2015 AHS Redesign Paper 4

Constructing Unit-Level Addresses in the Public and Indian Housing Information Center

Shawn Bucholtz
U.S. Department of Housing and Urban Development

Pramod Pamnani
Cloud Nine Technologies

March 2016

U.S. Department of Housing and Urban Development | Office of Policy Development and Research

Constructing Unit-Level Addresses in the Public and Indian Housing Information Center

Purpose

The purpose of this paper is to summarize the process of constructing U.S. Department of Housing and Urban Development (HUD) public housing unit-level addresses. Unit-level addresses were constructed from building entrance addresses and door numbers contained within the Inventory Management System (IMS)/Public and Indian Housing Information Center (PIC) system and from information contained within HUD Form 50058. The project to construct unit-level addresses was undertaken because unit-level addresses are necessary for linking public housing administrative records to the American Housing Survey (AHS), the American Community Survey, and other data sources.

1. Address Collection Background

This section provides background information on HUD's tracking and classification of public housing units and on the address collection procedures. The purpose of this section is to show the reader that, although HUD does not require the collection of unit-level addresses in IMS/PIC, the IMS/PIC requirement to capture building entrance addresses, and public housing authorities' (PHAs') optionally choosing to enter extra information in the form of apartment numbers, results in a de facto collection of unit-level addresses for as many as 98 percent of public housing units.

1.1 Public Housing Unit Tracking and Structural Classification

PHAs administer the public housing program. One system used to administer the program is called IMS/PIC. Within IMS/PIC, PHAs assign each unique public housing unit a key composed of four fields.

1. *Development code:* A unique code for the housing development.
2. *Building number:* A unique number assigned to each physical structure within the housing development.
3. *Building entrance number:* A unique number assigned to each entrance of a unique physical structure (building).
4. *Unit number:* A unique number assigned to each unit within a unique building entrance.

Development code is generally synonymous with an apartment complex or a housing development. *Building number* is generally used to number each distinct physical structure (building) within an apartment complex or housing development. *Building entrance number* is used when a building has more than one entrance, such as a "north" and "south" entrance. Finally, *unit number* is assigned for each unit associated with a building entrance number.

This classification system was likely designed to facilitate the tracking of units within multifamily apartment complexes. Only about one-half of all public housing units, however, are located within what most people would consider to be multifamily apartment buildings or complexes. The other one-half are located within rowhouses or townhouses, duplexes, and single-family detached homes. This distinction is important for reasons that will become clear in the next subsections. Table 1 shows the count of public housing units in each type of housing structure.

Table 1. Count of Public Housing Units by Structure Type, 2015

Structure Type	Count of Units	Percent of Total Units
Multifamily apartment buildings[a]	582,539	52
Rowhouses/townhouses	319,567	29
Duplexes	165,652	15
Single-family detached homes	46,446	4
Total	1,114,204	100

[a] The category multifamily apartment buildings includes two structure-type categories in the Inventory Management System/Public and Indian Housing Information Center, or IMS/PIC. The first is *ES*, which stands for structure with elevators. The second is *WU*, which stands for structures without elevators.

1.2 The Relationship Between Building Entrance Numbers and Unit Numbers

The four-field system for tracking public housing units ensures that each unit has a unique key that distinguishes it from all other units. For example, consider a public housing development (FL009) that structurally consists of one multifamily

apartment building. The building has two entrances, and each entrance has three units, for a total of six units within the building. The current coding/numbering scheme for the units would resemble the one described in table 2. As evidenced by table 2, all four fields are required to create the unique set of values necessary for tracking each unit in this multifamily structure.

