

## Data Notes

The National Housing Preservation Database (NHPD) provides users with the best assumption of subsidy status based on nationally available data sources. The accuracy, format, completion, and frequency of update for each data source vary by program and have changed overtime. Data for Section 8 Project Based Rental Assistance (PBRA) and US Department of Housing and Urban Development (HUD) Insured subsidies update monthly, while all other subsidy programs update once a year or less. Programs primarily tracked at the local level such as Low Income Housing Tax Credit (LIHTC) subsidies often have insufficient information to determine subsidy status at the national level. For instance, LIHTC subsidies can be kept affordable by state Housing Finance Agencies (HFAs) despite passing their extended use period or being listed as non-programmatic in HUD's LIHTC Database. Until September 2017, the true subsidy end date for Section 515 Direct Loans was not available and the Restriction Clause Expiration Date was used instead. Therefore, it is possible that many Section 515 Direct Loans were incorrectly classified as inactive in versions of the database before 2017.

Each record in the database describes a property. While each property can consist of multiple buildings and addresses, only the first address listed is displayed in the database. This may hinder searches by property address. Searching for individual property addresses can be particularly challenging for public housing properties. Public housing properties are reported in Asset Management Projects (AMPs), which tend to be larger groups of buildings that can be scattered among a larger area than other subsidy programs.

The number of active subsidies reported in the NHPD can change over time as the format of data files change, subsidy programs expand or are phased out, and as the logic applied to determine subsidy status is updated. The procedure for updating new data and changes in the number of active subsidies and factors that affect the comparability of data over time are noted below for each data refresh.

## Data Refresh Updates

### October 2020

The October 2020 data refresh incorporated one file update with no documentation changes.

#### *Number of Active and Inconclusive Subsidies Before and After October 2020 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	23,700	23,809	0%
<b>Section 202 Direct Loans</b>	1,218	1,209	-1%
<b>State HFA 236</b>	182	181	-1%
<b>HUD Insured</b>	9,632	9,617	0%
<b>LIHTC</b>	38,271	38,146	0%
<b>HOME</b>	20,381	20,038	-2%
<b>Section 515</b>	13,236	13,220	0%
<b>Section 538</b>	1,133	1,129	0%
<b>Public Housing</b>	6,352	6,335	0%
<b>State Subsidy</b>	2,539	2,536	0%
<b>Project Based Vouchers</b>	3,154	3,150	0%

<b>Mod Rehab</b>	515	514	0%
<b>Total</b>	120,313	119,884	0%

\*Data updated (previous data outdated by less than one year)

### August 2020

The August 2020 data refresh incorporated seven file updates, two new data files, and one documentation change. Buildings with 11 or more units assisted by Project Based Vouchers and Mod Rehab were newly added to the NHPD during the August 2020 data refresh. Additionally, HUD insured mortgages listed in HUD's terminated mortgage database with a maturity date in the future and a termination type code not equal to 20, 14, 17, 18, 60, 13, 26, 11, or 21 are now have a subsidy status equal to inconclusive rather than active.

### Number of Active and Inconclusive Subsidies Before and After August 2020 Refresh

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	23,700	23,700	0%
<b>Section 202 Direct Loans*</b>	1,277	1,218	-5%
<b>State HFA 236</b>	183	182	-1%
<b>HUD Insured*</b>	9,651	9,632	0%
<b>LIHTC*</b>	36,732	38,271	4%
<b>HOME</b>	20,956	20,381	-3%
<b>Section 515*</b>	13,301	13,236	0%
<b>Section 538*</b>	1,078	1,133	5%
<b>Public Housing*</b>	6,466	6,352	-2%
<b>State Subsidy</b>	2,543	2,539	0%
<b>Project Based Vouchers**</b>	0	3,154	N/A
<b>Mod Rehab**</b>	0	515	N/A
<b>Total</b>	116,006	120,313	4%

\*Data updated (previous data outdated by less than one year)

\*\*New data

The following updates were made to LIHTC end dates to correct for statewide affordability restriction requirements in Qualified Action Plans:

