An Overview of the Master Address File Coverage Study

Evan Boyd, Earl Letourneau, Aneesah Williams
U.S. Census Bureau, Decennial Statistical Studies Division,
4600 Silver Hill Road, Washington, DC 20233

Abstract

The Master Address File (MAF) Coverage Study (MAFCS) is one component of the 2020 Census address canvassing operation. The intent of the program is to produce coverage estimates of the MAF, provide continuous updates to the MAF, evaluate in-office address canvassing updates, evaluate in-field address canvassing targeting strategies, and test new in-field address canvassing procedures. This paper discusses an overview of the survey, including the objectives, sample design, and the analysis plan for the data collected. Approximately 20,000 blocks per year will be selected for an in-field canvass, beginning in 2016 through 2019. Field listers will canvass assigned blocks collecting address data for every place people might live or stay. The initial MAFCS uses a stratified sample based on geography and size of a block. Future sample designs will incorporate information from the in-office address canvassing operations as well. Key estimates include the numbers of added, deleted and verified addresses for the nation and for several domains of interest.

Keywords: address canvassing, coverage

1. Introduction

The Master Address File (MAF) Coverage Study (MAFCS) is a component of address canvassing that aims to identify and improve coverage of the U.S., including Puerto Rico on the MAF. The MAFCS will produce estimates for newly added, deleted, or validated addresses nationally and at some sub-national levels. Other applications of the MAFCS include assessing the In-Office Address Canvassing strategies used for MAF updates, testing in-field procedures, and updating the address list on the MAF. Topics to be discussed include the MAFCS relationship to re-engineered address canvassing, an overview of the MAF, methodology used in the MAFCS, and sample designs.

2. Background

The MAFCS conjoins with the Community Address Update System (CAUS) program which updates address lists used in the American Community Survey (ACS). In the past, CAUS selected up to 1,500 blocks per year where the United State Postal Service (USPS) Delivery Sequence File (DSF) was not currently providing reliable address information. The blocks selected by CAUS provide field verified address information used to update the MAF. Now, the MAFCS supersedes CAUS to select approximately 20,000 sample blocks on a yearly basis.

2.1. Relationship to Re-Engineered Address Canvassing

The MAFCS is part of the Re-Engineered Address Canvassing program that implements improvements to the quality of the Census Bureau's address list before the 2020 Census enumeration. The address list needs to contain complete and accurate address information, including the type and unit characteristics, for the physical location of each living quarter in the U.S. and Puerto Rico to ensure that residents will be invited to participate in the census.

Previous Censuses have included a very costly, full in-field address canvassing. This act of walking every block in the nation is no longer deemed necessary because of advances in technology that allow continual address and spatial updates to occur throughout the decade as part of an In-Office Address Canvassing effort. One major goal of In-Office Address Canvassing is to have no more than 25% of housing units canvassed through an In-Field operation. Some address information pertaining to structures in these blocks, such as number of units in an apartment building, are not accurately obtainable through imagery. Blocks with these characteristics are identified as target areas to potentially be selected in the MAFCS.

3. The Master Address File

3.1. What is the MAF?

The MAF is a database that contains both physical and descriptive attributes for all known living quarters in the U.S. and Puerto Rico. It serves as the base frame for Census operations and household surveys, such as the mail-out of questionnaires and postcards. Ideally, each address in the MAF is paired with a location in the Topologically Integrated Geographic Encoding and Referencing (TIGER) database to ensure census data is accurate in the corresponding geographic location.

3.2. History of the MAF

The MAF was originally created for the 2000 Census. Initial information contained in the MAF was based on the 1990 Address Control File and the United States Postal Services (USPS) Delivery Sequence File. The results of the previous Census, combined with the USPS Delivery Sequence File, made up the address list to be used for the 2000 Census.

The MAF has been updated with address canvassing information collected during Census 2000 and Census 2010, along with some updates through partnership programs that provide the MAF with address list updates.

3.2.1. Contents of the MAF

The MAF contains address information for single structures, or single units within a structure, that exist on the ground. The main types of MAF records are housing units, group quarters (such as college dormitories, prisons, and group homes), transitory locations (such as hotels and trailer pads), and nonresidential records. These MAF records store numerous attributes about the structure like its coordinates, when the record was added to the MAF, and its mailing address.