Table 2. Example of Current Coding/Numbering Scheme for Units in Multifamily Structures

Development Code	Building Number	Building Entrance Number	Unit Number
FL009	1	1	1
FL009	1	1	2
FL009	1	1	3
FL009	1	2	1
FL009	1	2	2
FL009	1	2	3

Public housing units in rowhouses or townhouses follow the same coding/numbering scheme but with one important difference: each building entrance number has only one unit. In relational database parlance, this situation is referred to as a one-to-one relationship between the building entrance number and the unit number. To understand why, consider a public housing development (GA008) consisting of one building composed of six rowhouses. The current coding/numbering scheme for the units would resemble the one described in table 3. As evidenced by table 3, each of the first three fields (*development code*, *building number*, and *building entrance number*) are required to create the unique set of values necessary for tracking each of the six units, but the fourth field (unit number) is not strictly necessary because each building entrance is the entrance to an individual unit.

Table 3. Example of Current Coding/Numbering Scheme for Units in Rowhouses and Townhouses

Development Code	Building Number	Building Entrance Number	Unit Number
GA008	1	1	1
GA008	1	2	1
GA008	1	3	1
GA008	1	4	1
GA008	1	5	1
GA008	1	6	1

Rowhouses and townhouses comprise 29 percent of all public housing units. About 25 percent out of that 29 percent exhibit the one-to-one relationship between building entrance number and unit number illustrated in table 3. The other 4 percent maintain the one-to-many relationship between building entrance number and unit number that is found with units in multifamily structures (illustrated in table 2). No single explanation is likely as to why the one-to-many relationship exists for those 4 percent. One possible explanation is that the structure type truly is a multifamily apartment building but was incorrectly coded as a rowhouse/townhouse. Another explanation is that the structure type does not fit neatly into the structure type categories available in IMS/PIC, and the PHA simply picked the structure type category they believed most consistent with the structural layout.

Public housing units in duplexes generally have the same one-to-one relationship between the building entrance number and the unit number as is exhibited by units in rowhouses and townhouses. Table 4 shows an example of the coding/numbering scheme for public housing units in a development (IA004) consisting of three duplex structures, for a total of six units. Each duplex structure (building) has two entrances—one for each unit. Duplexes comprise 15 percent of all public housing units. As was the case with rowhouses and townhouses, each of the first three fields (*development code*, *building number*, and *building entrance number*) are required to create the unique set of values necessary for tracking each of the six units, but the fourth field (*unit number*) is not strictly necessary.

About 13 percent out of the 15 percent exhibit the one-to-one relationship between building entrance number and unit number. The other 2 percent maintain the one-to-many relationship between building entrance number and unit number that is found with units in multifamily structures.

Table 4. Example of Current Coding/Numbering Scheme for Units in Duplexes

Development Code	Building Number	Building Entrance Number	Unit Number
IA004	1	1	1
IA004	1	2	1
IA004	2	1	1
IA004	2	2	1
IA004	3	1	1
IA004	3	2	1

As would logically be expected, public housing units that are single-family detached homes are coded/numbered as separate buildings. Consider the example in table 5 of a public housing development (MI002) consisting of six single-family homes. In this example, each physical structure (the single-family home) has a unique building number. As the example shows, one-to-one relationships exist between building number and building entrance number and between building entrance number and unit number. As such, each public housing unit that is a single-family detached home is unique based on only development code and building number; building entrance number and unit number do not provide any additional information that contributes to the uniqueness of the unit.

Single-family homes comprise about 4 percent of all public housing units. Virtually all of them exhibit the previously described one-to-one relationship between building number and building entrance number.

Table 6 presents the total percentage of units by structure type and, within each structure type, the percentage of units that have a one-to-one relationship between building entrance number and unit number and the percentage that have a one-to-many relationship. In total, 44 percent of all public housing units have a one-to-one relationship between building entrance number and unit number. The importance of this feature of the public housing tracking scheme will become apparent in subsection 1.5.