<b>State</b>	<b>Conditions</b>	<b>End Date</b>
<b>Alabama</b>	[YR_ALLOC]>2019	[YR_PIS]+35 or [YR_ALLOC]+37 if [YR_PIS] is blank.
<b>California</b>	[YR_ALLOC]>2000* Tax credits allocated earlier may be subject to extended affordability restrictions, but earlier QAPs are not available.	[YR_PIS]+55 or [YR_ALLOC]+57 if [YR_PIS] is blank.
<b>Connecticut</b>	[YR_ALLOC]>2010 and [CREDIT]=2 or 3	[YR_PIS]+40 or [YR_ALLOC]+42 if [YR_PIS] is blank.
<b>Hawaii</b>	[YR_ALLOC]>2015 and [CREDIT]=1 or 3	[YR_PIS]+45 or [YR_ALLOC]+47 if [YR_PIS] is blank.
<b>Kentucky</b>	[YR_ALLOC]>2004 and [YR_ALLOC]<2019	[YR_PIS]+33 or [YR_ALLOC]+35 if [YR_PIS] is blank.

<b>Maine</b>	[YR_ALLOC]>2003 and [YR_ALLOC]<2013	[YR_PIS]+90 or [YR_ALLOC]+92 if [YR_PIS] is blank.
	[YR_ALLOC]>2012	[YR_PIS]+45 or [YR_ALLOC]+47 if [YR_PIS] is blank.
<b>New Hampshire</b>	[YR_ALLOC]>2003 and [YR_ALLOC]<2019	[YR_PIS]+99 or [YR_ALLOC]+102 if [YR_PIS] is blank.
	[YR_ALLOC]>2018 & [YR_ALLOC]<2020 and [CREDIT]=2 or 3	[YR_PIS]+99 or [YR_ALLOC]+102 if [YR_PIS] is blank.
	[YR_ALLOC]>2020 and [CREDIT]=2 or 3	[YR_PIS]+60 or [YR_ALLOC]+62 if [YR_PIS] is blank.
<b>Oregon</b>	[YR_ALLOC]>2010 and [CREDIT]=2 or 3	[YR_PIS]+60 or [YR_ALLOC]+62 if [YR_PIS] is blank.
<b>Pennsylvania</b>	[YR_ALLOC]>2016	[YR_PIS]+35 or [YR_ALLOC]+37 if [YR_PIS] is blank.
<b>Utah</b>	[YR_ALLOC]>1999 and [YR_ALLOC]<2008	[YR_PIS]+50 or [YR_ALLOC]+52 if [YR_PIS] is blank.
	[YR_ALLOC]>2007 and [YR_ALLOC]<2013	[YR_PIS]+99 or [YR_ALLOC]+101 if [YR_PIS] is blank.
	[YR_ALLOC]>2012	[YR_PIS]+50 or [YR_ALLOC]+52 if [YR_PIS] is blank.
<b>Vermont</b>	[YR_ALLOC]>2000 and [CREDIT]=2 or 3	[YR_PIS]+99 or [YR_ALLOC]+101 if [YR_PIS] is blank.

March 2020

The March 2020 data refresh incorporated one file update with no documentation changes.

*Number of Active and Inconclusive Subsidies Before and After March 2020 Refresh*

Subsidy Type	Before Refresh	After Refresh	Change
<b>Section 8 PBRA*</b>	23,610	23,700	0%
<b>Section 202 Direct Loans</b>	1,286	1,277	-1%
<b>State HFA 236</b>	183	183	0%
<b>HUD Insured</b>	9,694	9,651	0%
<b>LIHTC</b>	36,732	36,732	0%
<b>HOME</b>	21,113	20,956	-1%
<b>Section 515</b>	13,301	13,301	0%
<b>Section 538</b>	1,078	1,078	0%
<b>Public Housing</b>	6,466	6,466	0%
<b>State Subsidy</b>	2,543	2,543	0%
<b>Total</b>	116,006	115,887	0%

\*Data updated (previous data outdated by less than one year)

December 2019

The December 2019 data refresh incorporated six file updates with no documentation changes.