3.3. State of the MAF

The MAF contains a nearly complete address list, however, there are some areas better covered than others. The number of living quarters is constantly changing across the nation. There are subdivisions being built and structures being demolished continuously,

which proves to be why the MAF is never perfect in real-time. There are also duplicate addresses in the MAF due to conflicting information from multiple sources. This can happen if a single apartment inside of an apartment complex is listed as 123 Main Street Apt. A in the MAF. However, the same living quarter could be delivered in an address list update from the USPS Delivery Sequence File as 123 Main Street Apt.1. These are often complex situations, as though it might seem that these are simply the same living quarter, there is the possibility that both addresses could exist on the ground as separate units.

3.4. Maintaining the MAF

To aid in the efforts of keeping the address list current between censuses, the Census receives address update files twice a year from the United States Postal Service. The Census uses household surveys, such as American Community Survey research, to update the address list on a continuous basis. The Census Bureau receives update files from geographic support systems. This data mainly comes from state and county governments. This partnership program has been successful at acquiring, evaluating, and incorporating current, high-quality address and road data from 618 partners that encompass more than 24,000 governmental entities, over 71% of housing units and 72% of the population in the 2010 Census. A visual representation of these geographic partners can be found in Attachment A. Commercial data sources have also provided the Census Bureau with address lists for MAF updating. Now, the Master Address File Coverage Study provides another source of regular address list updates for the MAF.

4. MAF Coverage Study

4.1. Purpose

The MAFCS is a survey that conducts an annual listing of 20,000 sample blocks as a component of the Address Canvassing operation. The MAFCS is intended to achieve the following objectives:

- Produce national and sub-national estimates of MAF coverage
- Continuously update the MAF for household surveys and the Census
- Refine parameters of the In-Office Address Canvassing operation
- Locate areas that are in need of in-field address canvassing

4.2. Methodology

The MAFCS occurs on a yearly basis from 2016 through 2019. The general outline of MAFCS processes are described next.

A sample of about 20,000 blocks is selected from a universe of addresses deemed valid by a specific filter that is applied to the MAF. This universe consists of records in the MAF that are believed to correctly represent up-to-date addresses. In 2016, these sample blocks were worked between the six-month period of April and September; however, future samples will be evenly distributed across a 12-month period.

The MAFCS is a dependent listing operation in which field workers are given a list of current addresses on the MAF. The field workers attempt to find the addresses on their list, just as they are listed in the MAF. If the MAF has the correct address information for their current location, they will simply verify the record. Field workers will also encounter structures on the ground that do not appear on their list. They will then add

these new units, collecting the address information for the structure as well as the residential or non-residential status, and contact information for group quarters. On occasion, they will be unable to locate an address on their dependent list. A building may have been demolished or converted into a business. At this point, these addresses will be marked as no longer valid, or a good address, for household surveys and the Census.

Using the information gathered by the field workers, the MAFCS will produce coverage estimates annually.

4.2.1. Questions and Estimates the MAFCS Aims to Answer

The MAFCS hopes to provide some insight to several areas of interest, including producing estimates of coverage on the MAF, comparing in-field address canvassing results to that of other census operations, as well as informing the operational design for the 2020 Census.

Using data collected from the field, the MAFCS will make estimates at the following levels:

- United States total (50 states and District of Columbia)
- Puerto Rico (beginning in 2017)
- Sub-national geography designation (central city, exurban, suburban, small town, and rural)
- American Indian reservation (on, or not on)

Other estimates will be made for internal use to better inform decennial operations:

- For In-Office address canvassing:
  - The blocks selected in the MAF Coverage Study sample will be worked through the first phase of In-Office Address Canvassing, or interactive review, prior to fieldwork being completed for the survey. This allows the ability to produce estimates to compare results from data collected in-field to that of In-Office Address Canvassing
- Inside or outside of a geographic support systems area
  - This enables comparison of in-field results to the information provided by geographic partners
- Comparison of address estimates made from field work to address estimates made from the address lists provided by commercial data sources
- Type of Enumeration Area (mailout/mailback, update methods)

4.3. Sampling

Sampling for the MAFCS enables sufficient data collection to reach the objectives and attempt to answer the research questions stated above. The main goal of the MAFCS is to produce MAF coverage estimates, but the design must also consider CAUS requirements and other appropriate guidelines for analyzing in-field processes.

Other components of the Address Canvassing program must be reflected in the sample design. The MAFCS will use CAUS requirements provided by the Geography Division (GEO) by focusing on rural blocks containing units with non-city style address that are unlikely to exist on the DSF. The MAFCS excludes some geographic areas for sample selection to ensure only blocks containing pertinent information for analysis and estimation purposes can be selected for sample. In-Office Address Canvassing results for

each block will be sent to DSSD in order to establish active, passive, and on hold blocks (those that do not have sufficient imagery available for accurate designation) for use in sampling rates. Some blocks in the MAF are inaccessible during winter months, thus, the sample design will schedule data collection for them during another season. GEO also provides DSSD with a list of partner identified blocks and a block-characteristic file to aid in sample selection. This file contains summarized data such as number of living quarters in the block, percentage of non-city-style addresses in the block, and percentage of addresses in the block included in U.S. Postal Service files.