Table 5. Example of Current Coding/Numbering Scheme for Units in Single-Family Detached Homes

Development Code	Building Number	Building Entrance Number	Unit Number
MI002	1	1	1
MI002	2	1	1
MI002	3	1	1
MI002	4	1	1
MI002	5	1	1
MI002	6	1	1

1.3 An Important Note About the Four Tracking Fields

As mentioned in subsection 1.1, PHAs track unique public housing units using a key composed of four fields with values assigned by the PHA: *development code*, *building number*, *building entrance number*, and *unit number*. Two important items are worth noting about these four fields as they relate to the address collection process discussed in subsection 1.5.

First, none of the four fields is designed to mimic how the U.S. Postal Service (USPS) assigns addresses to buildings within a multifamily property. It could be the case that HUD records a single building with four entrances, but USPS assigns all four entrances the same address (for instance, 123 Main Street). It could also be the case that HUD records a single building with four entrances, and USPS gives each entrance its own address (for instance, 123, 125, 127, and 129 Main Street).

Second, none of the four fields is required to contain numbers that correspond to actual street numbers or apartment numbers. In some cases, the unit number is an actual apartment number, but that is often the exception, not the rule. In short, the four fields are not required to contain, and often do not contain, actual address information.

1.4 U.S. Postal Service Address Definitions

For the subsequent parts of this paper, especially subsection 1.5, to be understood, some important terms must be defined.

- *USPS Deliverable Address:* An address recognized by USPS as a valid final delivery point for mail. In other words, an individual mailbox or a post office (PO) box.
- *Basic Street Address:* The portion of an address that includes the street number and street name. For single-family homes and townhouses, the basic street address is often the USPS deliverable address or the PO box.

Table 6. Percentage of Public Housing Units With One-to-One Relationship Between Building Entrance Number and Unit Number

Structure Type	Percent of Total Units	Percent of Total Units With One-to-One Relationship Between Building Entrance Number and Unit Number	Percent of Total Units With One-to-Many Relationship Between Building Entrance Number and Unit Number
Multifamily apartment buildings	52	2	50
Rowhouses/townhouses	29	25	4
Duplexes	15	13	2
Single-family detached homes	4	4	0
Total	100	44	56

- *Full Address*: The combination of basic street address plus any apartment, unit, or other identifier that is assigned by USPS. For most multifamily buildings, the full address is the USPS deliverable address.

1.5 Overview of Public Housing Address Information Collection

The previous subsections of this paper included necessary background material for understanding how PHAs record public housing addresses in IMS/PIC. Simply put, PHAs are required to collect building entrance number addresses. PHAs are not required to collect addresses at the unit-number level.

PHAs collect actual physical address information for each building entrance number in the IMS/PIC. For each unit, the PHA may enter information in the following five fields.

1. *address line 1*.
2. *address line 2*.
3. *city*.
4. *state*.
5. *ZIP Code*.

Generally speaking, IMS/PIC has no check to ensure an address is a building entrance number address, nor any post-data-entry system that enforces address quality.

It is useful to pause for a moment and consider what the address requirements and entry process means for the different types of structures discussed in subsection 1.1. Nearly all single-family detached structures, duplexes, and rowhouses and townhouses exhibit a one-to-one relationship between building number and building entrance number, which means that building entrance number address functionally should be the unit address. In other words, even though recording addresses at the unit level is not required, unit-level addresses should exist for 44 percent of public housing units.

The other 56 percent of public housing units exhibit a one-to-many relationship between building entrance number and unit number. Given the existing requirement to record only building entrance addresses, these addresses will definitely not be unit addresses. There is another data-entry field in IMS/PIC, however, called *door number*. This field is optional, meaning PHAs are not required to record any information in *door number* field. That being said, it is a logical place for PHA to record an actual apartment number for units with multiple unit numbers for a building entrance number. In fact, of the 56 percent of public housing units with a one-to-many relationship between building entrance number and unit number, 54 percent have some type of value in the *door number* field and 2 percent do not have a value.

That means PHAs are expected to have recorded unit-level address for 44 percent of public housing units and maybe to have provided enough information to construct a unit-level address for an additional 54 percent of public housing units.