*Number of Active and Inconclusive Subsidies Before and After December 2019 Refresh*

Subsidy Type	Before Refresh	After Refresh	Change
<b>Section 8 PBRA*</b>	23,681	23,610	0%

<b>Section 202 Direct Loans*</b>	1,329	1,286	-3%
<b>State HFA 236</b>	184	183	-1%
<b>HUD Insured*</b>	9,728	9,694	0%
<b>LIHTC</b>	36,957	36,732	-1%
<b>HOME*</b>	19,046	21,113	11%
<b>Section 515*</b>	13,423	13,301	-1%
<b>Section 538*</b>	995	1078	8%
<b>Public Housing</b>	6,471	6,466	0%
<b>State Subsidy</b>	2,556	2,543	-1%
<b>Total</b>	114,370	116,006	1%

\*Data updated (previous data outdated by less than one year)

### October 2019

The October 2019 data refresh incorporated one file update with no documentation changes.

#### *Number of Active and Inconclusive Subsidies Before and After October 2019 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	23,632	23,681	0%
<b>Section 202 Direct Loans</b>	1,333	1,329	0%
<b>State HFA 236</b>	184	184	0%
<b>HUD Insured</b>	9,769	9,728	0%
<b>LIHTC</b>	36,956	36,957	0%
<b>HOME</b>	19,252	19,046	-1%
<b>Section 515*</b>	13,423	13,423	0%
<b>Section 538</b>	995	995	0%
<b>Public Housing</b>	6,471	6,471	0%
<b>State Subsidy</b>	2,556	2,556	0%
<b>Total</b>	114,571	114,370	0%

\*Data updated (previous data outdated by less than one year)

### June 2019

The June 2019 data refresh incorporated updates from seven programs and changes to how Low Income Housing Tax Credits (LIHTC) subsidy status is classified. Tax credits that were allocated before 1990 or listed as non-programmatic are now listed as inactive, resulting in a large decrease in the number of active and inconclusive tax credits. Additionally, HUD insured mortgages listed in HUD's terminated mortgage database that do not impose affordability restrictions were added to the NHPD if the property received another subsidy. Previously these mortgages were excluded, resulting in a large increase in the number of HUD insured mortgages in the database.

#### *Number of Active and Inconclusive Subsidies Before and After June 2019 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	23,545	23,632	0%
<b>Section 202 Direct Loans*</b>	1,363	1,333	-2%
<b>State HFA 236</b>	183	184	1%
<b>HUD Insured*</b>	7,851	9,769	24%
<b>LIHTC*</b>	45,169	36,956	-18%
<b>HOME</b>	19,467	19,252	-1%
<b>Section 515*</b>	13,472	13,423	0%
<b>Section 538*</b>	911	995	9%

<b>Public Housing**</b>	6,587	6,471	-2%
<b>State Subsidy</b>	2,556	2,556	0%
<b>Total</b>	121,104	114,571	-5%

\*Data updated (previous data outdated by less than one year)

\*\*Data update (previous data outdated by more than one year)

## February 2019

The February 2019 data refresh incorporated updates from five programs and changes to how Low Income Housing Tax Credits (LIHTC) end dates are classified. Tax credits allocated before 1990 are now listed as inconclusive and have affordability periods of only fifteen years. Further, tax credit end dates were updated to reflect statewide affordability requirements beyond the federal minimum in place. The following updates were made to LIHTC end dates to correct for statewide affordability restriction requirements:

