

## Frequently Asked Questions

### **Why is the Quarter 4 Ending December 31, 2005 file missing counts in the majority of the vacancy and no-stat columns?**

The starting point for counting the number of days that an address is in no-stat or vacancy status is November 18, 2005. Therefore, all the no-stat and vacancy counts in the December, 2005 tract-level summary file fit into the 0 to 3 month no-stat and vacant categories. Because the USPS started counting in November there is no data for the remaining vacancy and no-stat categories in the December, 2005 file.

### **How do I find my community in the national tract-level file?**

The GEOID field in each tract-level file contains the concatenation of the state FIPS code, the county FIPS code and the Census tract code. By querying this field you can pick out states, counties, or specific tracts within a county. If you are unsure what your county or state FIPS code is, go to FIPS County Lookup page on the U.S. Census Bureau's website at <https://www.census.gov/geo/reference/codes/cou.html>. So for example, if you would like to select the census tracts in Bergen County, NJ you would construct a query that selects all the records where GEOID begins with '34003' because the FIPS code for New Jersey is '34' and the county code for Bergen is '003'.

### **How do I read the tract-level summary files?**

The tract-level summary files are downloaded as zipped DBF files. Once you've unzipped them, you can use a variety of programs to read the files. Excel is the most popular reader for DBF files, but we highly recommend that you DO NOT use Microsoft Excel to read these files. Microsoft Excel 2003 has a maximum row limit 65,536, although that limit may have been increased in later versions of Microsoft Excel. There are 65,688 records in each tract summary file, so by opening the file in Excel you risk losing some of the data. There are many other software packages that read DBF files, such as SPSS, SAS, or Access. Most GIS software will also read in DBF files. We are considering providing an option to download the tract summary files in text format.

### **Why am I getting a Microsoft Jet database engine error when I try to import the DBF file into Access?**

For some reason, Access only recognizes a DBF file if the name of the file is 8 characters or less. If you want import the DBF files into Access, then rename them using 8 characters or less.

### **Can I obtain the non-aggregated version of the data or a level of aggregation finer than Census tract?**

No. The ability to provide vacancy and no-stat information to HUD was thoroughly reviewed by the USPS Privacy and Legal departments in coordination with Address

Management. This review determined that the information could be provided only if use was aggregate to a Census tract level. Aggregates at a Census tract level virtually eliminate the chance of informational disclosure of individual addresses and their occupancy status.

### **Are P.O. Boxes counted as addresses?**

In urban areas, addresses whose residents collect mail at P.O. boxes instead of through regular delivery, are counted as no-stats, even if there may be someone living there. Addresses on rural routes that use P.O. boxes instead of delivery are not counted as addresses.

### **Will the business/residential/other distinction be made available for all extracts made prior to June 30, 2008?**

No. HUD has requested this from USPS and the request was denied

### **Should I include no-stats when calculating a vacancy rate?**

It can be very difficult to calculate a vacancy rate because technically some no-stats are vacant too. No-stats are addresses that have been more or less abandoned or addresses that are under construction and are not yet ready to be occupied. There is no real way to tell the two apart in the dataset. That's why it helps to have local intimate knowledge of the area you are studying. This can help you decide on whether or not to include no-stats as part of the vacancy calculation. In lieu of local knowledge, you might consider using other neighborhood characteristics to get an idea of what kind of neighborhood it is. For example, if you are dealing with a high poverty neighborhood, there is a good chance that many of the no-stats are vacant or abandoned structures. Another indication may come from some other fields in the dataset. For example, you might look at the no-stat 12 to 24 month count. If you see increasing numbers of addresses that are staying in the no-stat category over time, it might signify a neighborhood in decline. In another example, you might look at AMS (total addresses). If AMS is growing and you see a similar growth in the no-state 0 to 3 month category, it is possible that there is new construction occurring. But if AMS is holding steady or declining and you see increase in no-stat, then that could be a sign of a neighborhood in trouble. A straight up decline in AMS may suggest demolition.

### **I'm observing an implausibly large fluctuation in the number of addresses in a census tract over time. What's going on here?**

We recently discovered that our geocoding methodology coupled with the some of USPS business practices have been producing some anomalies in the data over time. First, we are not able to determine a tract code for every zip+4. We do for the vast majority (99%), but sometimes we can't because we don't have that particular zip+4 in our geocoding engine. However, our base files are updated regularly and some zip+4 that did not geocode in one quarter will geocode in the next and get added to the mix. The result of

this is that you may see spikes in the total address count of a tract that cannot necessarily be attributed to growth since the previous quarter.

Somewhat related to this problem, is the issue of splitting zip codes. Sometimes the USPS will split a zip code, which results in new zip+4 being created, while some are actually deleted. This can also manifest itself in weird spikes or drops in total addresses from quarter to quarter that again, cannot necessarily be attributed to immediate growth or decline. When the a zip code is split, our geocoding takes a while to catch up. The USPS publishes every year a product that identifies the major zip code boundary adjustments that were made. You can download it at <http://ribbs.usps.gov/>. Refer to this document if you suspect that this problem is affecting your area of analysis. We are working on a couple of ideas on how to deal with this problem, so there may be some changes in the data format coming soon.

### **USPS vs. American Community Survey**

HUD compared the 2006 ACS housing unit count (a special tabulation from Census) and residential USPS address counts at the tract level (including no-stats) and we found that, in general, the USPS count was consistently lower in non-metro counties than the Census' housing unit count. Meanwhile, in metro counties, the USPS count was consistently higher, but closer to the Census count than in the non-metro counties. In one particular county, for which we have a very good parcel coverage that includes zoning designations, we compared the county's residential address count with both USPS and Census and found that the Census count was consistently lower than the county's count at the census tract level, while the USPS count was consistently higher, but closer to what the county was reporting.

### **What types of addresses are in the “Other” category?**

Little is known about the mysterious other category except that it is used when the USPS is unable or unwilling to categorize an address as either business or residential. There are several possible reasons for this. One is that the address might function as both a business and residential dwelling. USPS may decide to remain ambiguous about it's classification to avoid having to make a decision on what to call it. It is also unclear what large public facilities or government agencies are categorized as. These addresses may be characterized as other.

### **What happened in 2010 that caused an increase the vacant address counts?**

In early 2010, USPS began using the 2010 Census mailing to improve the accuracy of their vacancy indicator. Because the Census mailing involved persistent mailing to non-respondent addresses, postal carriers were better able to determine an address' status as vacant. Carriers counted an address as vacant if they saw repeated Census mailings to that address, and would categorize an address as vacant even if the address was not flagged as vacant for the preceding 90 days—which is the normal USPS protocol for determining vacant status. As a result, the vacancy data shows an increase in the number

of vacant addresses in 2010. Because the Census mailings took place over several months, this change in definition affected the vacancy counts in quarters 2 and 3, not just in quarter 1. This change in methodology also effected length of vacancy: Addresses newly determined to be vacant were moved directly into the 0-3 month category. Consequently, there is a noticeable shift in the numberof vacant addresses in the 0-3 month category in 2010.