1.6 Potential Complications

As mentioned in subsection 1.5, PHAs potentially record unit-level addresses for as many as 98 percent of public housing units, even though they are not required to do so. The IMS/PIC system does not have any internal checks, however, to ensure the building entrance addresses are recorded in a format that is consistent with postal address standards. The same thing is true for PHAs who have chosen to populate the optional *door number* field.

In a data-entry system designed to ensure address entry consistency, PHAs would enter the basic street address in *address line 1*, the apartment number (as required when a one-to-one relationship exists between building entrance number and unit number) in *address line 2*, and the relevant city, state, and ZIP Code in their respective fields. PHAs, however, chose to enter the address information in numerous ways that are not consistent with that model. The following list includes a few examples.

- *Address line 1* includes the house number, and *address line 2* includes the street name (example: 5, Main Street).
- *Address line 1* includes the basic street address, but *address line 2* is missing an apartment number (example: 123 Main Street, null).
- *Address line 1* includes the name of the subdivision, and *address line 2* includes the full or basic street address (example: Dyson Homes, 123 Main Street).
- *Address line 1* includes the full address, and *address line 2* repeats the apartment number (Example: 123 Main Street Apt 101, Apt 101).
- *Address line 1* includes a range of full addresses (Example: 123–129 Main Street).

As a result, in some cases, the building entrance addresses are plagued with errors such that it is difficult to extract an address that is consistent with USPS standard. This situation is one in which address standardization and geocoding can help.

1.7 Geocoding Process and Output

After collecting or updating building entrance address information in the online IMS/PIC system, HUD IMS/PIC database managers send the building entrance addresses to HUD's Geocoding Service Center (GSC). Using the building entrance address information, the GSC completes the following steps.

Step 1. Fix common HUD-specific address entry mistakes using internally developed software/code.

Step 2. Conduct address standardization and a deliverability assessment using commercial software (CODE-1 Plus®).

Step 3. Estimate the physical location of standardized address using commercial software (GCP 5.2).

Many address entry errors or nonstandard inputs will cause the geocoding process to return a null result or return a result with less precision than desired. The HUD GSC has developed an inventory of common address entry errors and corresponding procedures to correct the errors. These procedures are implemented in step 1. Because HUD deals with addresses throughout the country, this process must be somewhat conservative so that fixing systematic errors in one county does not create errors in other counties (false positives). As a result, GSC is unable to correct some addresses.

GSC's commercial geocoding software's address standardization process (step 2) is designed to fix common formatting errors. For instance, the software recognizes that a "#205" input should be "Apt 205," but it cannot handle truly messy formatting errors. The output of the address standardization is an address that can be used for step 3. It is entirely possible that the address standardization fails, however, resulting in null output. It is also possible that the address standardization removes information supplied by the user to signify an apartment number. For instance, if the user supplied "123 Main St #%204," the software may return "123 Main St," discarding the apartment information because it is in a format the software is not designed to standardize.

After the address is standardized, the CODE-1 Plus® software determines if the standardized address is a valid USPS deliverable address. Similar to the output of the physical location estimation (step 3), the output includes information that can be used to assess the precision of the "deliverability assessment." For instance, the output will tell you if USPS recognizes the basic street address but does not recognize the apartment number.

Finally, in step 3, GSC attempts to estimate the physical location of the address using commercial software (GCP 5.2), a commercial base map (TomTom), and associated lookup tables. The important thing to note is that the GCP 5.2 software interpolates the spatial location based on street segment address ranges (highest precision), ZIP+4 centroid (next highest precision), or ZIP Code centroid (lowest precision). These precision levels have been termed "high," "medium," and "low," respectively.

2. Description of Process Used To Construct Public Housing Unit-Level Addresses

This section describes the process used to construct public housing unit-level addresses from the building entrance address and door number information in the existing IMS/PIC system and information extracted from HUD Form 50058 BLOB (or Binary Large Object) files, which are described in detail in subsection 2.3.

