Best Practices –Preparation of Address Data for Geocoding

Address Format

These address format guidelines are the same for all file formats. Details specific to a particular file format follow.

Separate your address information into different columns.
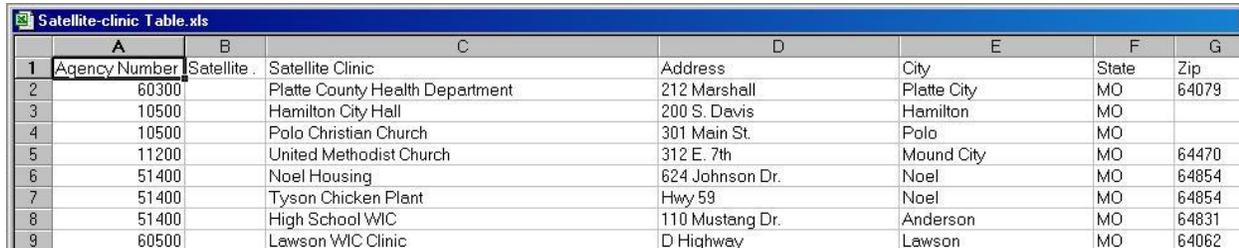
	A	B	C	D	E	F	G
	Agency Number	Satellite .	Satellite Clinic	Address	City	State	Zip
2	60300		Platte County Health Department	212 Marshall	Platte City	MO	64079
3	10500		Hamilton City Hall	200 S. Davis	Hamilton	MO	
4	10500		Polo Christian Church	301 Main St.	Polo	MO	
5	11200		United Methodist Church	312 E. 7th	Mound City	MO	64470
6	51400		Noel Housing	624 Johnson Dr.	Noel	MO	64854
7	51400		Tyson Chicken Plant	Hwy 59	Noel	MO	64854
8	51400		High School WIC	110 Mustang Dr.	Anderson	MO	64831
9	60500		Lawson WIC Clinic	D Highway	Lawson	MO	64062

Satellite-clinic Table.xls

- There must be separate fields for address, city, state and ZIP code. The use of two address fields is acceptable, assuming you put the proper content in each.
- Address fields should only contain address information, not Attention to someone's name etc. If you need to retain this information, create another column for it.
- When two fields are used for address the primary address must consistently be in one column (123 Main St) and the supplementation information in a 2nd field (Suite 203).   It is not OK to interchange the type of information between the two columns.
- For addresses that have both a street address and a PO Box, the street address is primary the PO Box is secondary.
- If the address in only a PO Box, it's OK to have it in the primary address column.
- PO Boxes will never match to a particular building; they will be placed in the center of the ZIP code.

Must have a column for state

All input data should include a column for STATE. It is common for users to leave this column out, assuming everyone will know the state is Missouri. (We actually routinely deal with data out of state.) The software used for geocoding can't make these assumptions. If there is no state column present in your data, you must add one, and populate it with a state abbreviation (MO).

Watch for Direction Prefix and Street Type

- Directional prefixes are often necessary to achieve a successful match. If there is both a 123 N Main St and a 123 S Main St, and you only enter 123 Main St, the result will match to a ZIP center.   The software will not pick a match out of the two possible candidates.
- Similarly, the street type matters. For example 123 Main St and 123 Main Ter. If you enter only 123 Main, the software will not pick a match out of the possible candidates

Watch out for abbreviations in your street name

- Street names containing abbreviations will frequently fail to find a match. You can spell it poorly, but you have to try to spell it out. Some common ones:

- o Martin L King
- o ML King
- o MLK
- o JFK
- o James CP Bell (this stands for James Cool Papa Bell)
- o MO Blvd (all state abbreviations must be spelled out – in this example change to Missouri Blvd).

- There are exceptions for Saint, Mount, Fort, and other commonly used abbreviations. Something like 123 St Louis Blvd is acceptable.

- Note: Some French or Spanish names may require a space within the name to match correctly. Examples: Right: De Soto, De Paul, Wrong: Desoto, Depaul.

St Louis City is not a city name. It's just St Louis.

St. Louis City is <u>not</u> a city name; it is a county jurisdiction. St. Louis County is also not a city name. For geocoding, make sure these names are edited in the city name columns to St. Louis.

Standard abbreviations for Saint, Sainte and Mount are allowed (can be St, Ste or Mt). All other city names should be spelled out. (Right: Kansas City, St. Louis, Springfield, and El Dorado Springs. Wrong: KC, STL, Spfd, and El Dorado Spgs)

Not all counties have street style addressing

- Counties that don't yet have 9-1-1 are not likely to have street addresses assigned in their county. There are still around a dozen counties in Missouri that haven't implemented 9-1-1.
- Addresses based on private streets have a tendency not to match. Private streets may or may not make their way into the databases that are used for geocoding.
- The chances matching 100% for your address file to a specific location is unrealistic. Anything over 90% is considered good.

