

# Georeferencing Workshop

March 21, 2008

## Georeferencing Tutorials and Recommended Practices

Topic	Page
Introduction	1
Preparing address data for georeferencing	1
Available address geocoders, pros and cons	1
Using Batchgeocode.com tutorial	2
Geocoding address data in ArcMap tutorial	4
Geocoding address data with Google Earth tutorial link	5
How to geocode non-address data	5
Creating and using a custom (census tract) geocoder in ArcGIS	5



Center for  
Geographic Analysis

Harvard University

## Introduction

The ability to assign specific geographic locations to information (known as georeferencing, or geocoding) is available to anyone with a computer and internet access. The relative ease of georeferencing and resulting accuracy can vary widely depending on a number of factors. What is the nature of the data? How 'clean' is it and what format is it in? What georeferencing technique will be used? Determining a georeferencing strategy that best suits a particular need is not always clear. This document describes various geocoding methods, how to use them, and suggests some best practices for data preparation prior to georeferencing. It is intended for anyone who wishes to perform georeferencing or gain insight to what georeferencing might involve.

The heart of georeferencing is comparing data in tabular form to the table of a Geographic Information System (GIS) data set (layer) which has been already mapped. When matches between the input table and the GIS table are found, the corresponding map coordinates are assigned to the input features, thus georeferencing them. A geocoding service is a program that allows for a user to input a batch of data contained in a table, search for matches as compared to a GIS reference layer, and output the result in map or GIS layer format.

## Preparing address data for georeferencing

As a geocoding service compares input data to a table in an existing GIS layer in search of an identical match, it is of critical importance to eliminate **misspellings**, **special characters** (such as “ ? # ‘ \ %), and **abbreviations** in the input dataset. Address data at a minimum should contain the Street Number, Street Name, Street Type, and Street Suffix Direction (if necessary – many street names don't have a Suffix Direction). Most address geocoders work on street intersections as well, and the intersecting Street Names and Street Types separated by an **&** in one field is the standard format for these. City and State are also required in separate fields, like the example below:

ADDRESS	CITY	STATE	ZIP	STORE_NAME	STORE_TYPE	SALES
1171 PIEDMONT AVE NE	Atlanta	GA	30309	Ace Market	Store	59812
MAIN ST & 41ST AVE	Atlanta	GA	30309	Andrew's Gasoline	Service Station	10843

Notice for the first address that the Street Number (1171), Street Name (PIEDMONT), Street Type (AVE), and Street Suffix Direction (NE) are all separated by a space, and there aren't any periods or other special characters that may lead to confusion. As most address geocoders refer to a streets GIS layer that uses standard U.S. Postal Service abbreviations, following these conventions is also necessary in order to 'match' addresses when geocoding. The U.S. Postal Service's standard Street Type, (or suffix) and State abbreviations are listed at:

<http://www.usps.com/ncsc/lookups/abbreviations.html#suffix>

Zip codes are good to have as well, but usually a good 'match rate' can be achieved without zip codes. Notice as well the additional table attributes STORE\_NAME, STORE\_TYPE, and SALES are included in the table. These will be retained in the output GIS table after the geocoding is performed.

## Available address geocoders, pros and cons

For the first time geocoder, **Batchgeocode.com** is the most useful, easiest to understand, and the CGA has found it to produce the highest match rate among geocoders. Also, it's a free service. Retaining additional attributes from the input address table for use in GIS is also possible when using this geocoder. Batchgeocode.com advises against geocoding more than 500 addresses at once, but we've successfully geocoded batches of more than two thousand.

**ESRI’s StreetMap USA** geocoder retains all address table attributes, and produces a shapefile ready for use in GIS. StreetMap USA requires ArcGIS software, which is very expensive (but free to all Harvard affiliates, and available at many other colleges). ArcGIS software also allows for the creation of a custom geocoder, very handy if one has a more detailed street network, or wishes to geocode non-address data.

**Google Earth** also offers batch geocoding of up to 400 addresses through its spreadsheet mapper utility – a free product. Results are displayed automatically in Google Earth with imagery in the background, and there are options to add informational and imagery content to each address. Converting mapped address data in Google Earth to GIS format is possible through a multi-step process.

**Using Batchgeocode.com tutorial**

1. Download Excel spreadsheet template from [http://www.batchgeocode.com/excel\\_example.xls](http://www.batchgeocode.com/excel_example.xls)
2. Read the instructions in the Excel spreadsheet.  
Key in the addresses into separate fields, or copy and paste addresses into these fields. Make sure to separate street address, city, state and zipcode. In case of European addresses, put the country name in the “State” column.

	A	B	C	D	E	F	G	H	I
1	# Note: This is an example template, keep the column names below but add your own address data.								
2	# not all address fields are required, but the more you have, the more accurate your geocoding will be								
3	# you can keep other fields here as well, such as a unique identifier that maps to a database								
4	# when you are done entering your addresses, copy/paste the data into Step #2 on BatchGeocode.com								
5	Address	City	State	Zipcode	Name	URL	Color	Image	Group
6	1600 Pennsylvania Avenue NW	Washington	DC	20500	The White House	<a href="http://www.whitehouse.gov/">http://www.whitehouse.gov/</a>	FF0000	<a href="http://www.batchg">http://www.batchg</a>	Group 1
7	350 5th Ave.	New York	NY	10118	Empire State Building				
8	1737 Cambridge Street	Cambridge	MA	2138	CGA	<a href="http://www.gis.harvard.edu">http://www.gis.harvard.edu</a>			
9									
10									
11									
12									

3. Once finished, select all the data as shown on the image below. Make sure that you select the field names as well. Then copy them by select “copy” from the “Edit” menu.

