



U.S. Census Bureau Geography Division

Final Report



GSS-I

Address Summit Pilot:

Data Sharing - Local, State, U.S. Postal Service (USPS), and Census Bureau

March 22, 2013

Version V1.0

Approval Log

This Pilot Project Plan has been reviewed and approved for use.

| Date | Version | Name & Area Represented | Signature | Description of Major Changes |
|------|---------|--|-----------|------------------------------|
| | V1.0 | Timothy F. Trainor Chief, Geography Division | | |
| | V1.0 | Deirdre Dalpiaz Bishop Mentor, Geographic Operations Advisor | | |
| | V1.0 | Paul Riley Mentor, Office of the Geographic Advisor | | |
| | V1.0 | Andrew Stanislaw, Geographic Areas Branch, Geography Division | | |
| | V1.0 | Michael Ratcliffe Mentor, ADC for Geocartographic Products and Criteria | | |

Table of Contents

| | |
|---|-----------|
| 1. Overview | 2 |
| 2. Introduction | 4 |
| 3. Objectives | 6 |
| 4. Methodology | 7 |
| 5. Application | 9 |
| 6. Discussion and Findings | 12 |
| 7. Conclusions | 21 |
| 8. Acknowledgements..... | 23 |
| Attachment A: Data Dictionary | 24 |
| Attachment B: Acronyms | 26 |
| Attachment C: References..... | 27 |
| Attachment D: File Requirements..... | 28 |

1. Overview

The Geographic Support System Initiative (GSS-I) is an integrated program of improved address coverage, continual spatial feature updates, and enhanced quality assessment and measurement. It will allow for a targeted, rather than full, address canvassing during 2019 in preparation for the 2020 Census. A pivotal piece of the GSS-I will be working with federal, state, local, and tribal governments, as well as other key stakeholders, to create an address list that is suitable for this purpose.

In preparation, the Census Bureau's Geography Division hosted a Census Address Summit in September 2011. Forty-four external experts in the fields of address list development, maintenance, and sharing attended the summit.

The goals of the Address Summit were as follows:

1. To educate Census Bureau partners about the GSS-I and the benefits of conducting a targeted address canvassing.
2. To gain a common understanding regarding the definition of an address.
3. To learn how Census Bureau partners are collecting, utilizing, and maintaining addresses.
 - i. What industry standards are they following?
 - ii. What are their best practices?
 - iii. What are their major challenges?
 - iv. What are their current practices for data sharing?
4. To brainstorm about potential pilot projects that will contribute to the improved quality of the Census Bureau's Master Address File.

During the Address Summit, attendees proposed six pilot projects. Geography Division determined that five pilots were feasible and began development in January 2012. One pilot was determined to be duplicative of other efforts occurring within the division and therefore, was not included. The goal of each pilot project follows:

1. Address Authority Outreach and Support for Data Sharing Efforts Pilot

To research and develop an approach for identifying and creating an inventory of address authorities which facilitates address data sharing activities and provides guidance on overcoming barriers (legal/policy) at the local level.

2. Federal Geographic Data Committee (FGDC) Address Standards and Implementation Pilot

To educate local authorities on the benefits, use, and implementation of the FGDC's United States Thoroughfare, Landmark, and Postal Address Data Standard (FGDC Address Standard).

3. Project for Federal/State/Tribal/Local Address Management Coordination Pilot

To create a formalized model to allow for the development, maintenance, and bi-directional (state-local-tribal and state-federal) sharing of high quality multiple use address data.

4. Data Sharing - Local, State, U.S. Postal Service (USPS), and Census Pilot

To create an address data exchange model that will allow for address data sharing between local governments, state governments, the USPS, and the Census Bureau. It will provide a business process that increases the accuracy and coverage of local government address lists, while streamlining the process of sharing those externally.

5. Hidden/Hard to Capture Addresses Pilot

To determine how to capture hidden and hard to capture addresses in the Master Address File and make them useful for enumeration purposes.

This report focuses specifically on the Data Sharing - Local, State, U.S. Postal Service (USPS), and Census Pilot.

2. Introduction

The partnerships between the Census Bureau, the United States Postal Service (USPS) and local and state government are essential for maintaining and updating a comprehensive national address list. The Census Bureau collects addresses from the USPS on a bi-annual basis in the form of the Delivery Sequence File (DSF), and from local governments through the Local Update of Census Addresses (LUCA) program, New Construction program, and field operations. The USPS, like the Census Bureau, depends on workers in the field to update addresses, while also increasingly working with local government to improve their address lists.

The primary goal of the Data Sharing—Local, State, USPS, and Census Bureau Pilot was to examine ways to improve address data sharing and build on existing partnerships with local and state government and the USPS.

Starting with preparations for Census 2000, the Census Bureau has collected address data from local governments through the once-a-decade LUCA program. The LUCA program allows local and state governments the opportunity to review and update city-style addresses on the Census Bureau's address list and challenge address counts in census blocks. The Census Bureau benefits from local review of address data to help ensure an accurate address list for conducting the decennial census. Local and state governments benefit from participation in LUCA by helping to ensure a more accurate census.

The USPS currently shares address data with the Census Bureau on a bi-annual basis in the form of the Delivery Sequence File (DSF). Residential addresses in the DSF are incorporated into the MAF and create the address framework for the decennial census, population estimates, the American Community Survey, and other current surveys. Currently the address updates received by the Census Bureau are primarily the result of postal staff recording updates they encounter in the field and then transmitting these updates for inclusion in the USPS DSF file. Title 13 restricts the Census Bureau from sharing any address data with the USPS.

The USPS works with local addressing authorities to improve address quality through the County Program. Local, county or state governments submit address lists to the USPS and the USPS provides local addressing authorities, free of charge, with:

- matched, standardized addresses,
- missing and invalid secondary addresses,
- congressional variances, and a
- no match list (address records that do not match the Address Management System (AMS) database)

However, after the USPS runs the matching software, there are restrictions that limit their ability to share address data with local and state governments. Restrictions under Title 39 limit the USPS from sharing primary addresses that are missing from local address lists.