2.1 Determining What Constitutes a Usable Address

It is important to understand the relationship, and differences, between geocoding and address matching. Geocoding, as described in subsection 1.7, is the process of taking a raw address, standardizing it, conducting a deliverability assessment, and estimating where it is physically located on the landscape. Address matching is the process of taking two addresses and determining if they represent the same physical place. In an address-matching project, it is common to standardize addresses first and then implement text matching between address elements.[1]

When conducting geocoding or address matching, the user determines when an address should be considered "usable." Recall from subsection 1.7 that the CODE-1 Plus® software determines whether an address is deliverable, and the GCP 5.2 software estimates the spatial location of the address. Each of these processes comes with levels of precision. The two processes are entirely separate, relying on different data sources.

For some uses, such as mapping, it is likely sufficient for a standardized address to be present and to have a medium to

[1] An alternative matching method is to use ZIP+6, which is often a unique 11-digit ZIP Code for each housing unit.

high level of spatial precision while not meeting any tests of deliverability. By contrast, for address matching projects, it is often very important that a standardized address be present and deliverable, regardless of the spatial location precision.

An additional consideration for address matching projects is the uniqueness of the standardized address. For example, suppose a researcher is seeking to match public housing units with health survey responses using the address of the respondent. Further suppose the researcher has a health survey response from 123 Main St. Apt 1. When researchers match to the HUD public housing data, they find two public housing units sharing the same address. This finding may cause the researchers to discard the survey response because they could not determine which of the two public housing units responded to the survey. In fact, the HUD public housing data have many instances of two different units having the same address.

For the remainder of this paper, a public housing unit-level address is considered usable if it is deliverable with high precision and a unique address among all public housing units. Appendix A describes the definition of "deliverable with high precision."

2.2 Initial Assessment of Address Usability

Before describing the process for constructing unit-level public housing addresses, it is useful to establish a baseline against which improvement can be measured. When the GSC geocodes addresses, they typically are using only the information available in the five core address fields: *address line 1*, *address line 2*, *city*, *state*, and *ZIP Code*. GSC does not use any information from the optional *door number* field that exists in IMS/PIC.

As mentioned in subsection 2.1, different users have different standards for declaring an address usable, and the standard adopted for assessment will depend on a user's needs. Although the usable address standard chosen for this paper is "deliverable with high precision and unique," baseline results are presented for address geocoding precision standards (table 7).

The baseline results for geocoding precision are consistent with previous expectations. Nearly 87 percent of addresses have high geocoding precision, meaning enough information was contained within the five address fields to estimate the location of the building entrance to a rooftop. This result suggests PHAs are generally doing an adequate, although not perfect, job of entering building entrance addresses.

Table 7. Baseline Results of Geocoding Precision and Address Matching Usability Using Typical GSC Process

Address Geocoding Precision Level	Percent of PH Addresses
High: "Rooftop" precision	86.5
Medium: ZIP+4 centroid	7.6
Low: ZIP Code centroid	5.8
Total	100
Address Matching Usability Level	
Usable: Deliverable with high precision and unique	4.7
Not usable: Not deliverable with high precision and/or not unique	95.3
Total	100

GSC = Geocoding Service Center. PH = public housing.

The baseline results for address matching usability are far lower than previous expectations. Recall that 44 percent of public housing units have building entrance addresses that are actually unit-level addresses, simply because of the processes in place for recording building entrances. As such, if PHAs are recording building entrance addresses correctly, as many as 44 percent of units could be expected to have deliverable with high precision and unique unit-level addresses. The actual result was 4.7 percent, suggesting that PHAs are recording building entrance addresses that are not deliverable, per USPS standards.

