codes.

---

[20]As cited in *Additional Steps Needed to Improve Local Update of Census Addresses for the 2000 Decennial Census,* IPE-10756, September 1998.

[21]*2000 Census, Local Address Review Program Has Had Mixed Results to Date,* GAO Testimony, Statement of J. Christopher Mihm, Associate Director, Federal Management and Workforce Issues, General Government Division, GAO/T-GGD-99-184, September 29, 1999 and Census Bureau Geography Division data spreadsheet.

C.   Insufficient Time Also Contributed to Erroneous Addresses on the Initial DMAF

In order to minimize the chances of omitting valid addresses from the decennial, the bureau included addresses that MAF operations had indicated were in error or that had conflicting indications from different operations on the initial DMAF. This inclusive approach allowed 5.2 million unresolved addresses to be delivered in July 1999 to the printing operation that labeled questionnaires for mail delivery. Unresolved addresses were those that block canvassing recommended be deleted because they were nonexistent, uninhabitable, or nonresidential or that had been added by LUCA 98 but not by block canvassing. The bureau decided that unresolved addresses would be reviewed under a LUCA 98 Field Verification operation.[22] However, this operation was not completed until several months after the address file was delivered to the printer. We believe that this approach contributed to the high amount of overcoverage reported in the Population Division study discussed in the first part of this observation.

The bureau's inclusive approach also resulted in including an unknown number of duplicate addresses in the decennial. We found two reasons why duplicates occurred. First, some addresses identified as duplicates by block canvassing were included because the results conflicted with LUCA 98. Second, a decision to alter software processing so that address updates of individual apartments, trailer park lots, and certain other housing unit types would not be incorrectly merged into a single address allowed duplicates of other addresses to occur.[23]

The high level of undercoverage and overcoverage that occurred due to the lack of time to verify addresses underscores the need for the bureau to ensure that sufficient time is allotted for MAF improvement operations in the future. As the bureau continues to update the MAF with Postal Service or local/tribal address information, preventing duplicate addresses will remain a challenge. Therefore, we are recommending that the bureau study the causes of duplicate addresses and implement methods to prevent them.

---

[22]LUCA 98 addresses that were not accepted by the bureau during block canvassing were to be rechecked during a subsequent inspection operation called reconciliation, where bureau employees were to verify the block canvass results between March and August 1999. However, the delays in LUCA 98 created the complication of some LUCA 98 addresses not included in block canvassing having to be matched to block canvassing results to determine their status. This process and other delays resulted in the delay of the reconciliation to verify LUCA 98 addresses.

[23]*Block Canvassing Address Merge Rules*, Memorandum from Robert Marx, Chief, Geography Division to Distribution List, June 25, 1999.

**II.   The Bureau Has Taken Steps to Improve Address Data Quality,
          but More Should Be Done**

The bureau took steps to identify erroneous addresses before starting NRFU. However, similar
to the bureau's lack of a clear policy on address eligibility for MAF-building operations, its
policy for defining address eligibility during the decennial is also not well-defined. In
particular, the specification that determines which housing units will be included in the final
decennial count has not been issued. The bureau could also make better use of information
already on the MAF to identify missing housing units and potential geocoding errors where a
MAF address does not link to the correct TIGER location. To improve address data quality for
Census 2000 and future censuses and surveys, the bureau should (1) define a clear and visible
policy for determining address eligibility during the decennial and (2) devise methods of using
MAF address history information to guide coverage improvement operations.

*A.   The Bureau Identified Some Erroneous Addresses and
        Flagged Them As Ineligible for Nonresponse Followup*

In our review of a sample of addresses in Prince Georges County, Maryland, we found evidence
of systemic overcoverage that we believe need to be quantified and targeted by the bureau. The
overcoverage included both duplicate addresses and addresses for nonexistent or uninhabitable
housing units.

For example, we found a street with 28 townhouses, of which 26 had their addresses listed twice
in the MAF. Initially, addresses for the 28 townhouses were added to the MAF with errors
caused by a slightly incorrect spelling obtained from the Postal Service file and an incorrect
block number caused by an error in TIGER. LUCA 98 added correct addresses and block
numbers for 26 of these townhouses. All 54 addresses were included in the initial DMAF.
However, as we pointed out earlier, the bureau's software had been modified to merge fewer
duplicates and thus allowed more of these duplicate addresses in the initial DMAF.

As we also pointed out earlier, another reason for overcoverage was the lack of a clear policy on
which addresses should be included for block canvassing. For example, we found an apartment
complex of some 500 uninhabitable units whose MAF addresses were excluded from block
canvassing lists but included in the initial DMAF. The September 1998 Postal Service file did
not contain these addresses, disqualifying them from block canvassing.[24] However, the

---

[24]*Master Address File Extract for Block Canvassing,* Bureau of Census memorandum, October 7, 1998.

