

## Guideline for Geocoding

Geocoding is the first step in using locational data with geographic information system software. Examples of public health studies utilizing geocoded data include tracking the spread of infectious and environmentally caused diseases, identifying factors related to low birth weight, and linking client populations to provider locations.

Geocoding is the process of assigning coordinates, most commonly latitude and longitude, to a street address. For example:

Original Address		Geocoded Address		
Address	ZIP	Address	ZIP	Latitude Longitude
1100 W 49 <sup>th</sup> St	78756	1100 W 49 <sup>th</sup> St	78756	30.8215 -97.735514

The coordinates represent the approximate location of the street address, but when it is not possible to locate the street address, the coordinates can represent the center of an area, commonly a ZIP code. Geocoding relies on two data sets, a reference street data set, which can include not only streets but also ZIP code boundaries, and the data containing the street addresses to be geocoded. If the street address data are not complete or correct or if the street reference files are not complete, an address may not geocode to street.

It is important to remember that ZIP codes can change on a continuous basis throughout the year. As postal delivery needs change in an area, the US Postal service can add or delete ZIP codes and modify the shape of ZIP code boundaries. Researchers often want to analyze data over time making ZIP codes less than ideal as a unit of analysis. It is recommended that the smallest spatial unit of analysis (i.e. street address) be used. Once data has been geocoded to a street address it is possible to aggregate it to different spatial units of analysis such as census tracts.

A complete physical address is needed in order to assign accurate map coordinates. Here are some rules for recording an address:

- Always use a street address. Records coded to post office boxes or rural routes cannot be used in most studies. A physical address and not a mailing address is necessary;
- Do not use punctuation in the address; and
- Use standard United States Postal Service abbreviations.

To enhance address usability for geocoding, address data elements ideally should be stored and exchanged consistently as text values and not as numeric values.

Element	Size	Comment
Address line	55	Follow United States Postal Service general addressing standards
City	28	
ZIP code	5	Use lookup tables to reduce data entry errors
ZIP+4 extension	4	
Longitude	11	Store as decimal degrees and not as degrees, minutes, seconds
Latitude	11	Store as decimal degrees and not as degrees, minutes, seconds
Match code	4	Geocoding accuracy results

For more information, contact [gis@dshs.state.tx.us](mailto:gis@dshs.state.tx.us)