

Historic Preservation, Process Development

Contributed by Bert Granberg
01, Aug. 2009
Last Updated 27, Aug. 2009

Notes on the building the overall process for address matching

First Crack at Geocoding Using AGRC's Geocoding Web Service (Barry):

- The complete dataset of addresses Heidi gave me to geocode is here: I:\AGR6\barry_work\HistAddresses2.gdb
- 74989 out of 88243 geocoded with a score of 80 or better (85%)
- 1882 of those remaining geocoded with a score between 60 - 79.
- It took the application about 1 hour 50 min to geocode 10,000 addresses. A total of approx 16 hours of processing time were needed for all 88,243 addresses.
- The application I made to batch process the addresses added 5 new fields:
 - UTM_X
 - UTM_Y
 - MatchAddr (The address in the streets fc the returned match is based on)
 - Score
 - Note (used only for marking those municipalities that were not found in the municipalities list)

- The geocoding application can handle either city or zip code as the zone. The original dataset did not have zip codes, so we added a field called "Zone" and calculated it to be either the city name if it was in the municipalities feature class in the SGID or the zip code that the place name from GNIS is within.

- Heidi corrected most of the municipality names so they matched the spelling in the SGID municipalities, which is used by the web service to search for possible zip codes. There are still some (3439) that are not the same for whatever reason, including some that are "Unknown". The largest part of these, however, are in "Richfield". I'm not sure why those weren't found in the municipalities list. These will probably geocode with a fairly high score if corrected.

- The "Comments" column in the original tables was removed since it did not get along with the delimited file formats I needed to use. It can be added back since the original ObjectID has been kept as a key since I was told that the ID field might not be unique.

Process components:

How to keep track of what's done, by whom, whether its worth keeping (how to build/keep metadata in some sort of standard)

Priorities