Another option that allows the USPS to share primary addresses is the USPS Computerized Delivery Service (CDS) program. The CDS provides customers with a complete list of confirmed addresses and E-911 converted addresses. In order to be eligible, customers must obtain or possess an address list that contains at least ninety percent of the current possible delivery addresses within a ZIP code and/or address group for which they wish to receive CDS updates. In addition, the CDS identifies addresses marked as vacant or seasonal, and provides customers with frequent updates. The CDS program could be a valuable resource for local governments in updating and maintaining local address lists. However, the ninety percent match rate is difficult for local addressing authorities to reach and requires significant investment of time and/or money.

When the USPS and local address lists improve, the Census Bureau benefits through both an improved DSF and when a local government participates in the LUCA program. The challenge for this pilot was to develop strategies that build on these existing data exchanges, create strategies for data sharing that more directly benefit all stakeholders, and meet the goal of continuous address updates.

3. Objectives

The objectives of the Data Sharing – Local, State, USPS, and Census team were to:

1. Identify and evaluate existing address data sharing models between local and state governments, the USPS and the Census Bureau.
2. Evaluate, understand and document current constraints of sharing address data and identify solutions for overcoming current constraints.
3. Develop recommendations and requirements for data sharing models that increase the accuracy and coverage of local and state government, USPS, and Census Bureau address lists.

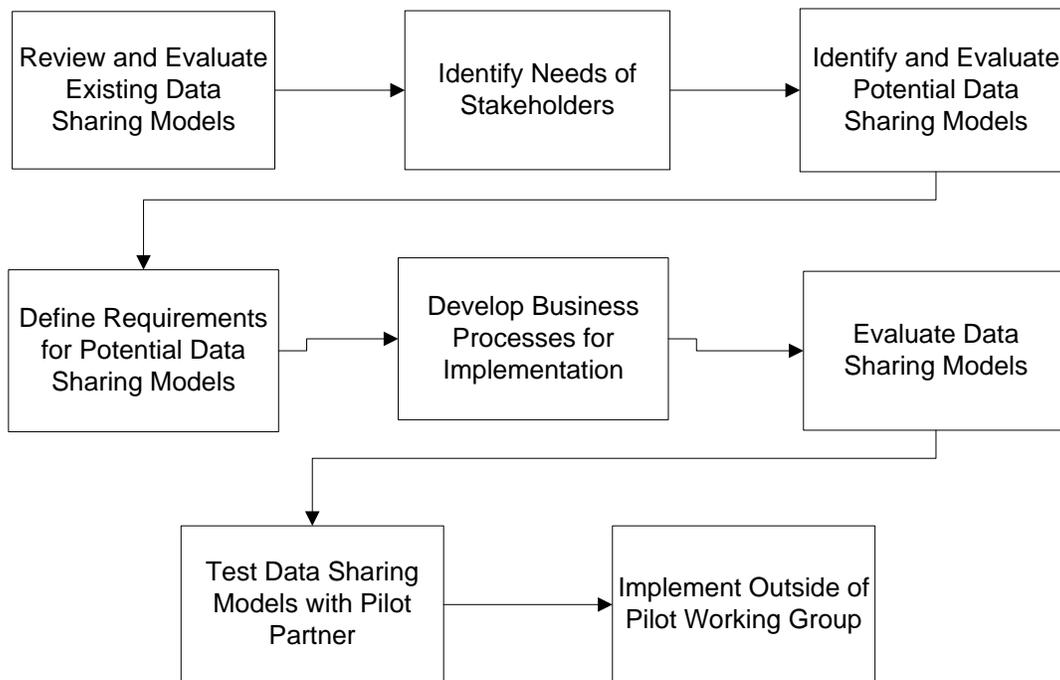
The In-Scope activities for the Data Sharing – Local, State, USPS, and Census team were to:

1. Establish a working group of participants, stakeholders and subject matter experts.
2. Research and evaluate existing address data sharing models between local and state governments, the USPS and Census Bureau and provide recommendations for improved data accuracy to Geography Division management.
3. Evaluate and document current constraints on sharing address data and explore strategies to deal with the constraints.
4. Evaluate the requirements of all the working group participants and the Census Bureau to ensure that the proposed pilot will meet and satisfy the necessary requirements.
5. Develop a business process that increases the accuracy and coverage of local government address lists, while streamlining the process of sharing those externally.
6. Develop a plan to communicate business objectives and partnership benefits with external partners.
7. Refine the proposed model based on feedback and experience gained from the pilot project.
8. Integrate the model within existing partners and seeking adoption of the model by other stakeholders outside the working group.
9. Identify and evaluate activities (e.g., NextGen E-911, Broadband Initiative) underway within other groups (NSGIC, NAPSG, USPS, etc.) and ensure coordination with those activities where appropriate.

4. Methodology

The project plan workflow consisted of eight high level steps to meet the goals and objectives of the pilot.

Figure 1: High Level Workflow



1. Review and Evaluate Existing Data Sharing Models. The team researched existing data sharing practices between state and local government, the USPS, and the Census Bureau. The team examined the benefits of current practices and explored how existing tools can be utilized and exploited. The team also identified and evaluated existing constraints and barriers to data sharing and explored ways to overcome these constraints and barriers.
2. Identify the needs of each stakeholder. Local and state governments, the USPS, and the Census Bureau each have unique needs. The team conducted a needs assessment to determine the project requirements.
3. Identify and recommend potential data sharing strategies.
4. Define requirements for the data sharing model based on needs of pilot participants.
5. Develop business processes for implementing the data sharing model. The team defined necessary resources and worked towards processes that share the benefits and burdens.

6. Test and evaluate data sharing model. The team tested the data sharing model with pilot participants.
7. Evaluate the data sharing model. The team evaluated success of the model based on requirements.
8. Implement the model beyond pilot participants. If successful, the final step is to expand the data sharing model to include external partners outside of the initial pilot participants.

5. Application

The pilot team identified and selected Anne Arundel and Talbot Counties and Laurel City in Maryland to participate in the address data sharing pilot project. In order to compare results, the team requested that the counties submit similar or identical address files to both the USPS County Project and Census Bureau. The USPS and the Census Bureau processed files from Anne Arundel County, Talbot County and the City of Laurel, Maryland.

