# BATCH GEOCODING: OVERVIEW OF TOOLS & OPTIONS

ROB PITINGOLO OCTOBER 2017



Geocoding is the process of converting a mailing address into geographical coordinates. When working with data that only includes the address of a building, business, or person, geocoding is the key to matching that person or place to its corresponding geography. This enables users to display the points on a map or create summary indicators for geographically small areas, such as Census tract or zip code. This brief summarizes the features of a set of geocoding tools, and offers advice on factors to consider when selecting a tool for your project or organization.

## **BATCH GEOCODING TOOLS**

The process of geocoding can either take place interactively or in a batch session. Interactive geocoding is the process of looking up addresses one-by-one on a map or through submitting a form. Most people geocode when they use Google Maps to find the location of a place where they want to go. Batch geocoding involves processing many addresses, sometimes thousands or tens of thousands, through a service that returns the geographic coordinates (latitude and longitude). Some services may return additional results, such as the Census tract or zip-code.

The batch geocoding tools that are available vary on several dimensions, including price, accuracy, and processing speed. In this memo, we only directly compare price. However, we do note differences in accuracy and features when they are commonly known.

- <u>ArcGIS Online</u> is the geocoding service provided by Esri. Even though it is called "ArcGIS Online" geocoding can be done using either their web applications or the ArcGIS desktop application (which must be logged into an ArcGIS online account). ArcGIS Online uses proprietary base files for its geocoding services.
- <u>Census Geocoder</u> is a web-based service provided by the US Census Bureau, which uses Census TIGER/Line files to match addresses to a specific street segment. Users can upload spreadsheets with a maximum of 1,000 addresses per file.
- 3. <u>Geocod.io</u> (rhymes with rodeo) is a web service that offers batch geocoding through their online tool, which is based on Census TIGER base files as well as several other data sources, including the <u>Open Addresses project</u>.

- Geolytics offers an offline <u>GeocodingDVD</u> based on the <u>2010 Census Redistricting data</u> file. The software comes on a DVD and can be run on a Windows PC.
- 5. Google Maps API offers geocoding services via software applications that connect to the Google Maps API and geocodes using Google's proprietary map data. One example of software that uses the Google Maps API is <u>R</u>, a statistical software program. The <u>"ggmap"</u> package is commonly used in R for geocoding.
- Local geocoders are offered in some cities, counties, and/or states that are specific to their local jurisdiction. For example, the New York City Planning Department offers <u>Geosupport Desktop</u>. The District of Columbia city government provides the <u>MAR Geocoder for Washington, DC</u> <u>addresses</u>.
- Nominatim is a tool created to geocode using <u>Open Street Map</u> data. An API is available for free for users who wish to use a tool like Python to create their own geocoding program. There are also third party vendors like <u>Mapzen</u> and <u>OpenCage</u> that offer paid solutions for using Nominatim.
- 8. <u>SAS Street Geocoding</u> is a procedure in the SAS statistical software that uses Census TIGER base files and is officially supported by SAS. This tool requires a license to use the SAS software.
- Smarty Streets is a web-based address verification and listing service which includes geocoding as a component of its product package. It uses TIGER base files for its match and reports the latitude and longitude centroid of the ZIP9 code associated with the address.
- Texas A&M Geoservices offers batch geocoding through their proprietary online tool that is based on locally-developed algorithms, using TIGER base files along with the <u>Boundary Solutions</u> 2012 National Parcel File.

## **GEOCODER PRICING**

The days of unlimited free geocodes (or unlimited geocodes included in a standard license) are over, so users today need to consider the cost of geocoding when selecting a system.

There are pros and cons to all geocoders, including the free options. For example, the Census geocoder does not require payment to use, but can only accept up to 1,000 addresses per batch submission. Similarly, the Google Maps API supports 2,500 free geocodes per day. If a user has a file with hundreds of thousands of records, but is willing to conduct their geocoding process over days or weeks, they

might not pay any service fees; but that also might not be reasonable for project demands and deadlines.

The table below describes what we know about the direct costs of these options. The record limits make it difficult to do a comprehensive apples-to-apples cost comparison of these tools, since we don't calculate the cost of labor.

