

**MASTER ADDRESS FILE UPDATING TEST:
PROCESS, PROCEDURES, AND RESULTS OF THE 1999 FIELD TEST**

Kyra Linse, Kenneth Stump, US Census Bureau*

Kyra Linse, US Census Bureau, 4700 Silver Hill Rd., FOB3-3747, Washington, DC 20332

Keywords: Master Address File, Automated Listing and Mapping Instrument, Updating, Targeting, Listing Procedures

I. OVERVIEW

In 1998, the Census Bureau created the Master Address File (MAF) by starting with the 1990 Census housing unit addresses and adding new addresses from the Delivery Sequence File (DSF)—a file of mail delivery addresses from the United States Postal Service. Since then, the MAF has been updated every six months with the DSF. As of April 1, 2000, the MAF was a complete list of addresses after decennial census update operations.

The MAF was developed for the 2000 Census. It needs to be kept current to provide an updated address file for the Census Bureau's new American Community Survey which will become a national survey in 2003 and collect information similar to the current long form of the Decennial Census. The American Community Survey needs an all inclusive and current residential address list in order to select a representative sample. The MAF will also be used in sampling operations for other new surveys. In addition, if the Census Bureau keeps the address list current, it will not have the costly and timely listing update for the 2010 Decennial Census.

For the reasons mentioned above, the Census Bureau now needs to develop general updating procedures that are part of an ongoing year round effort to keep the MAF up-to-date. For the Census 2000, in addition to the DSF, the MAF was updated from listings from other operations such as Local Updating of Census Addresses and the New Construction Operations. The DSF does not always update all areas. In order for ACS to have a representative sample from the whole country, we need to update in areas not covered or incompletely covered by the DSF.

The Community Address Updating System (CAUS) will be a listing operation that will update counties with housing units not on the MAF—e.g., new construction. CAUS targets a county that is in need of updating based

on certain criteria, such as the completeness of the DSF for that county. Within that county we will first contact the local governments in an area to secure lists of new housing units built in their area. The Census Bureau will match these lists to the MAF to identify new and potentially new units in some areas and identify general areas the Field Representatives (FRs) will need to update. This will allow us to pick up the high amounts of growth within an area without the cost of updating the entire county. These addresses we collect will then be put on the MAF to supplement the adds provided by the DSF.

There were three objectives for the 1999 MAF Updating Field Test:

1. Identify the most effective targeting sources to target both counties and areas within the county that needed updating.
2. Field test the new Automated Listing and Mapping Instrument. This automated instrument will allow data capture of addresses and mapping by our FRs in the field, instead of requiring all the information to be keyed.
3. Develop procedures for the FR to target areas of growth and update the address listings.

Overall, we needed the FRs to understand and do these general tasks:

1. Visit the local governmental offices that provide complete coverage of new growth. Visit other offices to get more information on growth in the county.
2. Compile all lists and look up addresses on the maps. Look for areas with heavy growth. Define targeted areas. Make a plan of where to update.
3. Visit targeted areas. Update the blocks where there is growth.

The question was how to do these tasks in a cost-effective and timely manner while maximizing the number of updates they collected.

The six counties were chosen based on the fact that they were mostly rural. As mentioned earlier, the MAF is currently updated every six months using the DSF. In

*This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

many rural counties, often the DSF contains non-city style addresses, such as PO Boxes and Rural Route numbers. These addresses cannot be added directly to the MAF because they require updating in the field in order to spot the new units into the proper census block. Starting with counties that were in the ACS test sample, we chose counties with a significant proportion of non-city style addresses used for mail delivery, according to the MAF, and that have the most growth. We chose one county based on the fact that it was non-permit issuing—permits are not required for new construction housing. The final list of test sites were Calvert, MD; Miami, IN; Schuylkill, PA; Starr, TX; Vilas, WI and Washington, MO.

II. AUTOMATED LISTING

The Automated Listing and Mapping Instrument (ALMI) is the software program developed to do listing and mapping on a computer. The ALMI uses extracts from the MAF as its source of addresses. The FRs use the ALMI to list and update the addresses from the MAF extract without relisting all of the housing units onto a paper listing as they do currently.