The USPS processed the files according to the County Project standards and provided the local participant with the following feedback:

- Matched, standardized addresses where the two databases (the USPS AMS and the local government) matched perfectly
- No Match List – this file identifies issues such as typos or incorrect suffixes
- Missing Secondary Addresses – structures that require the use of apt., unit, condo, suite, etc.
- Invalid Secondary Addresses - the USPS defines invalid secondary addresses as any secondary address discrepancy (i.e., County has units 1,2,3 and AMS has units A,B,C)
- Congressional variances
- County variances

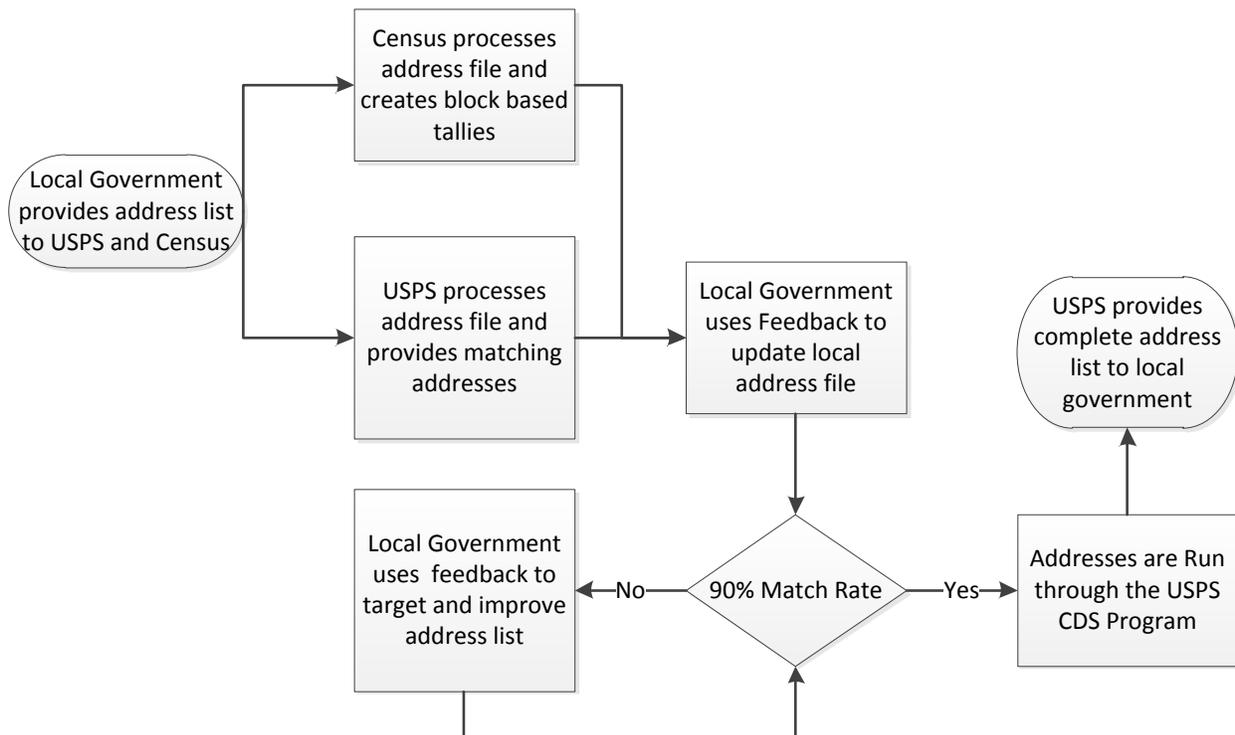
The Census Bureau processed and matched the address files and provided the local participant with the following feedback:

- Total Count of Census Addresses by place (incorporated place and CDP)
- Count of Matched Addresses by place
- No Match Count by place – the number of local addresses did not match the Census MAF
- Local match rate – the percentage of partner supplied address records that matched the MAF, divided by the total count of partner supplied address records.
- MAF match rate – the percentage of partner supplied address records that matched the MAF, divided by the total count of address records in the MAF (filtered to include only valid residential addresses).

The Census Bureau initially proposed providing feedback at the census block level, but was unable to do so due to Title 13 constraints. The Census Bureau also proposed providing a count of missing and invalid secondary addresses. This was not feasible given the relatively short duration of the pilot project and the additional programming resources required. Secondary or sub-address analysis is recommended for future data matching.

In this model, we worked directly with state and local partners. The local government posted its data file on an FTP site for the Census Bureau and the USPS each to download. The USPS and the Census Bureau provided the local partners with a preferred file format, but did not require that the local partner file meet the exact formatting requirements. This created some issues with formatting, particularly for the USPS and its need to run the data through their established software. As a result, the USPS asked the Census Bureau to assist the USPS by preprocessing address files supplied by local governments. This request was a high priority for the USPS and its ability to process local address files.

Figure 2: Data Sharing Workflow



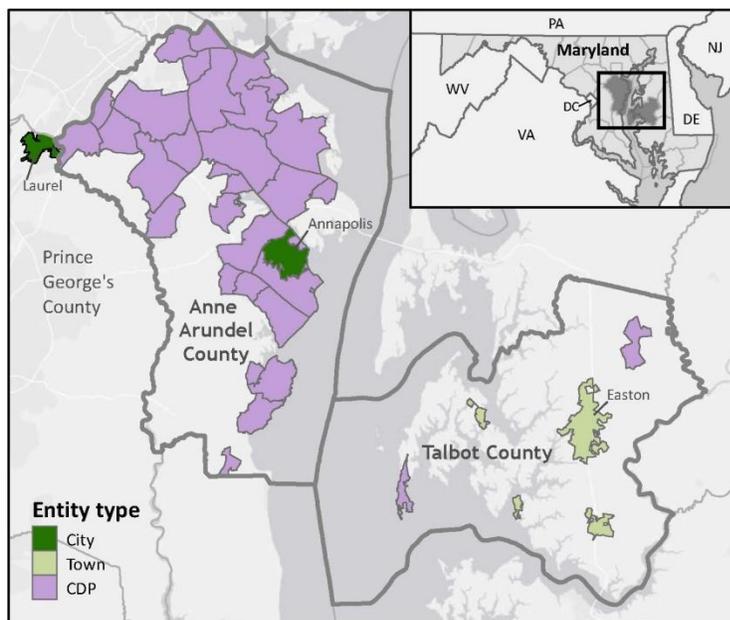
An initial goal of the pilot was to assist county governments in reaching a 90 percent match rate. The pilot went as far as providing feedback to county governments that could assist local partners in moving towards this goal. A potential next step is for counties or places to qualify for the USPS CDS program. To do so, a county or place must demonstrate that it has an address list containing at least 90 percent of the current possible delivery addresses in a zip code and/or address group. Once the local match rate is determined to be at or above 90 percent, a county or place may submit its address list to the CDS program. The USPS processes the updated address file under the CDS program and provides the local government with a complete list of addresses, including E-911 converted addresses, and various indicators such as vacancy, seasonal occupancy, and whether the address is a new delivery point. Under the CDS program, local

participants also are eligible for biweekly updates that list any changes to the USPS address database that have occurred.