Numbered Streets

Where street names are actually a number, for example 1st Street, you should use the numeric value instead (1st as opposed to First).   These may be returned as unmatched (depends on the geocoding software you are using). Change these names to their numeric equivalent (Right: 1st, 2nd , Wrong: First, Second).

Addresses based on highways

Addresses on highways, whether they are US highways, county routes or even frontage roads are problematic for two reasons.

- The source road network used by most common geocoding software is often lacking house numbers for highways. The highways appear with their name, but the house numbers are all 00000 (zeros).
- The highway name has to be very specific. For example, if the address is 1234 C Hwy, the geocoder may have it as 1234 County Route C.   The exact rules on the naming convention for

each 'level' of highway still need to be researched. The correct name is based on the US postal service.

- Counties which have their address style like 1234 Lawrence 2457 won't process correctly. Try changing these to 1234 County Rd 2457.

File Types

Excel
There are some specific formatting requirements to use Excel files with GIS. If you have multiple worksheets, process one worksheet at a time.

Delimited ASCII (text)
If you have delimited text, it is because you are likely extracting this data from some application that generates this format. If you have an option at all for a field delimiter, make sure that you choose something besides the comma (,). Address data often contains commas within a single column, and the data will be shifted over into the next column. The recommended alternatives are Tab or the pipe (|) symbol, generally found above the backslash (\) key.

A note about data that was copied from a web page
Data that was copied from a web page into Excel should always first be saved in the TXT file format. This will remove any formatting (HTML) that is still associated with the data.

File Preparation – Excel

Files that are used for geocoding frequently contain extra information besides just the address columns. Any extra columns you have will be carried forward into the geocoded output. However, Excel files tend to have header and footer information that needs to be modified before submitting a file for geocoding.

Here are some general guidelines for making any Excel file 'geocode ready', beyond the formatting of the address information.
- Keep a copy of your original file.
- Use only one line for your column headings, don't create 'stacked' headings, or multiple lines to represent the header.
- Brief headings are preferable.
- There should be NO additional header, footer or sidebar comments in the document. This includes summary lines, totals, references, citations, etc.
- Headings should start with a letter, not a number.
    Example: 2002rate should be changed to rate02 or rate2002
- Brief headings are preferable.
- Field names should be in all UPPER case or all mixed/proper; either is OK, as long as it is consistent.
- Field names must be unique
- The same consistency applies within the data. Text fields should contain data in all UPPER case (e.g. COLE, OSAGE, BOONE) or all mixed/proper case (e.g. Cole, Osage, Boone).
- For number data do not use 1000 separators. This can be changed under Format Cells > Number Uncheck 'Use 1000 separator (,)'.
- As long as you have all the required address fields, any other data you want to be associated with the output shape file can be included; you aren't limited to just the address data for geocoding.

- Remain consistent with your capitalization. If you have multiples files, they should contain data in all UPPER case (e.g. COLE, OSAGE, BOONE) or all mixed/proper case (e.g. Cole, Osage, Boone).
- None of the cells should be a calculated or summary value based on the data of other columns. Create a new column and Copy and Paste Special, selecting Values only as you paste to convert to plain numbers (or text).
- Within the data itself, particularly for character/text data make sure there are no leading blanks. Look closely at how the data aligns in a column. Text should appear justified to the left, without spaces before the first character.
- **Bold** text, boxes around cells and other cosmetics can be left as-is. This formatting will be stripped off automatically during the conversion from Excel to the other formats.
- Don't highlight cells or otherwise color code them to somehow classify your data; this information will not be available to you in ArcView. Instead, add additional columns and create a value to represent the meaning of the colors.

File Preparation – Text

Text data will not contain formatting, such as the use of color or bold as a method to classify some values. There are a few guidelines to consider which are in common with the Excel file formatting.

1. The first line should contain your field names.
- Brief headings are preferable.
- Field names should be in all UPPER case or all mixed/proper; either is OK, as long as it is consistent.
- Field names must be unique.
- For numeric data, do not use 1000 separators (1000 not 1,000).
- If you have commas within a field, either delete them, or use a delimiter such as Tab or pipe (|). The software will interpret all commas as the start of a new field value.
- Remain consistent with your capitalization. If you have multiples files, they should contain data in all UPPER case (e.g. COLE, OSAGE, BOONE) or all mixed/proper case (e.g. Cole, Osage, Boone).
- The same consistency applies within the data. Text fields should contain data in all UPPER case (e.g. COLE, OSAGE, BOONE) or all mixed/proper case (e.g. Cole, Osage, Boone).
- Particularly for character/text data make sure there are no leading blanks.
- For number data do not use 1000 separators.
- As long as you have all the required address fields, any other data you want to be associated with the output shape file can be included; you aren't limited to just the address data for geocoding.

Last Updated March 11, 2015
Tracy Schloss