The FR can review, verify, add, and delete any housing units within the tract as well as add and change the name of any features such as streets. A tract is a compact continuous land area that provides a geographic subdivision created by a group of blocks. The FR can work within one tract within a county at a time. Anywhere within that tract, they can accurately place the unit on a map.

When the updated files are loaded back at Headquarters, the addresses are loaded directly into the MAF where the information is updated both on the MAF and Topologically Integrated Geographic Encoding and Referencing (TIGER) files—the system and digital database developed at the Census Bureau to support its mapping needs for the Decennial Census and other Bureau programs. That address is then available for sampling for ACS and other surveys. Therefore, we want to collect as much information about the household as possible. Name, mailing address, phone number, and mapspots are important for finding and contacting the household in the future, but we also want as much information possible to be able to unduplicate addresses that are received to be sure that one household cannot be in sample twice.

The ALMI was loaded onto a pen-based computer rather than a standard keyboard laptop. We thought a pen-based computer would be easier for the FR to handle and record information. The pen is more accurate for mapping than

a laptop mouse which is often a peg in the middle of the keyboard. The pen allows the FR to comfortably hold the computer while recording the information at the door. It also comes equipped with a Global Positioning System (GPS) receiver, which we plan to test for mapping in the future. Overall, we felt the pen-based computer would be more FR-friendly and accurate than a laptop.

The ALMI has three main screens: the map view, the address view, and the block view. The *map view*—the main view within the ALMI—is the most complex. It allows the FR to view any part of the tract currently loaded, as well as to perform edits to the addresses. The *address view* shows the list of addresses in a selected block. A block is a smaller area within a tract often formed by physical boundaries such as streets. The FR can look at all information for an address in the address view. The *block view* is a list of all the blocks in the tract and the current number of addresses within each block, as well as the number of addresses the FR has edited or verified within the block.

We feel that the automated process made verifying existing addresses easier and more accurate, but the software needs to be refined. The following problems were some of the main observations pointed out by the FRs in the test. They are being addressed for next year's test.

1. We encountered problems where the file recognized housing units from different parts of the county as one mapspot—effectively making multiple single homes across the county a multi-unit.
2. The handwriting recognition was not up to the standard needed for interviewing at the door. The FR needs to be able to quickly record information while interviewing. Entering the information with the pen produced many mistakes and was frustrating for the FR. If the FR made a mistake, it was difficult to correct during the interview. The FR usually corrected the mistakes after the interview. All six FRs admitted using the on-screen keyboard to enter information at least some of the time. Three of them responded that they used it most, if not all, of the time.
3. The questions during interviewing did not flow in a logical order for the FRs. The FRs weren't sure what they could fill in by themselves and what to ask aloud to the respondent. Also, the respondent didn't always provide the information in the same order as the questions. The format of the interview questions has since been revised, so the questions can now be asked in any order by the FR.

4. Different features were hard to distinguish because they were all shades of black or gray. Some of the lines were barely visible. The political boundaries appeared as a light shade of gray. In direct sunlight, the line was not visible. This made mapping within a block very difficult. The boundary lines are being revised for the next test. The laptop being tested in 2000 will have a color screen.

5. Adding streets proved more difficult than intended due to some safeguards we had built into the software. In order to avoid “floating streets”—streets that are not connected to another street or trail, thus having no real access point—we required the street to be drawn within eighteen scale feet of another street. The software automatically connected those streets. In addition, the FR could not draw the street across a block boundary. This was to make sure the FR did not draw a short road on the other side of the block boundary (often a street) that didn’t exist.

The FRs had a difficult time accurately drawing streets with these constraints and were forced to redraw the roads several times. The main reason for this difficulty was that when working in a certain block the boundaries of that block were highlighted and the street appeared thicker than normal. This made it difficult for the FRs to start drawing a new street within eighteen scale feet.

Overall, the ALMI accomplished our goal of updating the county acceptably. The FR could review, verify, and correct information about existing units, add new units and delete units that no longer existed without having to use lists and paper maps. The ALMI requires some modifications to make it more user friendly. The problems that many FRs encountered were due to lack of training and understanding. The ALMI will be tested two more times before actual production for CAUS and will be modified for current surveys use as well.

