Census 2000 and the U.S. Postal Service Delivery Sequence Files

Miriam D. Rosenthal[1]
U.S. Census Bureau, Room 2501-2, Washington, DC 20233

Key Words: *address list development, housing unit coverage*

Abstract: There were many sources used to create the Census 2000 address list. The initial sources were the 1990 Address Control File and the November 1997 or earlier U.S. Postal Service Delivery Sequence Files (DSFs). Other sources included addresses provided by census field operations, local and tribal governments, and DSF refreshes. This paper examines the DSFs' contribution to the Census 2000 address list.

Public law paves the way to use DSFs in creation of Census 2000 address list

The United States Congress authorized the U.S. Postal Service to share its address list, the Delivery Sequence File, with the Census Bureau in the Census Address List Improvement Act of 1994. Prior to this law, the Census Bureau did not have access to the postal address list. Instead, the Census Bureau constructed a new list of addresses every ten years by buying commercial address lists. With the advent of the Census Address List Improvement Act of 1994, the Census Bureau could use the postal address list as a building block for a permanent address list.

The Master Address File (MAF) is the name of the Census Bureau's permanent address list. The Census Bureau uses the MAF to conduct censuses and surveys, such as the American Community Survey, a survey designed to supersede the census long form. The initial sources of the MAF were the 1990 Address Control File (the 1990 census address list) and the November 1997 or earlier U.S. Postal Service DSFs. The DSFs contribute only city-style addresses to the MAF because only city-style addresses are geocodable via automated address matching. That is, city-style addresses can potentially be linked to address ranges in the Topologically Integrated Geographic Encoding and Referencing (TIGER) database in order to determine the location of the address, which is required for census field operations.

For Census 2000, the Census Bureau specifies a census universe from the MAF, residential addresses that were geocoded. We called the census address list the Decennial Master Address File (DMAF). In conjunction with Census 2000, we updated the MAF and DMAF based on addresses provided by census field operations, local and tribal governments, and DSF refreshes. See Robin P. Pennington and Cynthia Rothhaas' paper in the 2001 ASA proceedings, "Final Status of Addresses on the Census 2000 Address List: Analysis of the Address List-Building Process" for more details about the DMAF.

U.S. Postal Service Delivery Sequence Files provided for Census 2000

The U.S. Postal Service provided the Census Bureau the following DSFs:
•November 1997 or earlier,
•September 1998,
•November 1999,
•February 2000, and
•April 2000.

The November 1997 or earlier DSFs category refers to a national DSF delivery to the Census Bureau in November 1997 and subnational DSF deliveries of earlier DSFs, such as for the 1995 census test sites. The February 2000 DSF was a file of added addresses only. The other DSFs were complete files: added addresses and addresses that hadn't changed.

In this paper, the addresses we track are residential housing units. For the November 1997 or earlier DSFs and September 1998 DSF, we consider housing units with residential status unknown to be residential as well. Residential status unknown addresses are addresses whose residential status was overwritten by later DSF deliveries.

There were 116,550,536 addresses on one or more DSFs. Eighty-five percent of the addresses were in Census 2000, that is, were enumerated as either occupied or vacant. Addresses were not enumerated because they couldn't be geocoded, weren't verified by two or more operations, or were duplicates discovered in a data processing operation undertaken to identify duplicates.

Matching the DSF addresses to the 1990 Address Control File (ACF), 64% of the DSF addresses were on the 1990 ACF. Of the DSF addresses that matched to the 1990 ACF, 97% were in Census 2000. Of the DSF

addresses that didn't match the 1990 ACF, 63% were in Census 2000.

Individually, the DSF deliveries provided the following number of addresses:
•November 1997 or earlier DSFs: 106,792,959 addresses; 87% of the addresses were in Census 2000.
•September 1998 DSF: 100,407,869 addresses; 94% of the addresses were in Census 2000.
•November 1999 DSF: 103,281,784 addresses; 94% of the addresses were in Census 2000.
•February 2000 DSF (adds only): 985,365 addresses; 53% of the addresses were in Census 2000.
•April 2000 DSF: 103,969,951 addresses; 94% of the addresses were in Census 2000.