6. Discussion and Findings

The Census Bureau and the USPS processed and matched address files provided by Anne Arundel County, Talbot County and the City of Laurel. The Census Bureau analyzed whether the data from local partners matched the Master Address File (MAF) in terms of formatting and structure, as well as numbers of addresses. The USPS ran the local address files through the County Project to determine if the local records matched data in the USPS Address Management System (AMS) database. The USPS worked with local officials and the local AMS office in Baltimore to receive address records in a format consistent with the County Project standards.

Figure 3: Pilot study area



The analysis and discussion of the data focuses on three match rates, two from the Census Bureau and one from the USPS. The Census Bureau match rates were calculated at the census block level; however, only place level geography is published in this report due to Title 13 restrictions. The USPS match rate is published at the county level. The first match rate, referred to as the Local Match Rate, compares the format of the local address file with the MAF. The Local Match Rate does not compare the count of addresses in the MAF with the total count. It does not measure coverage. For example, if a local government submits only one address, and the address matches an address in the MAF, the Local Match Rate would be 100 percent. The second match rate, referred to as the MAF Match Rate, measures both the format quality and coverage of the local address file. The third match rate, referred to as the USPS Match Rate, measures the format quality of the partner supplied address file compared to the AMS database.

The USPS Match Rate does not indicate level of coverage. Therefore, the USPS Match Rate is most similar and comparable to the Local Match Rate.

Local Match Rate

- Total count of partner supplied address records that matched the MAF, divided by the total count of partner supplied address records.
- All components of the local address must match the MAF in order to be identified as a match.
- Local address records that matched most elements of an address in the MAF were identified as an equivocated match (e.g., primary address matched but secondary address did not). Address records identified as equivocated matches were not included in the total count of matched records nor the total count of partner supplied records. The analysis below makes the case that equivocated matches should be included and counted as a match in the future.
- A high local match rate indicates that addresses in the local file match addresses stored in the MAF. A low local match rate likely points to some inconsistency in format. For this pilot, above a ninety percent local match rate is considered high and below seventy percent is considered low. The local match rate does not necessarily indicate the level of local or Census Bureau address coverage.

MAF Match Rate

- Total count of partner supplied address records that matched the MAF, divided by the total count of address records in the MAF.
- All components of the local address must match the MAF in order to be identified as a match.
- Local address records that matched most elements of an address in the MAF were identified as an equivocated match (e.g., primary address matched but secondary address did not). Address records identified as equivocated matches were not included in the total count of matched records nor the total count of partner supplied records. The analysis below makes the case that equivocated matches should be included and counted as a match in the future.
- A high MAF match rate indicates that the local file matches the MAF in both format and coverage. A low MAF match rate indicates that there are addresses in the MAF that were either missing in the local file or could not be matched. A low MAF match rate can also be due to incomplete or missing secondary/sub-addresses. For this pilot, above a ninety percent MAF match rate is considered high and below seventy percent is considered low.

USPS Match Rate

- Total count of partner supplied address records that exactly matched the AMS, divided by the total count of partner supplied address records.
- All components of the local address must match the AMS in order to be identified as a match.
- Address records are defined as a match, a no match, or a missing/invalid secondary address.
- The no match list identifies the number of records where there was no match or records that contain issues such as typos or incorrect suffixes.
- The total records received are the count of local address records that the USPS processed. This number differs from the number of records processed by the Census Bureau due to formatting changes and additional cleansing of the address records by local partners.

Anne Arundel County Results and Analysis
Table 1 – Census Bureau Local and MAF Match Rates, Anne Arundel County

| Census FIPS Code | Place Name | MAF Units | Local Address Units | Local Address Match | No Address Match | Local Match Rate | MAF Match Rate |
|-------------------------|---------------------|------------------|----------------------------|----------------------------|-------------------------|-------------------------|-----------------------|
| 38500 | Highland Beach | 76 | 76 | 76 | 0 | 100.00% | 100.00% |
| 38025 | Herald Harbor | 1178 | 1193 | 1173 | 20 | 98.32% | 99.58% |
| 02025 | Arden on the Severn | 789 | 783 | 782 | 1 | 99.87% | 99.11% |
| 22050 | Deale | 2232 | 2257 | 2188 | 69 | 96.94% | 98.03% |
| 66400 | Riva | 1590 | 1563 | 1543 | 20 | 98.72% | 97.04% |
| 60475 | Pasadena | 9434 | 9538 | 9137 | 401 | 95.80% | 96.85% |
| 51575 | Mayo | 3475 | 3355 | 3325 | 30 | 99.11% | 95.68% |
| 44975 | Lake Shore | 7580 | 7341 | 7238 | 103 | 98.60% | 95.49% |
| 12912 | Cape St. Claire | 3427 | 3260 | 3212 | 48 | 98.53% | 93.73% |
| 25050 | Edgewater | 4145 | 3896 | 3863 | 33 | 99.15% | 93.20% |
| 02275 | Arnold | 8942 | 8357 | 8294 | 63 | 99.25% | 92.75% |
| 66850 | Riviera Beach | 5280 | 4907 | 4869 | 38 | 99.23% | 92.22% |
| 01635 | Annapolis Neck | 4958 | 4577 | 4553 | 24 | 99.48% | 91.83% |
| 71200 | Severna Park | 14626 | 13592 | 13418 | 174 | 98.72% | 91.74% |
| 20875 | Crofton | 10800 | 10029 | 9861 | 168 | 98.32% | 91.31% |
| 30650 | Friendship | 166 | 160 | 151 | 9 | 94.38% | 90.96% |
| 31350 | Gambrills | 1115 | 1077 | 1010 | 67 | 93.78% | 90.58% |
| 10475 | Brooklyn Park | 5963 | 5598 | 5332 | 266 | 95.25% | 89.42% |
| 71150 | Severn | 17064 | 15617 | 14978 | 639 | 95.91% | 87.78% |
| 42550 | Jessup | 452 | 415 | 393 | 22 | 94.70% | 86.95% |