State	Conditions	End Date*
<b>Utah</b>	[YR_ALLOC]>1999 and [YR_ALLOC]<2008	[YR_PIS]+50 or [YR_ALLOC]+52 if [YR_PIS] is blank.
	[YR_ALLOC]>2007 and [YR_ALLOC]<2013	[YR_PIS]+99 or [YR_ALLOC]+101 if [YR_PIS] is blank.
	[YR_ALLOC]>2012	[YR_PIS]+50 or [YR_ALLOC]+52 if [YR_PIS] is blank.
<b>Vermont</b>	[YR_ALLOC]>2000 and [CREDIT]=2 or 3	[YR_PIS]+99 or [YR_ALLOC]+101 if [YR_PIS] is blank.
<b>California</b>	[YR_ALLOC]>2000* *Tax credits allocated earlier may be subject to extended affordability restrictions, but earlier QAPs are not available.	[YR_PIS]+55 or [YR_ALLOC]+57 if [YR_PIS] is blank.
<b>Kentucky</b>	[YR_ALLOC]>2004 and [YR_ALLOC]<2019	[YR_PIS]+33 or [YR_ALLOC]+35 if [YR_PIS] is blank.
<b>Maine</b>	[YR_ALLOC]>2003 and [YR_ALLOC]<2013	[YR_PIS]+90 or [YR_ALLOC]+92 if [YR_PIS] is blank.
	[YR_ALLOC]>2012	[YR_PIS]+45 or [YR_ALLOC]+47 if [YR_PIS] is blank.
<b>New Hampshire</b>	[YR_ALLOC]>2003 and [YR_ALLOC]<2019	[YR_PIS]+99 or [YR_ALLOC]+102 if [YR_PIS] is blank.
	[YR_ALLOC]>2018 and [CREDIT]=2 or 3	[YR_PIS]+99 or [YR_ALLOC]+102 if [YR_PIS] is blank.
<b>Oregon</b>	[YR_ALLOC]>2010 and [CREDIT]=2 or 3	[YR_PIS]+60 or [YR_ALLOC]+62 if [YR_PIS] is blank.
<b>Connecticut</b>	[YR_ALLOC]>2010 and [CREDIT]=2 or 3	[YR_PIS]+40 or [YR_ALLOC]+42 if [YR_PIS] is blank.
<b>Hawaii</b>	[YR_ALLOC]>2015 and [CREDIT]=1 or 3	[YR_PIS]+45 or [YR_ALLOC]+47 if [YR_PIS] is blank.
<b>Pennsylvania</b>	[YR_ALLOC]>2016	[YR_PIS]+35 or [YR_ALLOC]+37 if [YR_PIS] is blank.

\*An assumed end date of January 1st is used since only the year placed in service is provided in the originating dataset.

## Number of Active and Inconclusive Subsidies Before and After February 2019 Refresh

Subsidy Type	Before Refresh	After Refresh	Change
--------------	----------------	---------------	--------

<b>Section 8 PBRA*</b>	23,419	23,545	1%
<b>Section 202 Direct Loans*</b>	1,534	1,363	-11%
<b>State HFA 236</b>	183	183	0%
<b>HUD Insured*</b>	7,804	7,851	1%
<b>LIHTC</b>	45,216	45,169	0%
<b>HOME**</b>	17,286	19,467	13%
<b>Section 515*</b>	14,398	13,472	-6%
<b>Section 538</b>	907	911	0%
<b>Public Housing</b>	6,592	6,587	0%
<b>State Subsidy</b>	2,535	2,556	1%
<b>Total</b>	<b>119,874</b>	<b>121,104</b>	<b>1%</b>

\*Data updated (previous data outdated by less than one year)

\*\*Data update (previous data outdated by more than one year)

### June 2018

The June 2018 data refresh incorporated four file updates resulting in minimal changes compared to April 2018.

#### *Number of Active and Inconclusive Subsidies Before and After June 2018 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	23,393	23,419	0.11%
<b>Section 202 Direct Loans</b>	1,536	1,534	-0.13%
<b>State HFA 236</b>	183	183	0.00%
<b>HUD Insured*</b>	7,681	7,804	1.60%
<b>LIHTC*</b>	44,535	45,216	1.53%
<b>HOME</b>	17,682	17,286	-2.24%
<b>Section 515</b>	14,396	14,398	0.01%
<b>Section 538*</b>	748	907	21.26%
<b>Public Housing</b>	6,592	6,592	0.00%
<b>State Subsidy</b>	2,535	2,535	0.00%
<b>Total</b>	<b>119,281</b>	<b>119,874</b>	<b>0.50%</b>

\*Data updated (previous data outdated by less than one year)

### April 2018

The April 2018 data refresh incorporated one file update. The subsidy end dates for Section 8 contracts that received a Section 811 or 202 capital advance were updated to be 40 years from the date of initial occupancy since properties that receive Section 202 or 811 capital advances are required to continue renewing their Section 8 contracts for a period of 40 years.