November 1997 Postal Service file had designated these units as residential, qualifying them for the initial DMAF.[25]

The bureau agrees with our assessment that known duplicate addresses and addresses not verified by block canvassing and added solely on the basis of the November 1997 Postal Service file are likely to be erroneous. The bureau has acted to remove these addresses which include 0.5 million duplicate addresses, 0.9 million obsolete postal service addresses, 1.5 million addresses known as LUCA 98 provisional adds (out of 2.6 million), as well as 1.4 million addresses recommended for deletion by block canvassing.[26] We believe the bureau made the right decision in taking steps to not include these addresses in nonresponse followup, which saved the cost of enumerator visits and enhanced the enumerators' ability to obtain actual responses. Followup operations are finding additional nonexistent addresses. For example, as of June 14, 2000, the bureau has reported that its nonresponse followup operation found 5.9 million nonexistent addresses. The final number of nonexistent housing units will be determined after completion of coverage improvement followup.

B.   *The Bureau's Policy for Determining Address Eligibility*
       During the Census Is Not Well-Defined

Although the bureau has identified almost 20 million nonexistent and vacant addresses, 15.6 million in nonresponse followup plus the 4.3 million addresses before nonresponse followup, the decision about the criteria for which addresses to include in the final decennial count has not been made. The bureau does not have a written policy for determining address eligibility. Bureau officials cite a "double-strike" or "double-kill" policy, whereby addresses have to be deleted by two operations to be ineligible for remaining decennial operations. However, this policy has not been documented in the decision memorandum series that establishes and communicates bureau policy for the decennial. Instead, the policy is embedded in DMAF deliverability criteria found in software specifications, which often are not finalized until data processing for the operation is imminent or underway. We found that the actual policy implicit in DMAF specifications includes more than the double-strike rule described by decennial managers. For example, one set of DMAF specifications included criteria for deleting addresses not found on Postal Service files delivered after September 1998 that are not part of the double-

---

[25]*Specification of the Decennial Master Address File Deliverability Criteria for Census 2000*, DSSD Census 2000 Procedures and Operations Memorandum Series #D-1, Bureau of Census, June 30, 1999.

[26]Some overlap between the duplicates and the block canvass deletes and between the obsolete addresses and the LUCA provisional adds is possible.

strike policy.[27] In another set of DMAF specifications, addresses said to be seasonal or recreational only needed one strike to be ineligible for CIFU operations. In another case in this same specification, an address designated by block canvassing as nonresidential and as a delete by NRFU would still be included in CIFU because the block canvassing delete no longer qualified as a strike.[28] Finally, we observed address eligibility decisions being made up to the deadline for delivering NRFU-eligible addresses, and the bureau has yet to complete the DMAF specification for addresses eligible for inclusion in the final decennial results.

For future censuses and surveys, including the 2010 decennial, we are recommending that the bureau develop an address eligibility policy in advance that defines and explains the rationale to be used in selecting addresses. For the current decennial, we are recommending that the bureau issue a decision memorandum that explains the address eligibility policy for the final list of addresses to be included in the decennial.

C.   The Bureau Should Use the MAF as a Management Tool

Data on the MAF provides valuable information for increasing the accuracy and completeness of addresses. For example, we found a street with approximately 140 townhouses built since 1990 whose addresses were added by the Postal Service file, verified by LUCA 98, but not indicated on the MAF as verified by block canvassing. Especially where no LUCA 98 participation occurred, areas not block canvassed may have missing or inaccurate addresses. The bureau should identify areas with high percentages of addresses with no block canvassing action codes for future verification. As of April 2000, there were 9.5 million of these addresses in the DMAF. Possible reasons for their inclusion are that some were located by other operations or the housing units were built recently. However, areas with a preponderance of no block canvassing action codes also could have missing housing units.

---

[27]*Definition of the Nonresponse Followup Universe for Census 2000*, DSSD Census 2000 Procedures and Operations Memorandum Series #BB-3, Bureau of Census, November 1, 1999; *Specification for Identifying the Nonresponse Followup Universe for Census 2000*, DSSD Census 2000 Procedures and Operations Memorandum #BB-4R, Bureau of Census, February 2, 2000; and *Specifications for Updating the Decennial Master Address File in April 2000*, DSSD Census 2000 Procedures and Operations Memorandum #D-6, Bureau of Census, March 27, 2000.

[28]*Definition of the Coverage Improvement Followup Universe for Census 2000*, DSSD Census 2000 Procedures and Operations Memorandum Series #CC-2 Revision #4, Bureau of Census, June 28, 2000; *Revision Specification of the Coverage Improvement Followup Universe for Census 2000*, DSSD Census 2000 Procedures and Operations Memorandum Series #CC-3, Revised #4, Bureau of Census, June 28, 2000; and *Specification for Updating the Decennial Master Address File on June 15, 2000*, DSSD Census 2000 Procedures and Operations Memorandum Series #D-7 Revised, Bureau of Census, May 22, 2000.

The bureau could also use MAF data to identify addresses with incorrect geocodes. For example, we found two houses on an unpaved road not easily visible that were geocoded to the wrong block. These addresses were included in LUCA 98 and block canvassing. LUCA 98 verified them, but block canvassing deleted them, perhaps because they could not be found in the block to which they were assigned. Subsequently, LUCA 98 Field Verification verified that they existed. No operation corrected the block number, and the units were delivered to the DMAF assigned to the wrong block. As of April 2000, there were over one million addresses in the DMAF where LUCA 98 Field Verification verified addresses that block canvassing indicated should be deleted. We believe addresses that LUCA 98 Field Verification verified but block canvassing recommended for deletion may be housing units that are assigned to the wrong block.