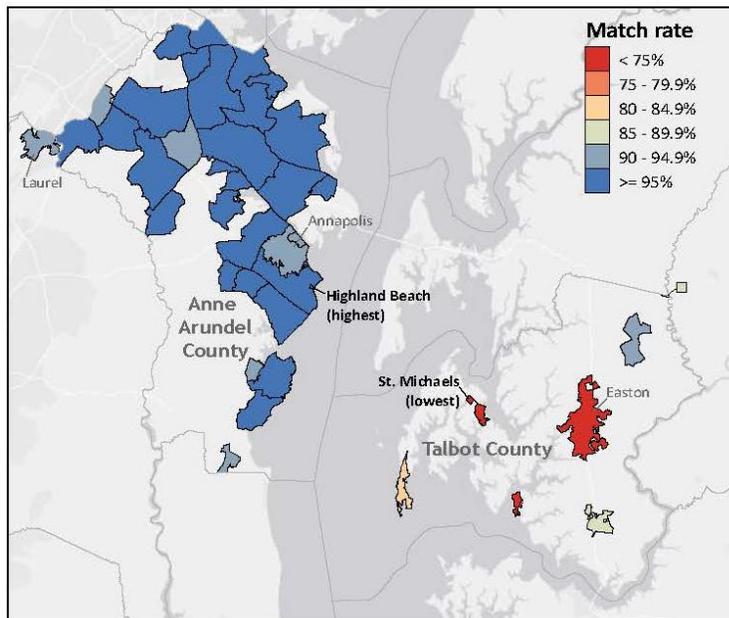
| | | | | | | | |
|-------|-----------------------------------|---------------|---------------|---------------|-------------|---------------|---------------|
| 21075 | Crownsville | 683 | 560 | 542 | 18 | 96.79% | 79.36% |
| 47125 | Linthicum | 4678 | 3787 | 3675 | 112 | 97.04% | 78.56% |
| 28075 | Ferndale | 7029 | 5315 | 5262 | 53 | 99.00% | 74.86% |
| 58300 | Odenton | 16162 | 12222 | 11965 | 257 | 97.90% | 74.03% |
| 29400 | Fort Meade | 2997 | 2283 | 2200 | 83 | 96.36% | 73.41% |
| 32650 | Glen Burnie | 30157 | 22295 | 21859 | 436 | 98.04% | 72.48% |
| 31275 | Galesville | 331 | 257 | 233 | 24 | 90.66% | 70.39% |
| 60325 | Parole | 9882 | 6791 | 6729 | 62 | 99.09% | 68.09% |
| 51075 | Maryland City | 7087 | 4829 | 4795 | 34 | 99.30% | 67.66% |
| 01600 | Annapolis | 20075 | 12822 | 12054 | 768 | 94.01% | 60.04% |
| 55050 | Naval Academy | 143 | 13 | 12 | 1 | 92.31% | 8.39% |
| | | | | | | | |
| | Total Places/CDP | 202516 | 168765 | 164722 | 4043 | 97.60% | 81.34% |
| | <i>Total Rural/non-place</i> | <i>23999</i> | <i>22473</i> | <i>21985</i> | <i>488</i> | <i>97.83%</i> | <i>91.61%</i> |
| | Anne Arundel County Totals | 226515 | 191238 | 186707 | 4531 | 97.63% | 82.43% |

Table 2 – USPS County Project Match Rates, Anne Arundel County

| County/ Place | Total Records Received | Exact Match – all address components | No Match List | Invalid or Missing Secondary Addresses | Match Rate |
|------------------|---------------------------|---|---------------------|---|---------------|
| Anne Arundel | 16383 | 15617 | 311 | 455 | 95.32% |

Highlights

- The overall local match rate for the county was 97.6 percent.
- The overall USPS match rate for the county was 95.3 percent.
- All 31 places received a local match rate of 90 percent or better.
- The overall MAF Match Rate for the county was 82.4 percent.
- Over half (17 of 33) of all places had a MAF Match Rate of 90 percent or better.
- Seven places - Highland Beach, Herald Harbor CDP, Arden on the Severn CDP, Deale CDP, Riva CDP, Pasadena CDP, Mayo CDP, and Lake Shore CDP - had MAF match rates over 95 percent.

Figure 4: Census Bureau Local Match Rates

Trends and Analysis

In Anne Arundel County, there are several places, such as Annapolis, Maryland City CDP, and Parole CDP, that have a high local match rate, but a low MAF match rate. Geographic areas and blocks with lower MAF match rates are areas that have a high percentage of secondary addresses, such as apartment complexes, military and/or university housing. As the analysis below indicates, discrepancies in secondary addresses were often minor.

In some cases, the count of local address units is higher than the count of MAF units. As a result, these places have a MAF match rate that exceeds the local match rate. This only occurs in Herald Harbor CDP, Deale CDP, and Pasadena CDP. However, when analyzing the data at the block and tract level, we see this phenomenon occurring in other regions of the county.

A number of factors lowered the MAF match rates in Anne Arundel County and warrant additional research. Analysis of the address discrepancies between the local file and the MAF revealed a number of issues, including:

- Secondary/sub-address contained matching unit identifiers (e.g., 101, 102), but the local file did not include descriptors (e.g., APT, UNIT, TRLR). There are over 6,000 records in Anne Arundel County that fall into this category.
- Street name pre and post directionals (e.g., E Bancroft Ln and Bancroft Ln E, E Glenshire Ct and Glenshire Ct E, Earleigh Heights Rd W and W Earleigh Heights Rd) did not match. Over 300 address records fell into this category.

- Street name post type (e.g., Crofton Blvd and Crofton Ln, Bowen Dr and Bowen Rd) did not match.
- Inconsistent or misspelled street names (e.g., Scots Ln and Scotts Ln, Eagle Landing Way and Eagles Landing Way, Teppers Rd and Tepper Rd).
- Unrecognized or unmatched abbreviations (e.g., Cobscook Harbour and Cobscook HBR, Englishman Harbour and Englishman HBR).
- Records that contained multiple units in the local file, but Census only had one unit.