2.3 Using HUD Form 50058 BLOB Files

Separately from the manual entry of building entrance addresses into the online IMS/PIC,[2] HUD collects public housing tenant information, including unit-level addresses, on form HUD Form 50058. Address information on form HUD Form 50058 is collected by PHAs within their respective record systems and electronically uploaded to HUD headquarters as raw text files, often referred to as "BLOB files." Most of the HUD Form 50058 BLOB information is parsed and stored in various tables in IMS/PIC. For public housing, however, the unit-level address information from HUD Form 50058 BLOB files is not currently parsed and stored in IMS/PIC. As such, the only building- and unit-level address information currently stored in IMS/PIC is the information entered by PHAs. It is worth noting that IMS/PIC system was not designed to store every field parsed from the HUD Form 50058 BLOB files, including the address fields. It was designed based on program requirements in place during system development.

[2] In practice, each PHA has its own software that collects address information and uploads it to IMS/PIC.

2.4 Process for Improving Unit-Level Addresses

This subsection describes the multistage process for constructing unit-level addresses. These processes were run on a September 2013 quarterly extract of IMS/PIC. It is important to note that, beginning with stage 2, the procedures in each stage are applied to addresses deemed not usable after the previous stage. For instance, stage 2 processes are applied to any addresses deemed not usable after stage 1 processes were completed.

Stage 1

Stage 1 of the process for constructing unit-level addresses consisted of two parts. Part 1 included adding door numbers to the existing information in the first address field (*address line 1*). Part 1 also included a manual review of the text information in the *door number* field for a small sample of units, followed by development and application of code to repair common issues preventing the construction of a usable unit-level address.

Part 2 of stage 1 consisted of a manual review of the text information in *address line 2* for a small sample of units, when present, followed by development and application of code to repair common issues preventing the construction of a usable unit-level address. As mentioned before, various PHAs enter addresses differently in the online IMS/PIC. As such, some information in the *address line 2* field represented apartment numbers, some information represented the full address, and some information was erroneous. Where feasible, information in the *address line 2* field was added to the *address line 1* information to form a full address.

Stage 2

Stage 2 of the process for constructing unit-level addresses consisted of extracting unit-level address information from HUD Form 50058 BLOB files. Each day, PHAs upload the HUD Form 50058 BLOB files to HUD. These daily BLOB uploads contain only records that have been updated since the previous upload (typically, the previous day). Given current public housing operating procedures, however, it is nearly a certainty that PHAs will have completed a new HUD Form 50058 sometime during the previous year for each household currently participating in public housing.

Section 5A of the HUD Form 50058 contains the unit address and apartment number. In effect, this source of address information is entirely separate from what is contained in IMS/PIC. As designed, however, IMS/PIC does not parse the HUD Form 50058 BLOB unit address information and store it in a table.

Scripts were written to extract the address fields from the HUD Form 50058 BLOBs. The process was conducted only for the units for which stage 1 improvement efforts did not yield a usable address (about 17 percent of units).

Stage 3

Stage 3 of the process for constructing unit-level addresses consisted of manually inspecting, correcting, and testing building entrance addresses in public housing developments that contained a large number of nonusable addresses. The process can generally be described as—

1. Extract the building entrance addresses for a single public housing development with many nonusable addresses and combine it with the *door number* field to form a full address.
2. Visually inspect the addresses to determine why the addresses are not usable in their current form.
3. Use the current address information, including any additional information from HUD Form 50058 BLOB files in stage 2, to develop a proposed unit-level address.
4. Test the usability of the proposed unit-level address against the USPS ZIP Code Lookup application, the GSC Geocoder, Google Maps, and general web searches.
5. When a proposed unit-level address is found to be usable, create SAS code to apply the revision technique to all the unit-level addresses in the development.

The process was followed for 88 separate public housing developments across numerous states. In some cases, the revision process to construct the unit-level address was as simple as concatenating a few fields or removing erroneous information. In other cases, the revision process was complicated. What held true through the revision process was that if one address in the public housing development required a particular type of revision, virtually all the addresses required the same revision.