While it is too late to carry out additional field operations to take advantage of MAF information for this decennial, we are recommending that in the future, the bureau maximize use of MAF information to identify areas where addresses are more likely to be missed or incorrectly geocoded.

III.   Improved Software Engineering Standards Could Improve Data Quality

Addresses in the MAF must be consistent with TIGER to enable these addresses to be geocoded to geographic locations accurately. However, because MAF and TIGER were developed separately and are not integrated, consistency between them cannot be easily maintained. TIGER data can be modified without ensuring that both databases have accurate and consistent information, causing some decennial addresses to no longer link to TIGER. In addition, some decennial addresses received a geocode from block canvassing but do not link to TIGER because either the geocode or the TIGER data was incorrect. As of April 2000, 4.5 million decennial addresses did not have a current link to TIGER. Although these housing units will still be in the census, they risk being inaccurately located. The bureau needs to take steps to ensure that decennial addresses are geocoded accurately. In the future, the bureau intends to modernize MAF and TIGER to fix consistency issues, among other things. To make the modernization a success, the bureau needs to adhere to its software engineering standards. These standards will guide analysts and developers toward building accurate, well-understood system components that work together to safeguard data consistency.

A.   Consistency Between MAF and TIGER Not Easily Maintained

Software systems build the links between the MAF and TIGER databases either automatically or by allowing bureau staff to access TIGER on-line and add or modify TIGER data so that each

address links to a location and can be geocoded.[29]   Inconsistencies occur when the name of a street or the range of house numbers associated with a street is modified or a street is deleted in TIGER. Such changes cause addresses in the MAF that are already linked to that street to no longer geocode. For example, in our field verification we found a street that was incorrectly listed twice in TIGER. An apparent attempt to correct this problem resulted in the name of another nearby street being changed to the same name as the street incorrectly listed twice. This action created three streets in TIGER with the same name: one real street, one fictitious street, and one incorrectly named street. The incorrectly named street had housing unit addresses on the MAF and DMAF for which questionnaires had been delivered. The bureau official assisting us with the TIGER queries explained that because MAF and TIGER are not integrated, changes made in TIGER can cause MAF addresses to no longer link to a TIGER block.

The addresses described above are typical of addresses that will not geocode to TIGER during production runs that occur during periodic Postal Service file updates. These addresses need to be sent to the clerical part of the process, which then has to correct TIGER so that these addresses will once again geocode. During the April 2000 update of the MAF, 4.5 million decennial addresses did not geocode to TIGER.

As bureau officials have explained to us, an operation has been implemented to resolve decennial addresses that do not geocode to TIGER when they are in blocks that border on or intersect jurisdictional boundaries. Since this effort applies to the cases with the greatest impact on decennial results, the bureau expects this operation to resolve the issue satisfactorily. However, the remaining addresses are not geocodable and may be located in error or without any geographic location when future TIGER updates are made. To prevent the loss of additional address links to TIGER in Census 2000, we are recommending that the bureau ensure that any further changes made to TIGER during the decennial are verified with the MAF. After the decennial, the bureau should devise methods to resolve all addresses that do not geocode to TIGER.

B.   *Software Engineering Standards Will Support Modernization*

The bureau's FY 2001 budget submission requests funding to replace MAF and TIGER with a modern system that will use current MAF and TIGER data. In the past, we have reported that the software development approach used by the bureau for the decennial was not based on software

---

[29]First, completely automated processing attempts to associate each address with a street and block in TIGER to geocode the address. Addresses that do not geocode on this first attempt are sent to a process called Master Address File Geocoding Office Resolution. This process consists of manual and on-line procedures through which clerks use maps and other local information to identify the location of an address and the corresponding block number. TIGER is then updated on-line with new streets, modified street names, or extended address ranges.

engineering standards for documenting requirements, reviewing software specifications and design, and ensuring that rigorous, independent testing is carried out.[30] This review found similar issues. Several specifications used by software programmers defined requirements in a narrative fashion, leaving room for ambiguity. After requesting written test plans, procedures, and results, we were also told that testing is informal and not well documented. Bureau Geography Division officials have stated that they would like to adhere to more formal software engineering standards and are exploring the use of CASE[31] tools to help set up a "top-down" software engineering approach to analyzing MAF and TIGER functionality before modernizing these systems. However, while CASE tools are useful, they are not a substitute for a complete life-cycle development process based on software engineering standards.

The Census Software Development Life Cycle manual[32] documents a process and standards for the first three phases of the life-cycle approach: system requirement definition, system requirements analysis, and software requirements definition and analysis. The manual, augmented by software engineering standards for the balance of the software life-cycle, would provide a helpful guide for developing a system that is soundly engineered and functionally correct. Using the rigor that these standards provide will help system developers realize system designs that meet user needs and maintain address data quality. Recognizing the importance of rigorous testing and quality assurance, the Decennial Management Division recently issued procedures to improve its software development process for CIFU[33] and is also improving the process for subsequent decennial software development. We believe that the bureau is doing the right thing by improving its software development process and that improvement should be expanded to the MAF and TIGER systems. Therefore, we are recommending that the bureau adopt software engineering standards as part of the MAF and TIGER modernization.

