Considerations for Generating Location for Mapping and Analysis

Background:
Mapping and analysis applications like ESRI's GIS products and Tableau can import address level data when it include fields for coordinates such as latitude and longitude (lat/long). Not only is a graphic generated, but all other fields in the table are available for mapping, analysis and query.

Ideally these coordinates are maintained as part of the location information of a record, along with address and county.

Products to validate addresses and generate lat/long are generally known as geocoding software. Geocoding software provides three main functions: 1) it returns corrected clean addresses (also known as address hygiene); 2) it provides the ZIP+4 for the address and 3) it returns a latitude/longitude coordinate pair. (Note: Not every geocoding products returns all three.)

When considering geocoding options, it's necessary to be able to execute a single address, as in the point of data entry within an application, or through some sort of batch routine. Within Missouri applications, the Pitney Bowes platform is commonly used for both.

Things to consider:

- A batch routine is useful when coordinates do not exist in a table, and address data is needed for input into a mapping or analysis product. Data can be extracted out of the system, and geocoded as a separate standalone process that generates the necessary coordinates. However since this creates a new output file, this becomes the input to the map/analysis. There is no way to live link to the original source.

- There are other ad hoc approaches. ESRI products have add-ins that can be used to convert addresses to points. As with a batch routine, this process generates a copy of your original data. Tableau has no tools to convert addresses to point; data must already contain the necessary coordinates.

- The Missouri Adaptive Enterprise Architecture has multiple standards for addresses and geographic locations. These can be found under https://oa.mo.gov/information-technology-itsd/it-governance/enterprise-architecture/information-domain

- Minimum fields needed in a schema for storing coordinate pairs.

Field Name	Field Type	Purpose
Latitude	Double	Store degrees of latitude to 6 decimal places. Example 38.123456
Longitude	Double	Store degrees of longitude to 6 decimal places. Example: -91.123456

		NOTE: Longitude is always a negative number in the US.
Location_Quality_Code	Text	Location code reflecting the positional accuracy of the coordinate
Census Tract (optional)	Text	Storing this data provides additional flexibility to aggregate data to a smaller unit within a county

- In some applications, fields for geographic coordinates were included in the schema, but never populated. When reviewing existing databases, don't just look at the schema and assume data is present.

- There are many ways of describing coordinates for X and Y, latitude and longitude are the most commonly known. Technically X = Longitude and Y = Latitude.

- Applications can be written to do geocoding on the fly, validating addresses and storing lat/long at the point of data entry. Sample code is available for developers.

- For existing applications that need to be mapping/analysis ready, these can be processed through a batch routine to populate the coordinates for existing records. Once existing data have been modified, a change to the data entry process for on-the-fly geocoding will populate all future addresses.

- A field for location quality is important to include, in addition to latitude/longitude. Some addresses, especially rural routes and PO Boxes will only be placed to the center of a ZIP code. Maps and analysis based on location need to know which points are placed at this lower level of accuracy.

- The geocoding process is not perfect, because the data we collect isn't perfect. In some instances, additional research yields a more accurate location. Somewhere in the application should be the option to allow for changes to the coordinates and location codes as a more manual process.

- When an application allows changes to these coordinates, a unique location code should be used, and a rule put in place so that the coordinates can't be overwritten by some future automated process.

- Besides returning clean addresses, there are advantages to a geocoding process that stores not only clean addresses and coordinates, but also the county. Applications that allow users to select the county, rather than selecting the county based on the street address can result in error, especially in the metro areas of Kansas City and St Louis. Projects that aggregate data by county will reflect incorrect summaries.

- For a few GIS maps, data has been extracted from an applications, batch processed through a nightly routine that generates the necessary format, appending

latitude/longitude into an output table that is then used for mapping and analysis. Custom scripts had to be written in order for this to work.

- o Since everything is refreshed nightly, this is not a suitable alternative for applications that have allowed manual overrides to the coordinates.
- o This can work for specific applications, but as an overall strategy it is short sighted. Each project would require a custom script, and in each instance a near replication of another data table would be created.

- Mapping software will allow points to be displayed straight from a table or view containing coordinate pairs. However the draw time compared to a native GIS file format is noticeably slower for large data sets (over 50K records).

- When developing schemas to store the output of a geocoding process determine whether addresses have to be stored as-is or if a standardized address is acceptable.

- o If the original address must be stored for legal reasons, there will need to be fields for both original and standardized addresses.
- o Even if the address must be stored once as-is and once standardized, the coordinate pair needs to be stored only once.

- Because Google is perceived as a free or cheap alternative for geocoding, please note the Terms and Conditions of the Google Map Geocoding is restricted.

- o You may only use the results of Google geocoding on a Google map, not with any other products. This means no ESRI, no Tableau, no other reporting software that uses data derived from this source.
- o The use of the Google Map API is not free. If it once was, changes to Google's conditions in 2018 have made this a product that the State is now paying for.

Tracy Schloss
Created 1/24/2003
Last Updated 12/11/2018