Tool	Advertised Pricing (As of April 2017)	Estimated Cost for 100,000 Geocodes
ArcGIS Online	40 credits per 1,000 geocodes.	\$400, in addition to ArcGIS license
Census Geocoder	No cost to use the service.	\$0 but limited to 1,000 addresses at a time.
Geocod.io	First 2,500 lookups per day are free, \$0.0005 per additional lookup.	\$48.75
Geolytics	Licenses are sold by number of lookups; 1 million lookups cost \$1,195.	\$120
Google Maps API	Free up to 2,500 requests per day. \$0.50 USD / 1,000 additional requests, up to 100,000 daily.	\$48.75
Local Geocoders	Not all local jurisdictions have geocoders. Those that do may release them to the public for free or charge users for their tool.	Varies
Nominatim	Users can write their own programs to connect to the API for free. Third party vendors also offer paid solutions.	Free (user created API connection) \$37.50 (Mapzen) \$500 (OpenCage)
SAS Street Geocoding	Included in SAS license.	\$0 beyond the SAS license cost
Smarty Streets	250 addresses for free. Monthly subscription with pricing based on maximum number of lookups.	\$500
Texas A&M Geoservices	Services are provided free of charge in allotments of 2500 transactions. For greater than 2,500 transactions, users can purchase credits on a tiered scale, with lower cost per transaction the more credits purchased.	\$175

## ADDITIONAL CONSIDERATIONS

In addition to comparing the costs of the geocoding options listed above, the accuracy, usability and security of systems should also be considered. There is no single best geocoding option, and the needs and resources of the organization should dictate which option is most appropriate.

#### Accuracy

While a direct comparison test was not performed on the above geocoding options, it is known that the accuracy of each service will vary, based on many factors, including:

- 1. Quality of the input address data
- 2. Quality of the base or lookup data
- 3. Ability of the tool to correct typos or errors
- 4. Ability for the user to correct unmatched addresses

For example, the Google Maps API is generally good at correcting typos or identifying the intent of the address input. You can see this in action by simply typing an address into Google Maps with a small typo and observe how it still finds the address. Google would also recognize that an address input on "MLK Avenue" is located on "Martin Luther King Jr Avenue".

On the other hand, the SAS street geocoder generally cannot correct mistakes like this, and will simply return a result of "unmatched". The ArcGIS geocoder can more easily correct unmatched addresses by allowing the user to use the "rematch" tool to manually geocode addresses that the batch geocoder did not successfully match.

#### Usability

The easiest to use geocoders are those that are web-based and simply require the user to format and upload a spreadsheet, like the Census Geocoder and Geocod.io. Tools like ArcGIS and the SAS geocoder require the user to have a basic understanding of those software packages. Using the Google Maps API requires a knowledge of how APIs work and of the tool that's used to connect to the API.

#### Security

Some users have address data that must be kept confidential. Always review any organizational policies, memoranda of agreement, or Institutional Review Board plans before using any geocoding tools. Generally offline geocoders like Geolytics and the SAS geocoder are the safest for preserving confidentiality since they do not require sending data through any online service. For online services, it may be possible to maintain confidentiality by creating a random identifier and then matching the geocoded data back to the source data after the fact.

### SUMMARY

The field of geocoding will continue to change as technology evolves. If you are aware of other useful geocoding tools, email <u>nnip@urban.org</u> and we will consider incorporating them into future versions of this brief. For more general guidance on data management, see Lessons on Data Management *Practices for Local Data Intermediaries* available on the National Neighborhood Indicators Partnership website at <u>https://www.neighborhoodindicators.org/node/5452</u>.



Rob Pitingolo is a research associate in the Metropolitan Housing and Communities Center at the Urban Institute.

Coordinated by the Urban Institute, the National Neighborhood Indicators Partnership (NNIP) consists of independent organizations in over 30 cities that have a shared mission to help community stakeholders use neighborhood data for better decisionmaking, with a focus on assisting organizations and residents in low-income communities.

Copyright © 2017. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.