---

[30]*PAMS/ADAMS Should Provide Adequate Support for the Decennial Census, but Software Practices Need Improvement*, OSE-11684, March 2000; *Improvements Needed in Multiple Response Resolution to Ensure Accurate, Timely Processing for the 2000 Decennial Census,* OSE-10711, September 1999; and *Headquarters Information Processing Systems for the 2000 Decennial Census Require Technical and Management Plans and Procedures*, OSE-10034, November 1997.

[31]Computer Assisted Software Engineering (CASE) describes automated tools that aid software developers throughout all phases of the life cycle.

[32]*The Census Software Development Life Cycle*, Bureau of the Census, November 7, 1994.

[33]*Coverage Improvement Follow-up Development Quality Assurance Plan*, Bureau of the Census, May 30, 2000.

IV.   **Success in Meeting Housing Unit Accuracy and Completeness Goals Should Be Reported**

The MAF is often referred to as the heart of Census 2000 because it is intended to identify the households that will be counted,[34] and the quality of addresses for those households has a direct impact on the accuracy and completeness of the decennial. As we noted earlier, the bureau found that in the 1990 decennial, nearly one person in every three who were missed were not counted because their housing unit was not in the address file. To the extent feasible, having an accurate and complete file will reduce the cost of the census. Duplicate and nonexistent addresses increase followup costs because of the added complexity to resolve duplicate addresses and the attempt to visit many nonexistent and, therefore, nonresponding addresses.

Recognizing the importance of a complete and accurate list of housing units, the bureau has created housing unit coverage performance standards and methods to evaluate if they are met. The Accuracy and Coverage Evaluation will measure housing unit coverage. However, the bureau has not clearly stated how evaluation results will be reported against the standard for this decennial. The Government Performance and Results Act of 1993 requires agencies to report performance goals, measures, and accomplishments to the President and the Congress in an Annual Performance Plan and an Annual Program Performance Report. These documents are appropriate vehicles for reporting on this important data quality measure and can help ensure that the bureau works to improve data quality to improve all of its operations that rely on the MAF.[35]

A.   *The Bureau Has Housing Unit Coverage Goals and Evaluation Plans*

The bureau has already cited housing unit coverage standards in its reengineering plan. The goal was a net housing unit coverage standard of 1 percent, with 2.5 percent missing and 1.5 percent erroneous.[36] This standard was based on the bureau's goal of doing better than the 1990

---

[34]*Census 2000 MAF Program Master Plan* draft, Bureau of Census, May 1999.

[35]*Performance Information Challenges,* GAO/GGD-00-52. Also, see P.L. 103-62, August 3, 1993, *Government Performance and Results Act* and OMB Circular A-11, Part 2, July 1999 for information on the preparation and submission of strategic plans, annual performance plans, and annual program performance reports.

[36]The reengineering plan cited no more than 1 percent duplicates as part of the overcoverage measurement goal. We estimated the total overcoverage measurement goal, including duplicates and other nonexistent housing units, as 1.5 percent based on the bureau's goal for missing housing units of 2.5 percent and net housing unit undercoverage measurement of 1 percent The Dress Rehearsal *Report Card* also cited a 1-percent net housing unit undercoverage standard for city-style addresses.

decennial, in which the net housing unit undercoverage was 1.4 percent, with 3.4 percent missing and 2 percent erroneous.[37]

The bureau has a wealth of information that can be used to assess whether MAF performance goals are being met and how performance can be improved. The bureau will have information about MAF data quality through the Accuracy and Coverage Evaluation, which is being conducted to determine not only the number of people missed, but also the number of housing units missed or incorrectly counted. In addition, the bureau's January 1999 operating plan contains plans to obtain information about the quality of decennial data through evaluations of key components of the decennial process. This plan outlines numerous evaluation studies and reports on how well operations worked in building the MAF. The evaluation results will be useful in fully understanding the results of the Accuracy and Coverage Evaluation, including identifying successful operations and those that should be improved.

B.   The Annual Performance Plan and Report Should Be Used to Report Housing Coverage

The Government Performance and Results Act requires federal agencies to report to the President and the Congress on performance goals for each major program activity and accomplishments in meeting those goals. Under this law, the head of each agency is to submit a plan and report each year on program activities. The plans are to establish performance goals to define the level of performance to be achieved by each program activity. The performance reports are to contain two main parts: (1) a report on the actual performance achieved as compared to the performance goals expressed in the performance plans and (2) the plans and schedules to achieve those goals that were not met. If a performance goal becomes impractical or infeasible, the agency is to explain why that is the case and what legislative, regulatory, or other actions are needed to accomplish the goal, or whether the goal ought to be modified or discontinued. Finally, the reports should relate performance measurement information to program evaluation findings in order to give a clear picture of the agency's performance and its efforts at improvement.