#### *Number of Active and Inconclusive Subsidies Before and After April 2018 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	23,104	23,393	-1.3%
<b>Section 202 Direct Loans</b>	1,536	1,536	0.0%

<b>State HFA 236</b>	182	183	-0.5%
<b>HUD Insured</b>	7,685	7,681	0.1%
<b>LIHTC</b>	44,323	44,535	-0.5%
<b>HOME</b>	17,948	17,682	1.5%
<b>Section 515</b>	14,371	14,396	-0.2%
<b>Section 538</b>	748	748	0.0%
<b>Public Housing</b>	6,588	6,592	-0.1%
<b>State Subsidy</b>	2,535	2,535	0.0%
<b>Total</b>	119,020	119,281	-0.2%

\*Data updated (previous data outdated by less than one year)

### December 2017

The December 2017 data refresh incorporated six file updates. The subsidy end dates for Section 8 contracts that completed a full mark to market agreement were updated to be 30 years from the date the agreement was completed. No other changes were made to how property records are processed which resulted in minimal changes to the number of active and inconclusive subsidies included the database.

### *Number of Active and Inconclusive Subsidies Before and After December 2017 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	22,970	23,104	1%
<b>Section 202 Direct Loans</b>	1,538	1,536	0%
<b>State HFA 236</b>	181	182	1%
<b>HUD Insured*</b>	7,896	7,685	-3%
<b>LIHTC*</b>	41,384	44,323	7%
<b>HOME</b>	18,461	17,948	-3%
<b>Section 515*</b>	14,092	14,371	2%
<b>Section 538</b>	748	748	0%
<b>Public Housing*</b>	6,778	6,588	-3%
<b>Total</b>	116,586	119,020	2%

\*Data updated (previous data outdated by less than one year)

### September 2017

The September 2017 data refresh incorporated 12 new file updates, a new matching protocol, and changes to how subsidy status is classified for all subsidies. These changes will impact overtime comparisons of earlier versions of the NHPD. Since many subsidies are either missing information on the subsidy end date or are not continually monitored by the government agencies that provide data on these programs, 'inconclusive' was added as a third subsidy status category. This change gives researchers more control over how conservative their estimates of active and inactive affordable housing properties are. Researchers can now choose to report inconclusive properties separately or re-categorize inconclusive subsidies using the subsidy status description field, which describes why a subsidy is labeled inactive or inconclusive. The number of active LIHTC, Section 515, and Section 8 PBRA, and State HFA 236 subsidies were most affected by the subsidy status change rule. Many subsidies missing key information, such as subsidy end date, that were previously classified as active are now classified as inconclusive. Similarly, the number of HUD insured mortgages increased drastically between

2016 and 2017 due to the inclusion of a handful of HFA risk sharing mortgages programs that were not previously tracked by the NHPD. Meanwhile, the large change in the number of active Section 538 subsidies can be attributed to data updates for these programs that were more than a few years outdated. To view a version of the Data Dictionary that lists all changes side by side, click [here](#).

#### *Number of Active Subsidies Before and After September 2017 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	21,574	19,989	-7%
<b>Section 202 Direct Loans*</b>	1,721	1,533	-11%
<b>State HFA 236</b>	206	35	-83%
<b>HUD Insured*</b>	5,826	7,896	36%
<b>LIHTC*</b>	37,865	33,352	-12%
<b>HOME*</b>	20,639	18,461	-11%
<b>Section 515*</b>	14,719	14,080	-4%
<b>Section 538**</b>	424	748	76%
<b>Public Housing*</b>	6,890	6,778	-2%
<b>Total</b>	<b>109,864</b>	<b>105,410</b>	<b>-4%</b>

\*Data updated (previous data outdated by less than one year)

\*\*Data updated (previous data outdated by more than one year)

#### *September 2017*

In September 2017, the NHPD was updated to incorporate new variables, prepackaged extracts, and enhanced user features. Data processing updates were also made to improve data accuracy and reduce the number of duplicate properties in the database. To learn more about these changes, click [here](#).

#### *February 2016*

The February 2016 refresh incorporated ten data file updates, which led to a significant change in the number of total active subsidies. The number of active Home Investment Partnership (HOME), Section 202, and Section 515 subsidies changed significantly because the data for both of these programs was outdated by more than three years. This refresh incorporated new subsidies that have been awarded or terminated during this time. Additionally, the HOME data used previously did not include subsidies awarded to properties with less than five units. The number of Section 202 Direct Loans decreased significantly because some owners prepaid their mortgages and there were no new Section 202 Direct Loans awarded. Section 202 assistance is now awarded through rental assistance contracts, which are tracked under the Section 8 PBRA program.