III. FIELD PROCEDURES

The FR’s job was to identify specific areas (i.e., tract, zip code, part of county, etc) and specific blocks within the county that contained new housing units and update them.

In this report, we refer to these areas as “targeted areas of growth” (TAGs). This overall job was split into many tasks.

In order to maximize the FR’s efficiency with the minimum of cost, we first provided the FRs a listing of various government offices, with addresses and telephone numbers. The FRs were to telephone the sources and gather lists from at least two of them, visiting those sources to confirm that the lists met the criteria. A

Primary Sources Form (PSF) containing a list of all known government sources was included in the appropriate FRs’ folder of materials. The PSF contained the sources the FRs had to contact and permitted the FRs to keep track whether the office would be helpful or not. We first asked the FRs to review this form and make telephone calls to those sources listed, using a Government Units (GU) Screener form for each source. The form was designed to obtain some general information about the GU source as well as screen any list of new construction and/or growth to see if it had the necessary information. This would be of valuable assistance to the FRs in determining which government sources to visit.

Once the initial telephone calls were made, the FRs visited those Governmental Units they thought would provide the most up-to-date lists of addresses of new construction and growth using these criteria:

1. The lists must cover the entire county.
2. The lists must provide information about new construction built from April 1998 to the present, or as close to this as possible.
3. The lists must include all types of housing units, including mobile homes, and multi-unit structures.

The FR met with the GU contact person, reviewed the lists and either proceeded to the next step if the lists were considered complete or went to additional sources to obtain complete lists.

In addition to the PSF, we also provided the FR with a Secondary Sources Form (SSF). This SSF was a “back-up sheet”, designed to be used if none of the sources on the PSF were able to produce a satisfactory listing of new housing units. Unlike the PSF, the FRs filled in the SSF as they contacted other offices within the county such as real estate offices. A side benefit of this form was to create a database field of additional sources that would be helpful to contact in future years. Only in a few instances was the SSF used. Most of the sources contacted were the governmental offices from the PSF.

Once the FRs obtained completed address lists from at least two government unit sources, they were instructed to merge their lists onto a paper form, called the Master Updating Sheet (MUS), and group (or cluster) the addresses by block or block groups. They then pinpointed areas using an atlas, which we provided for each of the six test sites.

The first step was to mark the areas of growth on maps that Headquarters provided, and record the location

information on the MUS. The MUS was designed to record new housing growth identification, sorted by the FR's preference to pinpoint duplicate addresses and to cluster contiguous areas. We suggested listing the addresses by tract, although other methods could be done.

Then, the FRs reviewed the map and the MUS to find areas that appeared to have a high amount of growth—combining contiguous blocks together where appropriate and forming their TAGs on the MUS.

We emphasized that the TAGs should be high growth areas. A high growth area was defined as a geographic area with a substantial amount of new growth. The final decision as to what constituted “high growth” rested with the FRs.

The intent of grouping areas of high growth on the map and the MUS was to assist the FRs in planning an efficient route of travel to the blocks selected for the listing and precanvassing phases of the test.

Once the FRs went through each of the source address list and recorded them on the Census Bureau maps and on the MUS, they should have had one complete list of new housing units and the corresponding maps should have been annotated. The FRs' next step was to determine which areas received priority for updating as TAGs. Hopefully, by looking at the maps, the FRs would see where the new growth was clustered.

When the FRs arrived at their TAGS, the first step for them was to precanvass the area and look for the new housing units or look for signs of growth. We gave them examples of signs of new growth such as newly paved roads, houses that do not have a lawn yet, or houses that are distinctly different than others. We asked the FRs to drive the roads of the blocks where the new housing units could be located.

Once the precanvassing was completed, the FRs were instructed to update the area, focusing on the streets or street segments that had new growth. In many cases, this would include driving every street that may have growth, going into an adjacent block, or it might mean driving every street in the TAG. In our test counties, which were predominately rural, we realized early on that asking the field staff to update the entire block might be too time consuming and exhaustive. For this reason we stressed that they should be updating the area(s) where they found new housing units or housing developments.