Talbot County Results and Analysis

Table 3 – Census Bureau Local and MAF Match Rates, Talbot County

| Census FIPS Code | Place Name | MAF Units | Local Address Units | Local Address Match | No Address Match | Local Match Rate | MAF Match Rate |
|------------------|------------------------------|--------------|---------------------|---------------------|------------------|------------------|----------------|
| 77912 | Tilghman Island | 645 | 783 | 636 | 147 | 81.23% | 98.60% |
| 19900 | Cordova | 246 | 266 | 241 | 25 | 90.60% | 97.97% |
| 64500 | Queen Anne | 37 | 40 | 36 | 4 | 85.71% | 97.30% |
| 78575 | Trappe | 504 | 466 | 415 | 51 | 89.06% | 82.34% |
| 24475 | Easton | 8831 | 7901 | 5526 | 2375 | 69.93% | 62.58% |
| 59450 | Oxford | 585 | 460 | 330 | 130 | 71.74% | 56.41% |
| 69825 | St. Michaels | 768 | 644 | 378 | 266 | 58.70% | 49.22% |
| | | | | | | | |
| | Total Places/CDP | 11616 | 10560 | 7562 | 2998 | 71.61% | 65.10% |
| | Total Rural/non-place | 9565 | 9955 | 8557 | 1398 | 85.96% | 89.46% |
| | Talbot County Totals | 21181 | 20515 | 16119 | 4396 | 78.57% | 76.10% |

Table 4 – USPS County Project Match Rates, Talbot County

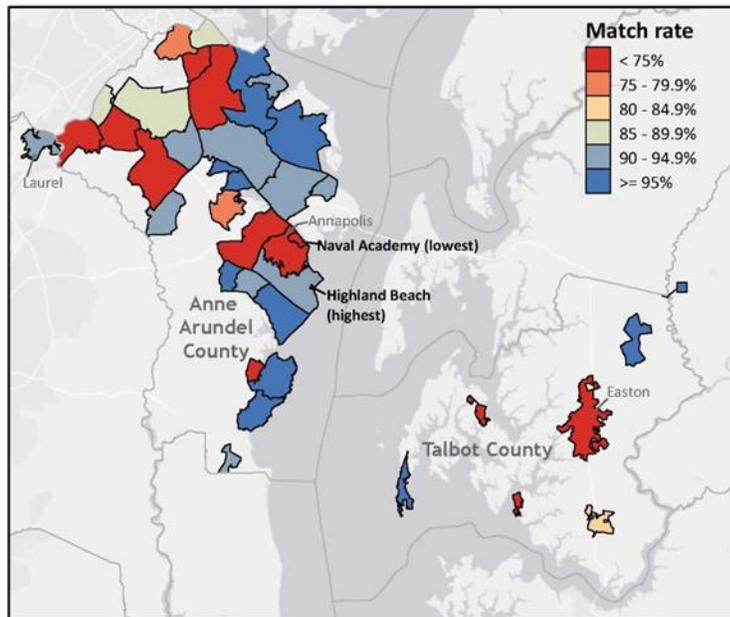
| County/Place | Total Records Received | Exact Match – all address components | No Match List | Invalid or Missing Secondary Addresses | Match Rate |
|--------------|------------------------|--------------------------------------|---------------|--|------------|
| Talbot | 23322 | 15332 | 5237 | 2690 | 65.74% |

Highlights

- The overall local match rate for the county was 78.6 percent.
- All 31 places received a local match rate of 90 percent or better.
- The overall MAF Match Rate for the county was 76.1 percent.

- Three places – Cordova, Queen Anne, and Tilghman Island had a MAF Match Rate of 97 percent or better.
- Three places – Easton, Oxford, and St. Michaels had MAF match rates below 63 percent.

Figure 5: Census Bureau MAF Match Rates



Trends and Analysis

Several places in Talbot County received a MAF match rate that was much higher than the local match rate. In the three places with the highest MAF match rate—Cordova, Queen Anne, and Tilghman Island—the number of local addresses outnumbered the count of addresses in the MAF. In these places, the local file has addresses that the Census MAF does not. This is an indication that the local file contains some non-residential addresses or that the Census MAF is lacking coverage.

The areas of the county outside of an incorporated place or CDP, presumably the more rural parts of the county, had a local match rate 86 percent and a MAF match rate of 89.5 percent. This indicates that the rural addressing in Talbot County is consistent with and has good coverage in comparison to Census Bureau data.

A number of factors lowered the overall match rate in Talbot County below 80 percent and warrant additional research. Analysis of the address discrepancies between the local file and the MAF revealed a number of issues, including:

- Street name pre and post directionals (e.g., N Locust Ln and Locust Ln N, Aurora St N and N Aurora St, W Chew Ave and Chew Ave W) did not match. Over 1,000 address records fell into this category.
- Street name post type (e.g., Audubon Ln and Audubon Dr) did not match.
- Inconsistent or misspelled street names (e.g., Bank St and Banks St).
- Secondary/sub-addresses not correctly transferred from the local file prior to processing by the Census Bureau. This artificially lowers match rates and increases no match counts.
- Talbot submitted 2,219 addresses with secondary/sub-addresses that returned as no match records. Correcting the format of these secondary/sub-addresses could potentially increase the local match rate to close to ninety percent.
- The USPS match rates are lower in part due to file structure and formatting. The USPS removed the secondary addressing from the local file prior to processing. The secondary addressing was contained in a separate field and was not compatible with the USPS processing software.

Laurel City Results

Table 5 – Census Bureau Local and MAF Match Rates, City of Laurel

| <i>Census FIPS Code</i> | <i>Place Name</i> | <i>MAF Units</i> | <i>Local Address Units</i> | <i>Local Address Match</i> | <i>No Address Match</i> | <i>Local Match Rate</i> | <i>MAF Match Rate</i> |
|-------------------------|-------------------|------------------|----------------------------|----------------------------|-------------------------|-------------------------|-----------------------|
| 45900 | Laurel | 12395 | 8565 | 7797 | 768 | 91.03% | 62.90% |

Table 6 – USPS County Project Match Rates, City of Laurel and surrounding area

| <i>County/Place</i> | <i>Total Records Received</i> | <i>Exact Match – all address components</i> | <i>No Match List</i> | <i>Invalid or Missing Secondary Addresses</i> | <i>Match Rate</i> |
|---------------------|-------------------------------|---|----------------------|---|-------------------|
| Laurel | 15683 | 9457 | 336 | 5889 | 60.30% |

Highlights

- The overall local match rate for the City of Laurel was over 91 percent.
- The lower overall MAF match rate for the City of Laurel of 62.9 percent was due to discrepancies in secondary/sub-addresses.
- The overall USPS match rate for the Laurel area was 60.3 percent. The USPS match included addresses outside of the City of Laurel boundaries.