The bureau has identified ways to measure housing unit coverage but has not stated how results will be reported to decision-makers for this decennial. One of the decennial's main goals is to provide quality data. According to the Department's Fiscal Year 2000 Annual Performance Plan, the bureau plans to measure success in achieving data quality by its ability to achieve a 0.1-percent net undercount.[38] Because the decennial is a census of housing as well as of the

---

[37]The bureau's 1990 Housing Unit Coverage Study, Preliminary Research and Evaluation Memorandum No. 193. Net housing unit undercoverage of 1.4 percent was derived by subtracting the bureau reported erroneous enumerations of 2 percent from the bureau report of omissions of 3.4 percent.

[38]The Department of Commerce Annual Performance Plan, Fiscal Year 2000.

population, the bureau's portion of the Department's Annual Performance Plan should include housing coverage as a performance measurement. The bureau planned to report on its Census 2000 accomplishments in a Report Card similar to that issued for the dress rehearsal; however, these plans have been canceled. The Annual Performance Plan and Annual Program Performance Report offer appropriate vehicles for providing this information.

Reporting Should Include Overcoverage and Undercoverage Components

In reporting MAF goals and performance, the bureau should include not just net housing unit coverage, but also the overcoverage and undercoverage components in order to provide a complete report on MAF data quality.[39] In the 1998 Dress Rehearsal, the bureau only reported its address goals in terms of a net housing unit undercoverage of 1.5 percent. This means that undercoverage and overcoverage combined, as calculated using the bureau's estimation techniques, should result in a net housing unit undercoverage no greater than 1.5 percent. For example, in Columbia, South Carolina, the bureau reported net housing unit undercoverage of 2.9 percent, which did not meet the net housing unit undercoverage standard of 1.5 percent. However, the two components that comprise the 2.9 percent net undercoverage measure provide important insight into the extent of overcoverage and undercoverage. The MAF missed 13 percent of the housing units in Columbia and erroneously included 10.4 percent of the housing units.[40] It is also possible for there to be both a large percentage of undercoverage and a nearly as large percentage of overcoverage, yet the net percentage of undercoverage could fall within the standard. The details behind the overall measure provide important insights into the extent of missing housing units and housing units counted in error that can be obscured by only reporting net undercoverage.

Therefore, we are recommending that the bureau include the housing unit coverage standards and results in achieving them—including overcoverage and undercoverage—as a second performance measure in its input to the Department's Annual Performance Plan and Program Performance Report for fiscal year 2002. By including housing unit coverage as a measure of decennial data quality, the bureau will attain a vehicle for reporting an important indicator of decennial success and will link the housing unit coverage standards with the evaluation of housing unit coverage. To justify steps to improve coverage, we are also recommending that the bureau retain the housing unit coverage measure as input to subsequent Annual Performance Plans and Program Performance Reports.

---

[39]See Reengineering Plan.

[40]*Results of the Housing Unit Matching Phase of the Integrated Coverage Measurement*, Bureau of the Census, DSSD ICM Dress Rehearsal Results Memorandum Series Number HU-1, September 9, 1998.

MAF Quality Improvement Program Should Be Continued

To ensure that it has the data needed for future annual reporting, the bureau should continue its quality assurance of MAF completeness and accuracy in terms of undercoverage and overcoverage by using either of the methods discussed below. In 1997, the bureau's evaluation staff designed an approach that would provide an annual measurement of MAF accuracy and completeness based on the methodology used in 1990 for statistical sampling and adjustment. The bureau used this approach in MAF quality improvement programs in 1997 and 1998. The *1997 MAF Quality Improvement Program Pilot Study* measured the completeness and accuracy of MAF residential addresses geocoded to a block in six counties. This pilot study concluded that, with a few modifications, the statistical adjustment methodology would provide an effective measurement of MAF housing unit coverage. The 1998 study confirmed relatively high MAF error rates and called for significant improvement in the MAF before the decennial. Although these studies provided important insights into MAF quality, the 1999 study designed to measure the results of the reengineered MAF-building approach was canceled. Census officials cited a lack of resources and schedule concerns. We believe, however, that these types of studies provide an important measurement of MAF accuracy and completeness. Also, the Population Division's methodology for comparing numbers of addresses in a county to independent demographic benchmarks has proved useful in determining areas with potential coverage problems. Therefore, we are recommending that the bureau continue projects designed to evaluate MAF housing unit coverage, which should be used throughout the next decade as a benchmark for basing MAF improvement and maintenance operations.

RECOMMENDATIONS

We recommend that the Director, Bureau of the Census, take the necessary actions to improve address data quality for Census 2000 and future censuses and surveys, including the following:

1.   **Explore causes for continuing overcoverage and undercoverage of housing units and use the resulting information as lessons learned when planning future MAF improvement operations.**

       The Census Bureau concurred with this recommendation.

2.   **Develop a MAF-building strategy that ensures sufficient time for MAF improvement operations and include as part of the strategy:**

       1.   **A consistent policy for address eligibility for improvement operations.**

       2.   **An approach for verifying addresses without block codes.**

       The Census Bureau concurred with this recommendation.

3.   **Study causes of duplicate addresses supplied by different sources, such as the Postal Service file, local/tribal governments, and block canvassing and implementing methods to prevent duplicates.**

       The Census Bureau concurred with this recommendation.

4.   **Develop an address eligibility policy that defines in advance the criteria to be used in selecting addresses during future censuses and surveys.**

       The Census Bureau concurred with this recommendation.