The November 1997 or earlier DSFs had 103,657,220 residential status unknown addresses. The September 1998 DSF had 193 residential status unknown addresses.

Definition of new, added, and deleted addresses
To understand better the contributions of each DSF, we defined DSF addresses in three more ways.

Addresses that were not on any preceding files, we defined as "new." In sequence, the files were:
•the 1990 ACF, the November 1997 or earlier DSFs,
•the September 1998 DSF,
•the November 1999 DSF,
•the February 2000 DSF,
•and the April 2000 DSF.

Addresses that were not on the immediately preceding file and were on the current file, we called "adds." In the case of the November 1997 or earlier DSF adds, we compared the 1990 ACF and the November 1997 or earlier DSFs. For the other adds, we compared the addresses on adjacent DSFs. Since the February 2000 DSF was a file limited to added addresses and the April 2000 DSF was a complete delivery, adds for the April 2000 DSF considered both the February 2000 and the November 1999 DSF as adjacent DSFs.

Addresses that were on the immediately preceding file and not on the current file, we designated "deletes." The November 1997 or earlier DSF deletes were addresses on the 1990 ACF (and one or more of the other DSFs) and not on the November 1997 or earlier DSFs. All other deletes compared addresses on adjacent DSFs. The February 2000 DSF was a file of adds only, so it had no deletes. The February 2000 DSF adds-only delivery also necessitated a modification of the definition of the April 2000 DSF deletes to include a comparison of the April 2000 DSF to both the February 2000 and the November 1999 DSFs.

New addresses
The five DSFs provided the following number of new addresses:
•The November 1997 or earlier DSFs provided 32,219,712 new addresses.
•The September 1998 DSF provided 3,000,454 new addresses.
•The November 1999 DSF provided 4,274,404 new addresses.
•The February 2000 DSF provided 836,431 new addresses.
•The April 2000 DSF provided 1,121,988 new addresses.

Added and deleted addresses
The five DSFs provided the following number of adds:
•The November 1997 or earlier DSFs provided 32,219,712 adds. By definition all new addresses were adds for this delivery.
•The September 1998 DSF provided 3,260,202 adds. Of the adds, 92% were new addresses.
•The November 1999 DSF provided 4,850,176 adds. Of the adds, 88% were new addresses.
•The February 2000 DSF provided 903,556 adds. Of the adds, 93% were new addresses.
•The April 2000 DSF provided 1,275,379 addresses that were not on the February 2000 or November 1999 DSF. Of the adds, 88% were new addresses.

To sum up the relationship between adds and new addresses, between 8-12% of the addresses were adds but were not new addresses. Adds are addresses not on the immediately preceding file. New addresses are addresses not on any preceding file.

The four DSFs that were complete files had the following number of deletes:
•The November 1997 DSF or earlier had 524,300 deletes.
•The September 1998 DSF had 9,645,292 deletes.
•The November 1999 DSF had 1,976,261 deletes.
•The April 2000 DSF had 1,490,768 deletes.

A graphic comparison of the adds and deletes helps to visualize the contribution of each DSF.
The November 1997 or earlier DSFs provided 76% of

the adds, and the September 1998 DSF provided 71% of the deletes. The November 1997 or earlier DSFs provided the first update of the 1990 ACF, so it would be expected to contribute a lot of addresses. Of the September 1998 DSF deletes, 7,414,334 or 77% were residential status unknown on the November 1997 or earlier DSFs. Again, the February 2000 DSF had no deletes because the delivery was limited to adds. The