The FRs, having determined which block(s) or street segments to update, would then do a dependent listing of

the block(s) or street segments using the ALMI. The FR would add a new unit, verify an existing unit, correct an existing unit or delete a housing unit that no longer existed. The examples of actual updating procedures were discussed in detail in Section II.

Finally, we also wanted to test other sources that might aid us in developing a more accurate database of new housing units. One source in particular, the American Business Index (ABI), contains a list of business establishments in the U.S., instead of housing units. Some businesses are connected to housing units—such as mobile home parks and apartment buildings—that may be missed by the government sources. We asked the FRs to follow up on a sample of business establishments that employees at Headquarters had telephoned initially in June, 1999. The pertinent information was listed on a Field Representative Verification (FRV) Sheet.

We looked at two main types of residential establishments listed as “business” on the ABI—mobile home parks and multi-unit structures. These are places that in the past have been difficult to keep updated.

We found, in some cases, that the addresses from the ABI were incorrect, or the phone numbers had changed. In most cases, the FRs were able to determine the correct location and make contact with the establishment.

The 1999 MAF Updating Field Test produced a number of results, mostly qualitative. In discussions with Headquarters Staff, Regional Office supervisory staff and the FRs who participated in the 1999 test, a number of issues were defined as needing attention prior to the next major test in 2000. These issues and concerns are outlined below:

1. We must realize that the GUs, for the most part, have limited time and staff resources to assist Census Bureau personnel. We need to allow more time for list acquisition. We plan to enlist the regional office CAUS Geographers to assist us in building working relationships with the GUs in their regions. The CAUS Geographers will be the lead contact with the GUs and offer assistance, when necessary, in the collection of the address lists.

2. Future tests should consider the MUS in an automated format. A general suggestion was to eliminate as much of the paper as possible, especially as more test counties are introduced and the CAUS system becomes more refined. Currently, we are considering preparing the targeting assignments at Headquarters from lists obtained through the CAUS Geographers in the regional offices.

3. The ABI was not an adequate source, according to a number of field staff and observers. Some establishments had misleading address information, phone numbers were incorrect, and some businesses had changed names more than once. The ABI proved to be costly for minimal results. We will not be using it in the future.

4. The County GU offices seemed to be the best source for the lists we wanted. Of the 13 offices we visited, 7 were county offices and 6 of the 7 provided us with lists.

5. Most offices had both electronic and paper files. Of the 12 lists collected, 7 were generated electronically from computer files. In the future, we plan to ask for the files rather than paper lists in order to save the time and burden of both printing and keying these lists. The files will only have to be reformatted.

6. The overall results of the listing and canvassing by county were:

County	Units Added	Units Verified	Units Corrected	Units Deleted
Calvert	573	1007	42	48
Miami	82	2	71	0
Schuylkill	32	241	3	0
Starr	67	0	24	1
Vilas	43	389	36	1
Washington	52	52	93	8
Total	849	1691	269	58

The columns for the most part are explanatory. The Units Verified column are those housing units that were located and found to be correct as indicated on the MAF.

IV. TARGETING WITHIN COUNTY

For this test, we tried to identify a method to efficiently locate new growth within a county since we will not have the resources to update the entire county for every county in America. As we described in Section III, FRs identified Targeted Areas of Growth (TAGs). The TAGs are where the FR actually updated based on the targeting sources. We identified three possible targeting sources: lists from the Permit Address Lists (PALs), FR Observations, and lists from local governmental units (GUs). In addition, Headquarters personnel compared two other targeting sources—Statistical Administrative Records System (StARS) and the Aggregate Targeting Algorithm (ATA)—to assess their usefulness in the future.

The Permit Address Lists (PALs) are a sample of addresses from recently issued permits that were chosen for our current surveys within a county. We thought a PAL could give a representation of where the new addresses in the county are. There are two problems with this source though. First, PALs are not created for every county in the nation. Second, since it is a sample, not all addresses from the permit offices are represented on the lists. It actually contains only about one-third of the addresses. For the two counties in the test with PALs, both FRs found it unprofitable and admitted that they did not use it as a targeting source.