5.   **Issue a decision memorandum that explains the address eligibility policy for the final DMAF delivery of addresses to be included in Census 2000.**

       The Census Bureau stated that it agrees that a Decision Memorandum should be issued describing the general policy for additions and deletions from the DMAF.

       We believe that the memorandum should also explain the specific criteria for including addresses in Census 2000.

6.   **Use information in the MAF as a management tool in the future to increase the completeness and accuracy of the address file (for example, to identify areas where addresses are more likely to be missing or incorrectly geocoded).**

The Census Bureau concurred with this recommendation.

We believe that the bureau should also consider additional techniques, such as applying address history data to locate geographic areas where addresses are likely to be missing and to pinpoint addresses likely to be geocoded in error for inclusion in its planned work.

7.   **Ensure that any further TIGER changes are verified with the MAF so that no additional decennial addresses lose their link to TIGER for Census 2000.**

The Census Bureau concurred with this recommendation and provided a clarification to the draft report. Specifically, the bureau noted that for Census 2000, MAF addresses were not "lost" from the census if they for any reason ceased to match to the TIGER database; block codes derived from other sources, such as field work, overrode or substituted for TIGER block codes according to a documented scheme.

In response to the bureau's concerns, we clarified our report where appropriate.

8.   **Devise methods to resolve addresses that do not geocode to TIGER.**

The Census Bureau concurred with this recommendation.

9.   **Adopt software engineering standards as part of the MAF and TIGER modernization.**

The Census Bureau concurred with this recommendation.

10.   **Report housing unit coverage standards and results, broken out by their overcoverage and undercoverage components:**

a.   **As performance measures in the bureau's input into fiscal year 2002 and subsequent Departmental Annual Performance Plans.**

b.   **As performance results in the bureau's input into fiscal year 2002 and subsequent Departmental Annual Program Performance Reports.**

The Census Bureau concurred with the recommendation that separate measurements should be made and reported for MAF overcoverage and undercoverage. The bureau further agreed that there should be a continuous process for monitoring and improving MAF quality and coverage. However, the bureau stated that it would not be cost effective or practical to make annual national or small area MAF coverage measurements because doing so would divert an unacceptable level of key staff resources from planning and implementing actual MAF/TIGER modernization improvements. The bureau noted that planners for the American Community Survey and other intercensal demographic data collections have identified no program requirement for annual MAF overcoverage/ undercoverage measures. According to the bureau these measures would be extremely useful at wider intervals and, funding permitting, it plans to generate these at several points in the decade in preparing for the 2010 decennial census.

We understand the bureau's concerns about the cost of implementing an annual national or small area MAF coverage measurement. However, other methods, such as comparing tallies of MAF addresses to demographic estimates as employed by Population Division, may serve as an alternative.

11. **Continue projects designed to evaluate MAF housing unit coverage that can be used throughout the next decade as a benchmark for basing MAF improvement and maintenance operations.**

The Census Bureau concurred with this recommendation.

Appendix A.

Acronyms Used in This Report

CIFU   Coverage Improvement Followup

DMAF   Decennial Master Address File

DSCMO   Decennial Systems and Contracts Management Office

DSSD   Decennial Statistical Studies Division

GAO   General Accounting Office

ICM   Integrated Coverage Measurement

LUCA   Local Update of Census Addresses

MAF   Master Address File

NRFU   Nonresponse Followup

PMP   Program Management Plan

PRED   Planning, Research, and Evaluation Division

TEA   Type of Enumeration Area

TIGER®   Topologically Integrated Geographic Encoding and Referencing
              System

September 29, 2000

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

Through: Robert J. Shapiro
                            Under Secretary for Economic Affairs

From: Kenneth Prewitt
                            Director

Subject: *A Better Strategy Is Needed for Managing the Nation's*
                            Master Address File
                            Draft Inspection Report No. OSE-12065

This is in response to your memorandum dated August 29, 2000, transmitting the above-referenced draft inspection report.

The U.S. Census Bureau generally concurs with the findings in this report, and many of the improvements to the Master Address File (MAF) recommended by the Office of Inspector General (OIG) either have already been implemented, or are part of our future MAF/TIGER® enhancement plans. However, the Census Bureau has concerns about comments on the following pages of the report:

Page i —   The reports cites a Census Bureau estimate that there are "112.5 million housing units nationwide." This estimate was established in July 1998. More recent estimates are available. We suggest you note the source and vintage of any estimate you cite.

Page ii—   The report states that: "The bureau did not have sufficient time available to ensure high quality address data. Making the decision to reengineer the MAF-building approach only two years before the addresses were needed for addressing questionnaires did not leave enough time to carry out the new strategy." This is not accurate. Difficulties in meeting the deadlines for the Local Update of Census Addresses (LUCA) occurred because implementing the operation required more time and effort than expected, particularly given the need to assure address confidentiality and quality. These delays were not related to the decision to reengineer the MAF-building approach, as the report suggests.