DSF Adds and Deletes
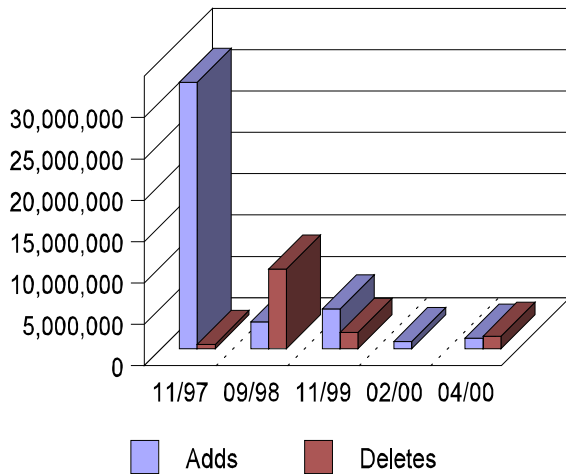
Adds   Deletes

April 2000 DSF had almost the same number of deletes as adds. By April 2000, when three DSF deliveries occurred within five months, the number of changes from delivery to delivery had subsided.

Breakdown of adds and deletes by size of structure
Overall, 73% of adds and 61% of deletes were single unit structures. Single unit structures have only one unit at the basic street address. Attached homes with different basic street addresses are single unit structures. Attached homes with the same basic street address and different unit designations are a multi-unit structure. For each DSF delivery, we added a higher percentage of single units than we deleted (Table 1, note: tables are at end of paper).

Blueline designation
The blueline demarcated the strategy used to construct the census address list. For areas "inside the blueline," the Census Bureau created the census address list by extracting addresses from the MAF. For areas "outside the blueline," the Census Bureau created and updated the census address list from scratch. Urban update/leave areas, a special enumeration strategy where questionnaires were hand-delivered in areas with city-style addresses, were inside the blueline but had the final address updating at the time of questionnaire delivery.

All other areas inside the blueline had their census questionnaires mailed, so they didn't have the opportunity of field updates of the address list. Areas outside the blueline had questionnaires hand-delivered and address updating occurred concurrently. Not all addresses had the geography needed to classify them as inside or outside the blueline. In this section, we limit the analysis to addresses classified as either inside or outside the blueline.

Most, 79%, of the enumerated addresses in the United States were inside the blueline. So, a greater percentage of the DSF adds and deletes should also be inside the blueline, which is the case, 73% and 87% respectively. The high rate of deletes in September 1998 may be associated with the large influx of adds in November 1997, since by November 1999 the delete rate levels off to 72%, similar to the overall 73% add rate of DSF addresses inside the blueline (Table 2).

Original source for areas inside the blueline
The census evaluation staff created a variable to indicate the original source of an address in the MAF. There could be multiple original sources because of overlapping address list updating operations. Also the original source designation didn't preclude later operations from identifying the address. In this analysis, we'll look at the original source for areas inside the blueline, where the DSF served as a building block for the census address list. Multiple original sources between a census field operation or local or tribal government updating process and a DSF will be attributed to the DSF.

Tables 3 and 4 present the original source for adds and deletes. The first couple of DSF deliveries had only the 1990 ACF and the DSF deliveries as original sources. Beginning in
November 1999 for adds and April 2000 for deletes, the local update of census addresses (LUCA) and block canvassing provided addresses. The lag time for these operations to appear as original sources for the deletes reflects the amount of time for a non-DSF original source address to cycle (be added and deleted) through the DSF.

Under the LUCA program, a provision of the Census Address List Improvement Act of 1994, participating local and tribal governments checked the Census Bureau's address list, identifying adds, deletes, and corrections. LUCA participation was voluntary, so coverage was not complete. In block canvassing, census field workers reviewed the census address list inside the blueline, suggesting adds, deletes and corrections to the list. Block canvassing occurred from January to May 1999. The address list reviewed in block canvassing had

addresses on it from the 1990 ACF, the November 1997 or earlier DSFs, and the September 1998 DSF.

DSF addresses on the MAF that were delivered to the census address list

The Census Bureau delivered to the DMAF DSF addresses that were geocoded and not deleted by two or more operations. Such addresses were said to be "DMAF deliverable." Overall, 70% of the adds and 33% of the deletes were DMAF deliverable.