4.3.1. 2016 MAFCS Sample

The 2016 MAFCS Sample was designed and selected before In-Office Address Canvassing began. This sample was selected based on the design of an earlier address canvassing test. It was largely a state-based design, stratified by the number of expected living quarters in a block.

Remote Alaska, water, median, and public land use blocks were removed from the universe of blocks eligible for sample. The remaining 11,036,716 blocks formed the sampling universe. Note that Puerto Rico was not in scope for the 2016 sample. The sample design required that each state, along with the top four most populated cities – New York City, Los Angeles, Chicago, and Houston – have a minimum of 300 blocks in the sample.

Blocks were selected using a stratified probability sample, with state and number of addresses in a block, or measure of size, as the primary strata. Varying sampling rates were assigned based on the measure of size of the block; those with a higher number of addresses in them were oversampled. Blocks were then sorted by geography (county and tract), leading to five stratum in each state based solely on the size of the block. Measure of size also determined the number of blocks that were grouped together to form clusters. Clustering reduces the travel time for field workers, saving money and maximizing the number of housing units the field worker can complete in a day. This also decreases the likelihood of selecting geographically isolated blocks. The clusters were numbered sequentially and were not allowed to cross state, county, or tract boundaries or oversampling strata.

The probability of selecting a given block to be included in sample was based on the product of a constant national sampling rate, the blocks assigned oversampling rate, and an adjusted rate required to ensure a minimum of 300 blocks per state.

Once blocks were selected for sample, assignment codes were placed on each cluster describing when it was to be worked. Blocks were divided into three two-month assignment periods. Adjustments in work assignment dates were made to ensure field worker's accessibility of particular areas during specific seasons. As previously mentioned, these blocks began being canvassed in April 2016 and will be completed in September 2016. A map of the MAFCS 2016 sample blocks can be found in Attachment B.

4.3.2. 2017 Sample Plan

The 2017 MAFCS sample plan continues to use a stratified probability sample to select the 20,000 blocks. It does not include any findings from the 2016 sample considering data collection was not yet complete at the time of sample selection. However,

information from in-office address canvassing was available. As such, those results were used in the design of the 2017 sample.

There are a few differences between the 2016 and 2017 sample designs. The 2017 sample eliminates the state minimum requirement of 300 blocks and the stratification by state. Puerto Rico is now in-scope for the 2017 sample. Blocks removed from the universe include exclusion areas that are now consistent with In-Office Address Canvassing exclusions, blocks selected for the Address Canvassing Test, and blocks in the 2017 Puerto Rico Test. The universe yielded 9,979,985 blocks eligible for sample selection.

The first level of strata used in 2017 are based on type of geography of a block; small town, rural, suburban, exurban, central city. Also, before any further stratification was done, blocks containing zero housing units were placed in their own strata and not placed in any of the geography strata. Blocks in this particular stratum will only have the first two levels of stratification. The second level of strata uses results from IOAC-IR that denote whether a block is active (9% of sample), passive (26%), or other (65%) which contains on hold blocks and those not yet processed through the first phase of In-Office Address Canvassing. The third level of strata is based on the measure of size of the block. The stratum consisting of zero blocks accounted for 10% of the 2017 sample, while small blocks (containing less than 100 units) made up 79%, and large blocks (100+ units) compose the remaining 11% of the sample.

This stratification allows the MAFCS to oversample a number of these strata which are of particular interest and may help to identify areas of poor coverage. Active blocks identified by IOAC-IR are likely to show some type of change, either growth or decline, in the number of living quarters in the block. Since the in-office operation that adds or deletes units was not underway before the 2017 MAFCS sample selection, these blocks were oversampled to allow the MAFCS an opportunity to provide accurate address information for units in these changing blocks. Passive blocks were also oversampled to ensure blocks deemed passive by IOAC-IR are indeed stable and show no active change occurring. Blocks with 100 or more living quarters, along with blocks in the central city stratum, were also oversampled for the 2017 MAFCS sample. These blocks are likely to contain multi-unit structures where the actual number of living quarters in these structures are more difficult to assess based solely on in-office procedures.