#### *Number of Active Subsidies Before and After February 2016 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	21,733	21,574	-1%
<b>Section 202 Direct Loans**</b>	2,231	1,721	-23%
<b>State HFA 236</b>	214	206	-4%
<b>HUD Insured*</b>	5,551	5,826	5%
<b>LIHTC*</b>	34,220	37,865	11%
<b>HOME**</b>	6,451	20,639	220%
<b>Section 515**</b>	10,877	14,719	35%
<b>Section 538</b>	422	424	0%
<b>Public Housing**</b>	6,813	6,890	1%
<b>Total</b>	<b>90,218</b>	<b>109,864</b>	<b>22%</b>



\*Data updated (previous data outdated by less than one year)

\*\*Data updated (previous data outdated by more than one year)

## February 2015

The February 2015 data refresh incorporated five new data files, an update to the logic used to determine which HUD Insurance programs would be included in the database, and an update on how LIHTC subsidy status is calculated. Previously, all properties that were HUD Insured were included in the database. During this refresh, selected HUD Insurance programs that are not affiliated with affordable rental housing were excluded from the database. This change resulted in a reduction of the number of active units in the database by one million. For a list of HUD programs included in the database, see the Data Dictionary.

Additionally, LIHTC subsidies listed as non-programmatic in the LIHTC Database were reclassified as inactive, which resulted in a significant reduction in the number of active LIHTC subsidies.

### *Number of Subsidies Before and After February 2015 Refresh*

<b>Subsidy Type</b>	<b>Before Refresh</b>	<b>After Refresh</b>	<b>Change</b>
<b>Section 8 PBRA*</b>	21,227	21,733	2%
<b>Section 202 Direct Loans</b>	2,266	2,231	-2%
<b>State HFA 236</b>	228	214,206	-6%
<b>HUD Insured*</b>	13,964	5,551	-60%
<b>LIHTC*</b>	37,184	34,220	-8%
<b>HOME</b>	6,595	6,451	-2%
<b>Section 515</b>	11,001	10,877	-1%
<b>Section 538</b>	422	422	0%
<b>Public Housing</b>	6,813	6,813	0%
<b>Total</b>	<b>101,407</b>	<b>90,218</b>	<b>-11%</b>

\*Data updated (previous data outdated by less than one year)

\*\*Data updated (previous data outdated by more than one year)

## Data Integration and Cleaning Protocol

### Procedure for Integrating New or Updated Data

The NHPD is updated three times a year. At these times, any updates made to source datasets by the dataset originator (such as the Department of Housing and Urban Development) are imported into the Database. A list of all data sources included in the NHPD and their most recent update date can be found [here](#).

As data quality and data format vary by data source and each data source may contain duplicate property entries, automated procedures have been created to standardize imported data and reduce the number of incorrect or duplicate entries in the NHPD. During the import process, data inconsistencies that cannot be corrected through the automated process are flagged for manual cleaning. Manual cleaning takes place at each tri-annual data update and on an ongoing basis. The procedures used for both automated and manual cleaning are described below.

### Automated Cleaning Procedures

Automated cleaning procedures center on correcting property addresses and latitude and longitude values, as these fields are the primary matching keys for identifying and linking all of a property's subsidies. As property addresses are imported into the Database, they are standardized according to

USPS standard address protocols and extraneous characters or words appearing in addresses are attempted to be removed. Likewise, property names are standardized and extraneous characters are attempted to be removed. These procedures improve the rate of positive property matches between data sources.

Once addresses are standardized, they are matched and compared to existing address records in the database using subsidy ID. If the address is new or has changed, it is entered into an address verification system. The system currently utilized is based on US Postal Service Coding Accuracy Support System (USPS CASS) certification provided by Smarty Streets and Melissa Data. These address verification systems return latitude, longitude, and other geographic information about the area the property is located in. Geocodes from Melissa Data and Smarty Streets are only kept for building level property matches. If the address verification systems do not return a match, the latitude and longitude provided from the originating data source are used if available.