The lower percentages of DMAF deliverable adds for DSF deliveries near Census Day, April 1, 2000, may be attributable to having less time to geocode the addresses. The September 1998 DSF had the highest quality for an individual DSF with the greatest percent of DMAF deliverable adds, 82%, and the lowest percent of DMAF deliverable deletes, 16% (Table 5).

DSF addresses on the MAF that were in Census 2000

Overall, 64% of the adds and 18% of the deletes were in Census 2000. The September 1998 DSF had the highest quality for an individual DSF with the greatest percent of adds in Census 2000, 77%, and the lowest percent of deletes in Census 2000, 7% (Table 6).

What percent of DMAF deliverable adds and deletes made it into Census 2000?

Overall, 91% of the DMAF deliverable adds and 54% of the DMAF deliverable deletes (but not ineligible) were in Census 2000. Once an add or delete was DMAF deliverable, it had a good chance of being in the census. This implies that the presence of an address on the DSF doesn't guarantee that it is an occupied or vacant housing unit. Conversely, addresses deleted for a particular DSF delivery may still exist.

The percent of DMAF deliverable adds in Census 2000 decreased to 81% as the DSF delivery neared Census Day (Table 7). By this point in the census process, addresses weren't subjected to the same extensive review as earlier addresses that had been submitted to block canvassing and LUCA.

A delete on the November 1997 or earlier DSFs was on the 1990 ACF but not on the November 1997 or earlier DSFs, different than other deletes which disappear from being on a previous DSF. The November 1997 or earlier DSF DMAF deliverable deletes had the highest rate for deletes in Census 2000, 91% (Table 7). This was another instance of a 1990 ACF address being a quality address, and, in this instance, not available on the DSF.

Conclusions

The DSF was an important source of addresses for Census 2000. The creation of a permanent address list required many updates. Following the initial creation, the file took shape as the Census Bureau received more deliveries of the DSF. By the time the census took place, the file was stabilizing.

Of the 116,550,536 addresses on one or more DSFs, 85% were in Census 2000. The best indicator for whether a DSF address would be in Census 2000 was whether it was on the 1990 census address list. Almost all, 97% of the DSF addresses that were on the 1990 ACF were in Census 2000.

As expected, the November 1997 DSF or earlier had the most addresses not on the previous file (the 1990 ACF), 32,219,712 adds. The September 1998 DSF had the most deletes, over 9 million addresses; 77% of the deletes were residential status unknown on the November 1997 DSF or earlier. By April 2000, the number of adds and deletes were comparable, 1,275,379 and 1,490,768, respectively.

Seventy-three percent of the adds were single units compared to 61% of the deletes. Seventy-three percent of the adds were inside the blueline; whereas 87% of the deletes were inside the blueline. Until the November 1999 DSF, the original sources were files: the 1990 ACF and the DSFs. Starting with the November 1999 DSF, block canvassing and LUCA sometimes identified the address before any DSF.

Addresses failed to be in Census 2000 because they were found to be duplicates, couldn't be geocoded, or were verified as nonexistent by two or more operations. Overall, 70% of the adds were DMAF deliverable, and 91% of the DMAF deliverable adds were in Census 2000. Adds that couldn't be geocoded were not DMAF deliverable. The percent of adds that were DMAF deliverable might have been even lower. Census field operations, such as block canvassing, improved DMAF deliverability by providing physical locations for some of the DSF addresses missing this information.

DSF deliveries close to Census Day had less time for addresses to be geocoded and fewer opportunities to be field verified, resulting in decreased quantity and quality of DSF addresses on the DMAF.

Less than half of the deletes, 33%, were DMAF deliverable. Of the 33% of DMAF deliverable deletes, 54% were in Census 2000. Since more than half of the DMAF deliverable deletes were in Census 2000, it was good that we delivered these addresses to the DMAF. The addresses were delivered to the DMAF because an address remained DMAF-eligible until more than one operation identified it as a delete.