Page ii – In the second paragraph on page ii, the report states that: "The bureau has taken steps to improve address quality and has potentially identified almost 20 million nonexistent and vacant addresses, 4.3 million before nonresponse followup and 15.6 million during nonresponse followup. However, the policy for determining the addresses eligible for several previous and ongoing census operations has not been well-defined, and the decision about which addresses to include in the final decennial results has not been made." These processes, particularly as they relate to the Nonresponse Follow-up (NRFU) and Coverage Improvement Follow-up (CIFU) operations, are not within the scope of MAF development. Moreover, it is misleading to include the 15.6 million vacant addresses that existed during NRFU as MAF quality enhancements. The more important figures are those that resulted subsequent to the completion of the NRFU and CIFU operations, after potential vacants and nonexisting addresses have been confirmed, converted to occupied, or removed from the census.

Page ii – Here and elsewhere in the report, reference is made to the "Accuracy and *Completeness* Evaluation." This operation is called the Accuracy and *Coverage* Evaluation.

Page 1 – The report stresses the importance of the MAF for those receiving forms in the mail. The MAF is equally important for the Update/Leave operation as well.

Page 3 – The Address List Development Operations Planning Group, cited in the report, is responsible for designing all address list development activities and communicating the requirements of these operations—not simply "resolving decennial address issues," as the report suggests.

Page 4 – The MAF database does not "track the nation's addresses and their geographic locations" as the report states. This is not a MAF function. The MAF records the status and census block location confirmed by Census Bureau field operations for each address.

Page 6 – The report states that "About a year before Census 2000, the bureau planned a Local Update of Census Addresses (LUCA) program to provide an opportunity for participating local/tribal governments to review the address lists and provide updates and corrections." This is misleading. While the Census Bureau originally intended to implement LUCA one year before Census 2000, the idea was developed much earlier in the decade, and the city-style address portion of the program—known as LUCA 1998—was implemented nearly two years before Census 2000.

Page 7 – The report states that "the bureau subsequently modified some aspects of the reengineered plan. It replaced the Postal Service validation with a review and correction of addresses in the summer of 1999." It is worth noting that the traditional

United States Postal Service (USPS) validation was replaced with a *letter carrier* review and correction of the USPS's Delivery Sequence File and that this operation was repeated in January 2000. In addition, the Census Bureau used updated address data from the USPS to update the MAF in February 2000.

Page 13 – The report states that "LUCA 98 schedule slips and bureau policy regarding which addresses were eligible to be included in block canvassing hampered this major operation designed to improve address quality." Census Bureau staff do not believe that LUCA 98 schedule changes undermined the block canvassing operation. In addition, it is not clear that the inclusion of unresolved addresses in the DMAF complicated decennial operations, as the report suggests. It would be helpful if the OIG provided specific examples of such complications.

Page 21 – With respect to the report's contention that "consistency between MAF and TIGER® [was] not easily maintained," it is important to stress that if an address is assigned a block code through a field operation (as would be a block canvass add, for example), the address retains its field-assigned block code, regardless of its geocodability to the TIGER database and, thus, is included in further census processes. Further, if an address that once geocoded to TIGER loses its linkage to the TIGER database through some change, the address retains its block code previously assigned by TIGER.

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

1. **"Explore causes for continuing overcoverage and undercoverage of housing units and use the resulting information as lessons learned when planning future MAF improvement operations."**

The Census Bureau concurs with this recommendation and already has established groundwork for MAF coverage analysis by:

- Designing and maintaining extensive address history and source flags in the MAF itself.

- Designing and delivering MAF evaluation extracts that can be readily used by the Census Bureau's statistical analysis staff.

- Designing and implementing a MAF Accuracy and Coverage Evaluation that will provide data on housing unit coverage and geocoding accuracy.

- Designing a set of comprehensive Census 2000 MAF evaluations that will measure the impacts of various MAF improvement operations as well as the quality of the MAF itself.

- Designing and implementing a MAF Quality Improvement program that will be carried forward for nationwide quality measurement during the coming decade.

- Designing and implementing a system for comparing the MAF to independent demographic benchmarks.

The Geography Division, in developing its proposals for modernization of the MAF/TIGER System, is seeking information from the private sector on address list-building approaches that achieve list completeness and avoid the creation of duplicate addresses. The Geography Division and staff in the Field and Demographic directorates are cooperating in the planning and implementation of the Community Address Updating System (CAUS), which will take advantage of the on-the-ground presence of American Community Survey field representatives to effect improvements to the MAF.

2. **"Develop a MAF-building strategy that ensures sufficient time for MAF improvement operations and includes as part of the strategy a consistent policy for address eligibility for improvement operations and an approach for verifying addresses without block codes."**

The Census Bureau concurs with this recommendation. The address source and history flags in the MAF will allow different internal customers for the MAF to set appropriate and consistent address eligibility specifications, given their program requirements, and the results of the various measurement and evaluation activities described above will inform the setting of those requirements. The Census Bureau is planning future MAF improvement operations with lessons learned from Census 2000 in mind. Various tools and approaches described in its proposal for MAF/TIGER modernization reflect this, including:

- Automating address list improvement field work using portable computers that allow field workers to view, record, and edit the full set of MAF addresses, their locations, and the related street information in the TIGER database for their assignment area.

- Continuous or "rolling" LUCA, rather than a LUCA conducted only close to the census in competition with other census activities.

- Targeted field work by CAUS staff throughout the decade.