After addresses have been standardized and verified using Smarty Streets and Melissa Data, they are matched to comparable addresses in the database using the following fields:

- Property address, city, and state
- Property name, city, state, zip code, and total units +/- 2
- Property ID
- Latitude and longitude
- Referencing subsidy ID

Any subsidy record that matches based on these fields is considered to be awarded to the same property. If two records with different property addresses are matched by these rules, an admin user select which property address to display in the database. Any property records missing street address, city, state, total units, latitude, or longitude are flagged for manual review and are withheld from the database until these fields can be populated.

### Manual Cleaning Procedures

Several types of data issues lead to manual review and cleaning. First, all properties that do not CASS certify with a valid USPS address in the automated cleaning process are flagged for manual review. These properties are checked using Google Maps to validate the address and are manually cross checked to the NHPD to ensure that there are no duplicate properties located in the database once the address is updated. Any changes made to property and subsidy records in the database are retained after subsequent data updates. Several common address errors and their corresponding cleaning protocol are listed below.

#### **Incomplete or Incorrect Address: Incomplete or Incorrect Address:**

- Case 1: Address is incomplete and does not contain a house number. (Ex. Main St.)
- Case 2: Address is a set of cross streets. (Ex. 5th and Vine)
- Case 3: Address contains no street address. (ex. apartment name or city is repeated in street address line)
- Case 4: Address contains misspellings (ex. 100 Mairn St., Phonix, AZ)
- Case 5: Address contains incorrect information. (ex. 100 Main St., Phoenix, AR)
- Case 6: Address contains a range of street numbers. (ex. 1-100 Main St., Phoenix, AR)

**Solution:** The correct property address is researched by googling the apartment name and location to identify the official address and by using google maps to verify the address and identify a corresponding building footprint. Once a correct address is found, it is cross checked to the NHPD to ensure that a duplicate property with the correct address is not present. If a duplicate is found, the subsidy information for the duplicate property is merged. Each of the corrected addresses are CASS certified using SmartyStreets. If the corrected address does not CASS certify the latitude and longitude provided by Google Maps is entered into the record. If an address is too incomplete to identify a property's location, the property is flagged as 'incomplete' and remains flagged for cleaning. It cannot be updated until more information is received from the source data.

- Case 6: Address does not offer a mail receptacle.

**Solution:** The address is viewed on google maps to determine that the building footprint is viewable. If the property address is confirmed, the latitude and longitude provided by Google Maps is entered into the record. If the footprint cannot easily be confirmed, the latitude and longitude provided by the source data is retained and the property remains flagged.

Second, all properties that have received a comment from users or staff from [databasefeedback@preservationdatabase.org](mailto:databasefeedback@preservationdatabase.org) are flagged for manual review. Comments may pertain to incorrect address information as described above, indicate that a property is a duplicate property, state that a property's name has changed, or other data issues that require change and verification. Several common duplicate and name discrepancy scenarios and their corresponding cleaning protocol are listed below.

#### **Property is a Duplicate Entry:**

- Case 1: Properties with same street name and city/state, same or similar name, unit count is +/- 2.
- Case 2: Properties with same property name, city/state, unit count +/- 2, different street address.

**Solution:** The main property address is validated as described above. Then the subsidies located at the duplicate property are attached to the property with the valid address or more subsidies.

#### **Property Name is Incorrect or Has Changed:**

- Case 1: Property name on Google Maps is different.
- Case 2: Property name on official website site is different.

**Solution:** The correct name is verified using Google and Google Maps and is replaced.

- Case 3: Property name is correct, but can be linked to a duplicate property with different address and different property name (most likely from a new data source).

**Solution:** The correct name is verified using Google and Google Maps. The subsidy from the older dataset with the incorrect name is linked to the property with the updated name from the newer dataset.

Third, properties may be flagged for manual review if there are major inconsistencies in property information between data sources, but the property can be linked to another property from a different data source through the use of (HUD) IDs. Likewise if there are major inconsistencies in property information from one update to the next in a particular data source, the property is flagged for manual review. The discrepancies are verified and if they cannot be verified, one source is chosen based on staff's assessment of data quality and the radical nature of the change.