We defined both adds and deletes by looking at consecutive DSFs. Some addresses popped in and out of the DSFs, as evidenced by not all, but rather 88% to 93%, of the added addresses being new addresses. While the DSFs do provide a sizeable number of addresses in the census, the DSFs are not 100% DMAF deliverable. Without being able to put a postal address on a census map, the postal address was not DMAF deliverable and so didn't make it into the census.

Regular DSF deliveries between censuses combined with regular use of the MAF by ongoing surveys, such as the American Community Survey, should make the DSF a more familiar and predictable file.

Reference:Pennington, Robin P. and Cynthia Rothhaas. "Final Census Status of Addresses on the Census 2000 Address List: Analysis of the Address List-Building Process," ASA JSM Proceedings, August 2001.

Table 1. Number and percent of single unit adds and deletes by DSF delivery

DSF	# of single unit adds	% of adds	# of single unit deletes	% of deletes
11/97	23,252,944	72%	259,092	49%
09/98	2,541,031	78%	5,822,366	60%
11/99	3,682,642	76%	1,247,763	63%
02/00	676,366	75%	----	----
04/00	909,460	71%	951,086	64%

Table 2. Number and percent* of adds and deletes inside the blueline by DSF delivery

DSF	# of adds inside the blueline	% of adds	# of deletes inside the blueline	% of deletes
11/97	23,254,139	75%	442,861	85%
09/98	1,798,937	61%	8,361,970	92%
11/99	2,659,569	68%	1,289,642	72%
02/00	410,054	65%	----	----
04/00	700,771	76%	676,147	71%

*This analysis is limited to addresses classified as either inside or outside the blueline.

Table 3. Original Source (OS) for adds by DSF delivery

OS / DSF	1990 ACF	11/97 DSF	09/98 DSF	11/99 DSF	02/00 DSF	04/00 DSF	LUCA	Block canvassing	LUCA and block canvassing
11/97	NA	100%	—	—	—	—	—	—	—
09/98	12%	NA	88%	—	—	—	—	—	—
11/99	11%	5%	NA	31%	—	—	9%	32%	12%
02/00	9%	0%	2%	NA	67%	—	5%	14%	3%
04/00	10%	4%	1%	NA	NA	52%	7%	21%	5%

— = not considered, NA=not applicable

Table 4. Original Source (OS) for deletes by DSF delivery

OS / DSF	1990 ACF	11/97 DSF	09/98 DSF	11/99 DSF	02/00 DSF	04/00 DSF	LUCA	Block canvassing	LUCA and block canvassing

11/97	100%	NA	—	—	—	—	—	—	—
09/98	9%	91%	NA	—	—	—	—	—	—
11/99	58%	35%	7%	NA	—	—	—	—	—
04/00	58%	25%	3%	7%	3%	NA	1%	2%	1%

— = not considered, NA = not applicable

Table 5. Number and percent of DMAF deliverable adds and deletes by DSF

DSF	# of DMAF deliverable adds	% of adds	# of DMAF deliverable deletes	% of deletes
11/97	21,941,529	68%	505,270	96%
09/98	2,684,904	82%	1,497,288	16%
11/99	3,737,517	77%	1,481,143	75%
02/00	561,989	62%	----	----
04/00	833,056	65%	1,072,522	72%

Table 6. Number and percent of adds and deletes in Census 2000 by DSF

DSF	# of adds in Census 2000	% of adds	# of deletes in Census 2000	% of deletes
11/97	20,307,368	63%	461,383	88%
09/98	2,495,645	77%	640,938	7%
11/99	3,123,386	64%	820,814	42%
02/00	463,266	51%	----	----
04/00	676,601	53%	525,265	35%

Table 7. Percent of DMAF deliverable adds and deletes in Census 2000 by DSF

DSF	% of DMAF deliverable adds in Census 2000	% of DMAF deliverable deletes in Census 2000
11/97	93%	91%
09/98	93%	43%
11/99	84%	55%
02/00	82%	----
04/00	81%	49%