

## pdGeoTIGER – Getting Started Guide

**pdGeoTIGER**, now at version 2, is a comprehensive United States national Address Range database developed to permit exceptionally precise assignment of Geo location designations (more technically known as Topographical Faces).

This product covers all 50 states, the District of Columbia and Puerto Rico and is drawn from the most recent edition of the U.S. Census Bureau *TIGER/Line Shapefiles*.

Latitude and longitude coordinates can be assigned along with multiple GeoCoding location designations for both 2010 and 2000, including: State, County, County Subdivision, Puerto Rico Subbarrio, Census Tract, Census Block, Census Block Group, Urban Area, Consolidated City, Place, Congressional District, State Legislative Districts, School Districts, Voting District and Zip Code Tabulation Area.

*pdGeoTIGER 2* continues to be in development while the U.S. Census Bureau completes its current work on the *Shapefile* data. Additionally, redistricting is not completed for all states, and there are many challenges in the courts to the current political district line designations. A finalized version of this product will be released in 2012. However, even after finalization, updates will be made available on a regular basis.

Note that all purchasers of a pre-finalized version of *pdGeoTIGER 2* will have an opportunity to download the finalized version **free of charge** once released.

### Included Database Files

Three different file formats are included to insure compatibility with any database system. Each of the file formats contains the same data. 53 separate Address Range files are provided in each format for each state, the District of Columbia and Puerto Rico. Texas is divided into two files due to its size. The xxxxx in the list below is a placeholder for the state name:

- xxxxx.CSV (comma-delimited ASCII text—MS Excel compatible)
- xxxxx.TXT (fixed length ASCII text)
- xxxxx.DBF (dBase/FoxPro—easily imports into MS Access)

## Included User Documentation

- README.txt (general information)
- This Getting Started Guide
- pdGeoTIGER-Layout.pdf (layout & data definitions)
- Distance-Formulas.pdf (distance calculation formulas & instructions)
- Site License (see Licensing below)

## Version Number

The version number of each copy of *pdGeoTIGER* is written in the first or second row (depending on format) of the first column of the database files in **X.X.X.X** format. The first number is the version of the release. The number after the first dot is the update for the version indicated. The number after the second dot references a minor revision.

During the period this product is being further developed towards its final version in 2012, a fourth indicator will be sequentially added to the version number: Alpha, Beta, Gamma, Delta, etc. (We could have used A, B, C, D, etc., but the Greek letters are more fun.) When the fourth indicator is gone, *pdGeoTIGER* will be finalized until version 3. However, updates will continue to be made available on a regular basis.

## Layout & Data Definitions

File specifications can be found in the included *pdGeoTIGER-Layout.pdf* file. This gives the name of each field, the character type (character/alphanumeric or numeric), the width (including any decimal points, the starting point and the ending point) and the definition of the field contents.

## Instructions

*pdGeoTIGER* is drawn from the most recent edition of the U.S. Census Bureau *TIGER/Line Shapefiles*. 53 separate Address Range databases are provided for each state, the District of Columbia and Puerto Rico. Texas is divided into two files due to its size. This product permits exceptionally precise assignment of Geo location designations (more technically known as Topographical Faces). Matching is based on Address Range.

Address Ranges point to a sequential line of potential addresses and not individual addresses. All possible house numbers are included in the range, from the first structure to the last and all numbers of the same parity (odd or even house number) in between, regardless of if the actual structure currently exists.

The files contained in this database product are set up for easy Address Range matching. Users can simply match the **MATCHKEY**, **FROMHN** and **TOHN** fields against their lists to append the Geo location information contained in the fields beginning with **STATENUM**. Alternatively users can match directly against the parsed address fields: **PREQUAL**, **PREDIR**, **STREETNAME**, **STREETTYPE**, **SUFDIR**, **SUFQUAL**, **ZIP** and **PLUS4**, along with the **ADDRPARITY**, **FROMHN** and **TOHN** fields.

The parity of an address (contained in the **ADDRPARITY** field) is important because Address Ranges only include addresses of the same odd/even parity.