The Census Bureau will carefully examine its Census 2000 experience with address canvassing activities and consider such potential improvements as including all structures (and not just those presumed to be residential) and segmenting canvassing into specialized modules or phases that will more effectively accomplish tasks such as:

- Establishing unit inventories within multiunit structures.

- Verifying the existence of specific addresses.

- Verifying and determining block locations for ungeocoded addresses, possibly using 5-, 7-, or 9-digit ZIP Codes on addresses as a starting point.

3. **"Study causes of duplicate addresses supplied by different sources, such as the Postal Service file, local/tribal governments, and block canvassing and implementing methods to prevent duplicates."**

The Census Bureau concurs with this recommendation. Address duplication is one type of coverage problem that will be addressed in the actions related to recommendation 1.

4. **"Develop an address eligibility policy that defines in advance the criteria to be used in selecting addresses during future censuses and surveys."**

The Census Bureau concurs with this recommendation and describes its proposed actions to meet it in the discussion of recommendation 2 above. It is worth noting that there will likely not be a single set of eligibility criteria that will apply to every MAF extract delivered to support a census, survey, or other use of the MAF, and requirements will vary over time as well. Nonetheless, the Census Bureau agrees that appropriate criteria should be determined in advance of their implementation and documented thoroughly.

5. **"Issue a decision memorandum that explains the address eligibility policy for the final DMAF delivery of addresses to be included in Census 2000."**

The Census Bureau agrees that a Decision Memorandum should be issued describing the general policy for additions and deletions from the DMAF. This will be issued in the near future.

6. **"Use information in the MAF as a management tool in the future to increase the completeness and accuracy of the address file (for example, to identify areas where addresses are more likely to be missing or incorrectly geocoded)."**

The Census Bureau concurs with this recommendation and addresses its planned actions in the discussion of recommendations 1 and 2.

7. **"Ensure that any further TIGER® changes are verified with the MAF so that no additional decennial addresses lose their link to TIGER for Census 2000."**

The Census Bureau concurs with this recommendation, but believes the draft report needs to be clarified in one regard. For Census 2000, MAF addresses were not "lost" from the census if they for any reason ceased to match to the TIGER database; block codes derived from other sources, such as field work, overrode or substituted for block codes currently TIGER-derived according to a documented scheme.

The Census Bureau proposal for MAF/TIGER modernization calls for greater integration between the two databases through tools that would provide for concurrent viewing and modification of individual addresses in concert with the street feature and locational information currently stored in the TIGER database.

8. **"Devise methods to resolve addresses that do not geocode to TIGER®."**

The Census Bureau concurs with this recommendation and, in fact, already has devised a system for identifying Census 2000 addresses without a current linkage to the TIGER database and resolving them through a combination of office and field research. Implementation has begun, with priority on those discrepancy cases involving collection blocks split by tabulation block boundaries, which are those with the potential to affect governmental and statistical entity-level counts. Not all discrepancy cases will lead to census errors. The Census Bureau must also resume resolving **new** addresses that do not geocode to TIGER, and these efforts will likely be combined.

9. **"Adopt software engineering standards as part of the MAF and TIGER modernization."**

The Census Bureau concurs with this recommendation and believes that a more formal life-cycle development process must be an integral part of the modernized MAF/TIGER System. More immediately, efforts are underway to implement more rigorous testing and improve quality assurance. In addition to new software development quality assurance processes, such as those implemented for CIFU and Census Unedited File creation, rigorous testing plans linked to requirements documents are being put into place for other processes such as tabulation block geocoding and development of geographic products for DADS. We have instituted use of PVCS Configuration Management software for upcoming coding operations, as well as for the Geography Division Production Control System.

10. "Report housing unit coverage standards and results, broken out by their overcoverage and undercoverage components, as performance measures in the bureau's input into fiscal year 2002 and subsequent Departmental Annual Performance Plans and as performance results in the bureau's input into fiscal year 2002 and subsequent Departmental Annual Program Performance Reports."

The Census Bureau concurs with the recommendation that separate measurements should be made and reported for MAF overcoverage and undercoverage; these are distinct problems with distinct solutions. These separate measurements will be reported for the Census 2000 MAF. The Census Bureau further agrees that there should be a continuous process for monitoring and improving MAF quality and coverage. However, the Census Bureau **strongly** disagrees that it would be cost-effective or practical to make annual national or small area MAF coverage measurements. Doing so would divert an unacceptable level of key staff resources from planning and implementing actual MAF/TIGER modernization improvements. Planners for the American Community Survey and other intercensal demographic data collections have identified no program requirement for annual MAF overcoverage/undercoverage measures. These measures would be extremely useful at wider intervals and, funding permitting, the Census Bureau plans to generate these at several points in the decade to plan for the 2010 decennial census.

11. "Continue projects designed to evaluate MAF housing unit coverage that can be used throughout the next decade as a benchmark for basing MAF improvement and maintenance operations."

The Census Bureau concurs with this recommendation. The Census Bureau's intent is to use existing evaluation tools and develop improved tools to identify, analyze, and monitor progress toward eliminating various specific MAF limitations and deficiencies throughout the decade. Funding permitting, these approaches would be combined with periodic coverage measurement activities.

cc: US/EA