Blocks were grouped together with sizes of one, two, or five to form clusters dependent upon the first level strata. Central city blocks were given a cluster size of one block, suburban and exurban blocks were clustered in groups of two blocks per cluster, while zero housing unit, small town, and rural strata consisted of five blocks per cluster. A sort by stratum, state, county, tract, large multi-unit identifier (if a block did or did not contain a multi-unit with 25 or more units in a single structure), and block was performed before selecting clusters via systematic random sampling.

The blocks selected for sample were split into six two-month assignment periods, and as in 2016, adjustments were made to ensure certain areas are reachable by field workers during their assignment period. Using the two-month assignment period is necessary for field procedures, but also allows the MAFCS to produce national estimates every six months.

5. Summary

The MAFCS is a vital element in maximizing the success of the 2020 Census and Address Canvassing operations throughout the decade. The MAFCS is an important component in achieving the goal of reducing expenses for the 2020 Census without sacrificing the quality of the data.

The functionality of the MAFCS improves overall accuracy of the MAF, leading to greater confidence in the address information on the MAF. Simultaneously, the MAFCS provides a means of recognizing blocks and types of areas that are not well covered by the MAF, requiring in-field address canvassing to properly update the address information. The data received from the sample blocks will enable improvements to be made throughout the In-Office Address Canvassing process.

Upcoming milestones for the MAFCS include completion of data collection for the 2016 sample. Once this procedure is finished, the MAF is updated with the newest address information allowing estimation and analysis to be conducted. This analysis will include calculating variances and sampling error. Preliminary findings are projected to be complete in March 2017. Beginning with the 2017 MAFCS, core estimates will be released on a six month cycle. For the 2018 sample selection, the first phase of In-Office Address Canvassing is expected to be complete. This will allow the sample design to make full use of In-Office Address Canvassing results.

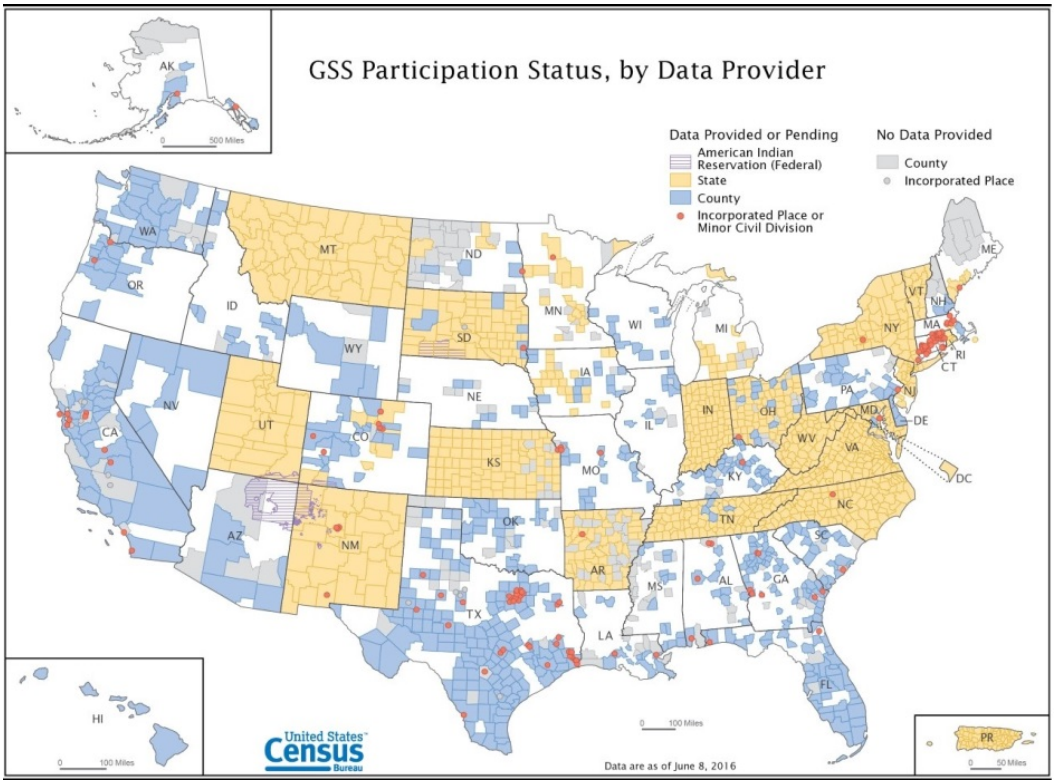
References

2020 Study Plan: Master Address File Coverage Study

Address Canvassing Detailed Operational Plan

2016 MAF Coverage Study Sample Design, Code Assignments, and Recommendations

Attachment A.



Attachment B.