Stage 4

Stage 4 of the process for constructing unit-level addresses consisted of a simple assumption and subsequent implementation of that assumption. The assumption was that a small fraction of building entrance addresses that, after three previous stages of improvement attempts, were still not deemed usable perhaps had been updated in the IMS/PIC with better addresses. The original building entrance addresses for this address construction project were derived from a September 2013 IMS/PIC extract. These original building entrance addresses then underwent the

previous three stages of development to turn them into usable unit-level addresses. It was hypothesized, however, that other building entrance address revisions (improvements) or additions of door numbers could have been entered into IMS/PIC by PHAs after September 30, 2013. If PHAs revised building entrance addresses or door numbers after September 30, 2013, these revisions should be present in a current IMS/PIC extract.

A short process was developed that included the following steps.

1. Determine which building entrance addresses were still not usable after stage 3.
2. For those building entrance addresses, extract the current IMS/PIC building entrance address information (where applicable) from a June 30 snapshot of PIC.
3. Evaluate the June 30 addresses to determine if they were usable. When usable, replace the September 30, 2013 address with the June 30, 2015 address.

Stage 5

Stage 5 of the process for constructing unit-level addresses was nearly identical to stage 3. The only major difference was that building entrance addresses (including door numbers) were evaluated on a PHA basis rather than a development basis. The move to PHA-based evaluation was undertaken because it was discovered after stage 4 that building entrance address entry issues (or door number entry issues) preventing the construction of usable unit-level addresses were often consistent at the PHA level, meaning that many or all of the public housing developments within the PHA shared the same address entry issue.

The stage 5 process can generally be described as—

1. Extract the building entrance addresses and door numbers for a PHA with many nonusable addresses.
2. Visually inspect the building entrance addresses and door numbers to determine why they are not usable in their current form.
3. Use the current building entrance addresses and door numbers, including any additional information from BLOB files in stage 2, to develop a proposed unit-level address.
4. Test the usability of the proposed unit-level address against the USPS ZIP Code Lookup application, the GSC Geocoder, Google Maps, and general web searches.
5. When a proposed unit-level address is found to be usable, create SAS code to apply the revision technique to all the building entrance addresses and door numbers in the PHA.

The process was followed for 18 separate PHAs across numerous states. In some cases, the revision process to construct the unit-level address was as simple as concatenating a few fields or removing erroneous information. In other cases, the revision process was complicated.

3. Result of Address Construction Procedures and Conclusion

The final results of the unit-level address construction procedures are presented in table 8. Stage 1 efforts dramatically improved

Table 8. Final Results of Geocoding Precision and Address Matching Usability

Address Geocoding Precision Level	Initial Percent of PH Addresses	Stage 1 (%)	Stage 2 (%)	Stage 3 (%)	Stage 4 (%)	Stage 5 (%)
High: "Rooftop" precision	86.5	91.5	92.8	92.9	93.0	92.9
Medium: ZIP+4 centroid	7.6	3.5	3.5	3.5	3.5	3.7
Low: ZIP Code centroid[a]	5.8	5.0	3.7	3.6	3.5	3.4
Total	100	100	100	100	100	100
Address Matching Usability Level						
Usable: Deliverable with high precision and unique	4.7	83.1	89.6	91.9	92.1	92.8
Not usable: Not deliverable with high precision and/or not unique	95.3	16.9	10.4	8.1	7.9	7.2
Total	100	100	100	100	100	100

PH = public housing.

[a] In stage 5, the percentage of unit-level addresses with high address geocoding precision actually went down by 0.1 percent. Although a curious result, it is easily explained. In stage 5, about 6,000 building entrance addresses were fixed. About 1,000 of these building entrance addresses had high address geocoding precision but actually were not a reflection of where the unit was located. It is possible the original address was an address of a main building. Regardless of the reason, after the building entrance address was fixed and verified as deliverable with high precision and unique, the resulting unit-level addresses within the building entrance had less address geocoding precision than the original address. In other words, the geocoding software could physically locate the original building entrance address, but the address was not where the unit was physically located.

the number of usable unit-level addresses, and moderate gains were made in each subsequent stage. In addition to the large improvements in the number of usable unit-level addresses for address matching, small gains were made the in the geocoding precision level.