Trends and Analysis

The high local match rate indicates that the partner supplied file closely matches the MAF in terms of primary address formatting. The low MAF match rate is due mostly to issues with secondary/sub-addresses. The USPS match rate is lower in part because it includes some addresses outside of the Laurel incorporated place boundary, whereas the Census Bureau match rates only included the addresses within the Laurel incorporated place boundary. The USPS match rate may also be low due to same issues that the Census Bureau encountered with secondary/sub-addresses (see below).

A number of factors lowered the MAF match rate in Laurel and warrant additional research. Analysis of the address discrepancies between the local file and the MAF revealed a number of issues, including:

- Secondary/sub-address contained matching unit identifiers (e.g., 301, 301B), but mismatched descriptors (e.g., local file identified address as a UNIT and Census identified the address as APT or STE). This occurred frequently. Over 3,000 records fell into this category. These addresses received an equivocated match, but did not count towards the match rate. If these records had counted as a match, the local match rate would increase to almost 93% and the MAF match rate would increase to approximately 88%.
- Primary address matched, but secondary/sub-address identifier did not match. Over 100 records fell into this category. These addresses did not count towards the overall match rate.
- Street name was missing pre- or post-directional (e.g., Washington Blvd S).

7. Conclusions

The pilot was successful in examining ways to improve address data sharing and build on existing partnerships with local and state government and the USPS. Data sharing between the Census Bureau, USPS, and local and state governments adds value to all stakeholders. However, in order to build on existing partnerships, the Census Bureau must continue to look for ways to assist local and state governments in their addressing efforts and provide incentives for local and state participation in address data sharing. The Census Bureau should also look for ways to support the relationship between USPS and local and state governments. These relationships are critical for successful data sharing and maintaining a comprehensive national address database. The following recommendations build on the findings of the address pilot and from feedback provided by our external partners.

Recommendation 1: The Census Bureau should make block and tract level MAF counts and match rates available to local governments.

Rationale: Block level data would be valuable for partners to investigate where local address records differ from MAF address records and make appropriate corrections. If block level data is not possible, the Census Bureau should release MAF counts and match rates at the tract level. This would provide incentives for local and state governments to participate in data sharing with the Census Bureau.

Recommendation 2: The Census Bureau should assist the USPS and local government in evaluating and filtering address files.

Rationale: Any assistance that the Census Bureau can provide the USPS or local governments in evaluating or editing address files will directly benefit the Census Bureau. The Census Bureau may be able to assist the USPS in identifying invalid addresses and non-structural addresses, such as cell phone towers and fire hydrants. The evaluation could take place outside of or prior to entering the MAF. Enhancements to how the USPS evaluates and processes local address files through the County Project will improve the quality of the DSF, which in turn will improve the quality of the MAF. Local governments should also be encouraged to remove any invalid and non-structural addresses prior to submitting files to the USPS County Project or Census Bureau.

Recommendation 3: The Census Bureau should run the matching and geocoding software with state level data.

Rationale: The state of Maryland is interested in receiving tract-level match counts using the *MdProperty View* data. This would allow the state to evaluate if and where holes exist relative to the MAF. In addition, the state would like see how their data compares to county level data.

Running the state level file for Anne Arundel and Talbot would highlight any differences between their data.

Recommendation 4: Match rates and tract level counts should inform targeted address canvassing.

Rationale: Tract level analysis of match rates between local files, USPS and the MAF can support the case for targeted address canvassing. For this to work, the local government must be convinced that the local address file and the MAF meet certain quality standards. Conversely, in order to convince local governments that its data are of high quality, the Census Bureau should release tract level address counts and match rates.

Recommendation 5: The Census Bureau, USPS and local government should work together to standardize the format of secondary/sub-addresses.

Rationale: Secondary/sub-addresses are the primary reason for lower match rates. Formatting issues with secondary addresses also impact match rates. For example, in Talbot County, the USPS removed secondary addresses from the file before processing because they were formatted in a separate field. This is an area again where the Census Bureau could assist the USPS in preprocessing data.

Recommendation 6: The Census Bureau should continue to investigate methods to evaluate and fix common equivocated matches.

Rationale: In many instances, match rates were lower because of minor differences between the data. Discrepancies in secondary/sub-address descriptors were responsible for the majority of equivocated matches. In Laurel, for example, the local file used ‘Unit’ to describe many secondary/sub-addresses, while the MAF, and presumably the USPS, used ‘Suite’.

Recommendation 7: The Census Bureau and USPS should rerun periodically run matching software in order to track the progress and value of address updates.

Rationale: The local governments are cleaning up their data using feedback from USPS and the Census Bureau. When complete, the plan is for USPS to rerun the local data through the County Project. The Census Bureau should do the same in order to measure the level of value added by the feedback.

A goal of the pilot was to evaluate whether the feedback provided by the Census Bureau and USPS is useful to local and state governments. The Census Bureau should continue to pursue partnership activities that provide and evaluate feedback, similar to this pilot project, as a means to further improve and enhance data sharing.

8. Acknowledgements

Mentors

Deirdre Bishop (GEO, Chief, Office of the Geographic Operations Advisor)

Paul Riley (GEO, Office of the Geographic Operations Advisor)

Team Members

Andrew Stanislaw (GEO, GAB)

Michael Clements (GEO, GAB)

Chad Toves (GEO, APMB)

Michael Horgan (Boston RO/RCC)

Charles Pruitt (External Partner, USPS)

Barney Krucoff (External Partner, MD Geographic Information Officer)

External Data Partners

James Cannistra, Maryland Department of Planning

Mark Cohoon, Talbot County Public Works

Caroline Gaulke and Matt Webb, Anne Arundel County Office of Information Technology