Since the FRs work in the county year round, we asked them to record where they spotted new housing in the county from March to June of 1999, as they completed their other assigned work. They did not make any special visits to complete observations. We combined their observations into a list and gave it to them to help identify TAGs once the test started. The FRs did not find these lists useful. Part of the problem was that not all FRs completed the observations. In some counties, FRs did not record any new units. If this source is going to be effective, the FR observations during the rest of the year will have to be more consistent, so these lists will be more complete. We will not ask them to do this for the next test, since the impact was negligible for the amount of work it would require the FR throughout the year.

There were lists provided by the local GUs—such as planning and zoning offices and the permit offices—for the county. Each county has its own system of recording new housing and each governmental unit's list contained different information. Not all permit lists were complete or accurate. Some counties don't issue permits, forcing the FR to visit each township and city in the county to look for records or a list of permits at that level. The lists from the local governments were the best source of information and the main lists that the FR used during the test, having the most information and being most accurate. Overall, 13 GUs provided us with 1,726 addresses. Of the 6 FRs, 5 used the lists from the GUs as their main source. The other used a real estate office. We have decided that GUs will be one of our main sources for targeting for both counties and within county. We are currently developing a procedure to key the lists in a central area to help make this process more efficient.

Statistical Administrative Records System (StARS) is a database of administrative records from various sources maintained by the Census Bureau's Planning, Research and Evaluation Division (PRED). They use it as a data file of all housing units in the United States. PRED matched the StARS database to addresses in the MAF.

Headquarter's observers were provided with a list of all the nonmatches between StARS and the MAF—as well as all StARS addresses in a sample county—with the addresses sorted by ZIP code or, when available, ZIP+2. The observers canvassed the county separately from the FR to evaluate only StARS and its accuracy. Results were mixed. In some counties, nonmatches were duplicates of units and didn't target areas with new units. In other counties, the nonmatches were new units.

The Aggregate Targeting Algorithm (ATA) was a linear combination of several variables which were supposed to represent growth. The variables represented growth at different levels. For example, some represented growth within a ZIP code while others represented growth within a town or a census block. Several of the variables represented the difference in counts at two points in time for a specific area (e.g., ZIP code, Census block). We felt these would be good indicators of growth. The algorithm and variables were run through a SAS software program to assign each block in the county a ranking. The following shows the actual updating results compared to ATA rankings:

ATA Ranking	Number of Blocks with adds	Number of Units added	Percent of Units added	Average Number of Units added per block
HIGH	30	178	21 %	3.4
MEDIUM	57	468	55 %	7.99
LOW	35	203	24 %	5.71

The results suggest the ATA did not do a good job of predicting areas of high growth. About 33% of the total blocks were ranked high growth but only 21% of adds were in these blocks. More adds were found in low growth blocks and even more adds were found in the medium growth rank. The results are inconclusive though. The FRs did not have enough time to visit all their TAGs. Had they had enough time, the results might have been different. On the other hand the FRs visited the areas where they felt they would find the most growth and they visited 92 medium and low growth blocks versus 30 in high growth. This shows that they saw potential growth 75% of the time in the lower ranked blocks.

We will attempt to analyze parts of the ATA in this upcoming test, since we are using lists from the GUs as a primary source. We will look to see if any one of the variables from the ATA, by itself, is a good indicator of growth.

V. 2000 PLANS

We will be conducting our second test in 2000. List acquisition will begin in May 2000 and field work will begin in September 2000. It will be conducted in 24 counties throughout the country.

The objectives of the 2000 test are more defined and narrow than those of the 1999 field test. Our main objectives for the 2000 test are:

1. Develop a system for making listing and updating assignments to the FRs.
2. Develop a centralized targeting system.
3. Explore additional methods of targeting areas of growth within a county.
4. Test enhancements to the ALMI.
5. Refine the listing and updating procedures.

There are still some issues to be resolved, such as the development of a database of GUs and the actual targeting of counties, but the 200 test is another step closer to the actual production system. Overall, the procedures, resources, and tools have been developed. This test should give us the information needed to fine-tune them.