It is worth putting the results in the context of the effort necessary to achieve them. In subsection 1.5, it was hypothesized that unit-level addresses could be constructed from existing IMS/PIC information for as many as 98 percent of public housing units. Of course, this conclusion assumes that the building entrance address information currently in IMS/PIC and the supplemental apartment number information (door number) optionally entered into IMS/PIC by PHAs were correct.

The baseline assessment of address usability concluded that only 4.7 percent of addresses were deliverable with high precision and unique. This result, however, was based on the current GSC processes, which made no use of the supplemental apartment number information contained in IMS/PIC and which attempted to correct only a handful of systematic errors in address entry.

In stage 1, a major improvement occurred in the usability of unit-level address (from 4.7 to 83.1 percent). This dramatic result essentially was achieved by two steps. The first step was to use the supplemental apartment number information by adding it to the existing building-level address. The second step was to fix many systematic errors in the way building entrance addresses were entered into IMS/PIC. These two steps did not require a significant amount of manual inspection.

Stage 2 improved usability from 83.1 to 89.6 percent. This improvement was achieved by making use of the HUD Form 50058 BLOB files, which are in effect a secondary source of unit-level address information. Much like stage 1, stage 2 did not require a significant amount of manual inspection.

Stages 3 and 5 consisted of a significant amount of manual inspection to repair development- and PHA-specific address entry errors to construct usable unit-level addresses. The net return on these two stages was 3.2 percent. The return on Stage 4 was 0.2 percent.

Given the experiences with stages 3 and 5, it can safely be concluded that a large amount of manual effort would be required to construct additional unit-level addresses and move the needle from 92.8 percent to the hypothesized 98 percent.

Appendix A. Defining What Is "Deliverable With High Precision"

Unit-level address matching between HUD public housing units and the American Housing Survey (AHS) or the American Community Survey is conducted using the U.S. Census Bureau's Master Address File (MAF). The MAF is a master list of all residential addresses in the United States used by the Census Bureau for conducting its censuses and surveys. It is primarily based on data supplied by the U.S. Postal Service (USPS) but is augmented with canvassing operations conducted by the Census Bureau and feedback from local jurisdictions. In the MAF, each unit-level address is assigned a unique number called the *MAFID*. The Census Bureau takes the standardized HUD public housing unit addresses and matches them to the MAF such that each public housing unit receives a unique number. The process is repeated for the AHS. When the process is complete, the match between HUD public housing units and the AHS is a simple process of matching on MAFID.

The Geocoding Service Center's commercial geocoding software includes a Delivery Point Validation (DPV) determination. Generally speaking, this determination reflects if USPS delivers mail to the address. To be more specific, the determining takes the values described in table A-1.

On first glance, it is tempting to conclude that *Y* is the only DPV code that truly means deliverable. HUD, however, conducted an analysis of how often HUD public housing units uniquely matched with the MAF and how the match rates varied by DPV code value. The results showed that public housing units with DPV codes *Y*, *N*, and *D* nearly always produced unique matches with the MAF, but DPV code *S* did not. Although it is somewhat difficult to explain this result for DPV codes *N* and *D*, it is nonetheless an important result.

Using the results of this analysis, HUD made the determination that public housing unit-level addresses with DPV codes *Y*, *N*, and *D* would be classified as "deliverable with high precision."

Table A-1. DPV Codes

DPV Code	Meaning
Y	Confirmed; entire address was DPV confirmed as deliverable.
N	Not confirmed; address could not be DPV confirmed as deliverable.
S	Confirmed by dropping secondary information (apartment, suite, etc.).
D	Confirmed, but missing secondary information (apartment, suite, etc.).

DPV = Delivery Point Validation.