Theresa Martin, Department of Information Technology, City of Laurel

Attachment A: Data Dictionary

| Term | Definition |
|-----------------------------|---|
| Census FIPS | The five-digit Federal Information Processing Series (FIPS) place code is assigned based on alphabetical sequence within a state. FIPS Codes are unique within state. |
| Census Place Name | The name of the incorporated place or Census Designated Place (CDP). CDPs are the statistical counterparts of incorporated places, delineated to provide data for settled concentrations of population that are identifiable by name, but are not legally incorporated under the laws of the state in which they are located. |
| MAF Units | The count of address units stored in the Census Bureau Master Address File. The MAF unit count has been filtered to include only valid residential addresses, as determined by Census. |
| Census Local Address Units | The count of complete residential address units supplied by the local partner that were determined to be an exact match or no match to the MAF. Equivocal matches not included. |
| Census Local Address Match | The count of local partner supplied addresses that match a MAF address unit. |
| Census No Address Match | The count of local partner supplied addresses that do not match a MAF address unit. |
| Census Local Match Rate | The percentage of partner supplied address records that matched the MAF, divided by the total count of partner supplied address records. |
| MAF Match Rate | The percentage of partner supplied address records that matched the MAF, divided by the total count of address records in the MAF. |
| USPS Total Records Received | The count of local address records that the USPS processed. |

| | |
|-----------------------------|--|
| USPS Exact Match | The count of matched, standardized addresses where all address components from the partner supplied file matched exactly with the AMS database. |
| USPS No Match List | The count of records where there was no match between the partner supplied file and the AMS database, or records that contain issues such as typos or incorrect suffixes. |
| Invalid Secondary Addresses | The count of records that contain mismatched secondary addresses between the partner supplied file and the AMS (e.g., the county has units 1,2,3 and AMS has units A,B,C). |
| Missing Secondary Addresses | The count of records that require the use of apt., unit, condo, suite, etc., but are not provided in the partner supplied file. |
| USPS Match Rate | The count of exact matches divided by the count of total records received. |

Attachment B: Acronyms

| Acronym | Meaning |
|---------|--|
| AMS | USPS Address Management System |
| APMB | Address Program Management Branch |
| AUSB | Address Update Software Branch |
| CDS | Computerized Delivery Service |
| DSF | Delivery Sequence File |
| GEO | Geography |
| GSSI | Geographic Support System Initiative |
| LUCA | Local Update Census Address |
| MAF | Census Bureau Master Address File |
| NAPSG | National Alliance for Public Safety GIS |
| NSGIC | National States Geographic Information Council |
| USPS | United States Postal Service |
| WG | Working Group |

Attachment C: References

| Document Title, Author | Document # | Version, Date | Location |
|--|-------------------|----------------------|-----------------|
| Address data text file and location points in x,y coordinates, Anne Arundel County | N/A | July 2011 | |
| Address data text file and location points in x,y coordinates, City of Laurel | N/A | August 2012 | |
| Address data shapefiles and location points in x,y coordinates, Talbot County | N/A | March 2012 | |

Attachment D: File Requirements

The Census Bureau and the USPS provided the following requirements to partners prior to requesting address data:

File Requirements – USPS (from County Project SOP)

Local, county or state government would provide a file to the USPS in the following format or a format agreed upon by the USPS. The USPS will work with local governments that require or request to submit address data in a different format.

Flat non-indexed text file (fixed-width) with records separated by carriage return/line feed or CVS comma delimited.

The file name must:

1. Begin with the three letters assigned by NCSC following initial registration for the service.
2. Be followed by four numbers, reflecting the date the file was submitted, (mmdd), and end with the file extension.txt.

For example, if your three letters NCSC assigned characters are YYY and the 4 numbers representing the submit date are 0925, if it were September 25th, then YYY0925.TXT would be a valid file name.

Files must be compressed with the PKZip compression utility and must include the .zip file extension. The first seven characters of the zipped file name must be the same as the first seven characters of the text file name (e.g., if your text file name is YYY0925.TXT, the ZIP file name must be YYY0925.ZIP).

Input File Layout

Use the following file layout to prepare your file.

| Field Name | Start Position | End Position | Length | Required Fields | Field Type |
|--------------|----------------|--------------|--------|-----------------|------------|
| NAME | 1 | 42 | 42 | File Owner | A/N |
| COMPANY NAME | 43 | 108 | 66 | Business Name | A/N |
| ADDRESS LINE | 109 | 174 | 66 | Required | A/N |
| CITY | 175 | 202 | 28 | Required | A/N |
| STATE | 203 | 204 | 2 | Required | A/N |

| | | | | | |
|---|-----|-----|----|------------------------|----------------|
| ZIP5 | 205 | 209 | 5 | Required | NUMERIC |
| ZIP4 | 210 | 213 | 4 | Optional | NUMERIC |
| CONG CODE | 214 | 217 | 4 | AMS Congressional Code | Example: TN02 |
| COUNTY CODE | 218 | 222 | 5 | AMS County Code | Example: TN157 |
| FILLER | 223 | 241 | 19 | Spaces | A/N |
| KEY* | 242 | 291 | 50 | Required | A/N |
| CR/LF | 292 | 293 | 2 | Required | |
| TOTAL RECORD LENGTH 293 BYTES | | | | | |
| <i>All optional fields should be left blank if no data exists.</i> | | | | | |
| * A customer supplied identification code or record ID. (Example: PUBLICJQD5) | | | | | |

File Requirements – Census Bureau

The Census Bureau will accept the format of the USPS County Project requirements outlined above. This will reduce the burden on local government and work towards standardization of addresses. However, the Census Bureau will accept additional address fields when available. Below are the required and optional preferred address fields.

| | |
|----------------------------------|----------|
| House Number (required) | Required |
| Primary Street Name (required) | Required |
| Alternate Street Name(s)* | Optional |
| City (required) | Required |
| State (required) | Required |
| ZIP5(required) | Required |
| ZIP4 (optional) | Optional |
| Block* (Census Tabulation Block) | Optional |

| | |
|---|--------------------------------------|
| Tract* (Census Tabulation Tract) | Optional |
| Spatial Coordinates* (Longitude/Latitude, decimal degrees to precision of six decimal digits) | Optional |
| Building Name, ID* (Identifier assigned to a structure, i.e., Washington Towers, Building A) | When Available or Appropriate |
| Unit* | When Appropriate |
| Address Type* (e.g., group quarter, business, public building, fire hydrant...etc.) | When Available or Appropriate |
| Mile Marker* | Optional |
| Parcel ID* | Optional |
| <i>Rural Route Descriptor</i> | <i>Non-city Style Addresses Only</i> |
| <i>Rural Route ID</i> | <i>Non-city Style Addresses Only</i> |
| <i>Box Number</i> | <i>Non-city Style Addresses Only</i> |