Latitude and longitude coordinates can be assigned along with multiple GeoCoding location designations for both 2010 and 2000, including: State, County, County Subdivision, Puerto Rico Subbarrio, Census Tract, Census Block, Census Block Group, Urban Area, Consolidated City, Place, Congressional District, State Legislative Districts, School Districts, Voting District and Zip Code Tabulation Area.

Latitude and longitude coordinates are provided in three formats: six-decimal, degrees/minutes/seconds (DMS) and converted to radians (for use in trigonometry functions—15 numeric places). The decimal and radian formats are most useful when calculating distances and radiuses while the DMS format is ideal when printing out the coordinates in documents and on websites.

For those calculating distances and radiuses, formulas and instructions are contained in the included *Distance-Formulas.pdf* file.

Peacock Data, Inc. offers Address Range databases instead of ZIP+4 databases because Address Range matching is significantly more accurate. The reason for this is USPS ZIP+4 boundaries have no relationship with U.S. Census lines, and many ZIP+4 Codes fall into multiple Census Geo locations or *Shapefile* polygons.

There are almost 62 million records contained in this product, which is 25 million more Address Ranges compared to version 1. However, while coverage is very high, users should be aware the Census *Shapefiles pdGeoTIGER* is based on continue to have gaps and limitations. There are Address Range overlaps, odd/even reversals and low/high orientation reversals. With the exception of overlaps, these may be valid. Peacock Data has corrected most of low/high orientation reversals.

The *Shapefiles* also do not provide data for every field in every record. For example, Voting District numbers (**VTD**) are not assigned when a state or locality does not follow Census Bureau guidelines when drawing ward and precinct lines. And, as another kind of example, the **SUBMCD** field is only filled in the Puerto Rico database; this is the Subminor Civil Division (Subbarrio) FIPS code number.

The *Shapefiles* generally contain only Address Ranges with *house number-street name* style addresses. It does not have rural route and post office box addresses because these do no reference a specific location latitude and longitude coordinates can be applied to. It also does not include information for some small

places where the USPS provides only post office box service, as well as for Single Address-Address Ranges, including out-of-parity and out-of-sequence ranges that cover only a single house number.

The included databases, however, may contain structure numbers assigned for use by local emergency services, but not recognized by the USPS for mail delivery.

Address Ranges can include house numbers with alphabetic characters, which help to uniquely identify addresses, or hyphens, which separate avenue numbers, private road designators and grid cell numbers from structure numbers.

Some Address Ranges may have a full 9-digit ZIP Code, with the first five digits contained in the **ZIP** field and the final four digits in the **PLUS4** field. The purpose of these is to differentiate addresses that are duplicated within the same 5-digit ZIP Code. When a full 9-digit ZIP Code is made available, it is advised to first try to match against the Address Range with the full code, and only secondarily against the Address Range with the 5-digit ZIP Code.

Because so much of *pdGeoTIGER* is made up of Peacock Data's own variations and proprietary information, a new field (**SOURCETYPE**) is provided with this version as a source type indicator. The values in this field are as follows:

- 1 = Census *Shapefile* data
- 2 = Census *Shapefile* data algorithmically enhanced to accommodate new address ranges
- 3 = Peacock Data corrected Census *Shapefile* data
- 4 = Peacock Data variation or proprietary data - high confidence
- 5 = Peacock Data variation or proprietary data - medium confidence

Other new fields in version 2 of this product include: 2010 Census Urban Area Code (**URBANAREA**), corrected 2000 Census Urban Area Code (**UA00COR**) and 2010 and 2000 Consolidated City FIPS code numbers (**CONCITY** and **CONCITY00** respectively).

## Licensing

If purchased as a standalone product, the Site License for this product can be found in the included *pdGeoTIGER-Site-License.pdf* file. If purchased as part of *pdSuite Master Collection*, it can be found in the included *pdSuiteMC-Site-License.pdf* file. Users are allowed to install *pdGeoTIGER* on all computers within a single company or organization (see the license for details). A Developer License is also available for separate purchase granting users the right to incorporate the information in for-profit services and products.

*pdGeoTIGER* is Copyright © 2009-2011 Peacock Data, Inc. All Right Reserved.