	A	B	C	D	E	F	G	H	I
1	# Note: This is an example template, keep the column names below but add your own address data.								
2	# not all address fields are required, but the more you have, the more accurate your geocoding will be								
3	# you can keep other fields here as well, such as a unique identifier that maps to a database								
4	# when you are done entering your addresses, copy/paste the data into Step #2 on BatchGeocode.com								
5	Address	City	State	Zipcode	Name	URL	Color	Image	Group
6	1600 Pennsylvania Avenue NW	Washington	DC	20500	The White House	<a href="http://www.whitehouse.gov/">http://www.whitehouse.gov/</a>	FF0000	<a href="http://www.batchg">http://www.batchg</a>	Group 1
7	350 5th Ave.	New York	NY	10118	Empire State Building				
8	1737 Cambridge Street	Cambridge	MA	2138	CGA	<a href="http://www.gis.harvard.edu">http://www.gis.harvard.edu</a>			
9									
10									
11									
12									

4. Go to the Batch Geocoder website:  
<http://www.batchgeocode.com/>  
Select all of the text in the “Step #2” box, and Delete.  
Right click in the box and Paste your data into the box, as shown on the next page:

**Step #2**

Copy/paste addresses into the table below, or try it out with the example data provided:  
(if you only have one address, you can use our [Single Address Lookup Tool](#))

Address	City	State	Zipcode	Name	URL	Color	Image	Group
1600 Pennsylvania Avenue NW	Washington	DC	20500	The White House	http://www.white			
350 5th Ave.	New York	NY	10118	Empire State Building				
1737 Cambridge Street	Cambridge	MA	2138	CGA	http://www.gis.harvard.edu			

(example file will load automatically upon selecting a format, click here to [turn off example loading.](#))

**Step #3**

Click "Validate Source" to ensure data format is readable and populate column data:

**Step #4**

Select appropriate columns from the drop-downs below. Use address/city/state for exact map, or city/state or zip can be used for regional map:

Location Fields		Map Fields (not required)	
Address	<input type="text" value="Address"/>	Title	<input type="text" value="Name"/>
City	<input type="text" value="City"/>	Description	<input type="text" value="All Remaining Fields"/>
State / Province or Country (Europe)	<input type="text" value="State"/>	Group By	<input type="text" value="Group"/>
Zip / Postal Code	<input type="text" value="Zipcode"/>	URL	<input type="text" value="URL"/>
		Image URL	<input type="text" value="Image"/>

(Note: URLs and Image URLs should start with "http://" or they will not work)

- Go to "Step #3" and click Validate Source to validate your source data. It is OK to continue if it warns you that you have more than 500 records. But keep in mind that large amount of data will take some time for the website to process, in some cases might crash the system.
- Go to "Step #4" to accept default setting. If you didn't follow the template, you might need to adjust the field names.

**Tip:**

If you want to calculate the straight line distance from many addresses to one address, put that single destination address as the first record in the list. And check "Calculate (straight line) distance" box in "Step #4". The website will return a table with one more field that contains straight line distance.

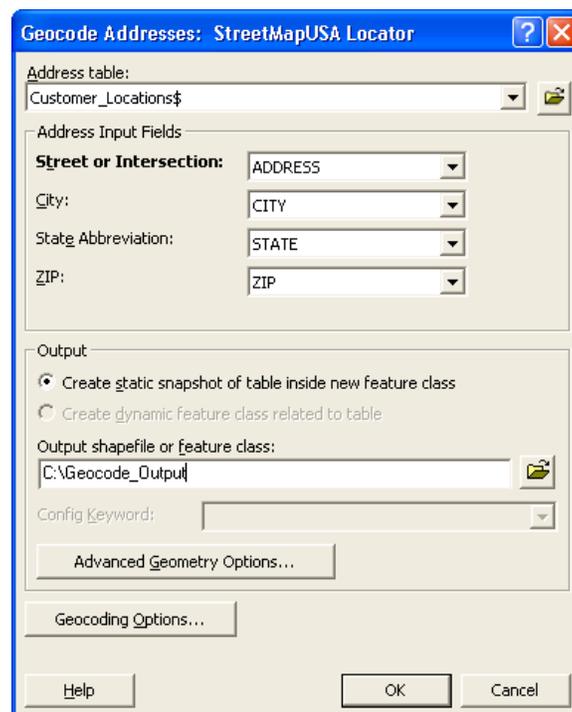
\* The website does not do driving distance calculation.

- In "Step #5", run the model. You will see how many records are geocoded.
- If you want to get the table back with newly added coordinates, in "Step #6", right click in the box and press "Ctrl"+"A" on your keyboard to select all. Then open the Notepad program on your computer, paste it and save it as ".txt" file. You can use this tab delimited text file in other programs, and easily import this data into ArcMap using the AddXY function, which converts it into a GIS layer.
- The website will generate an image with the first 100 geocoded points on Yahoo Map. If you want to show your result on a website, you can select "Save Map to a Web Page" in "More Map

Options”. The website will generate a webpage which has an interactive Yahoo map and the list of your addresses with descriptions and images if you have input such information in the Excel spreadsheet.

### Geocoding address data in ArcMap tutorial

1. Start ArcMap by clicking Start > All Programs > ArcGIS > ArcMap. Click OK to open a new map.
2. Click Tools > Geocoding > Geocode Addresses. In the Add Address Locator window, click the Add button, and browse to the usa/streetmap\_usa/streets folder, choose the StreetMapUSA Locator, and click Add. Note: the usa folder is part of the dataset that comes with ArcGIS. In Harvard’s HMDC computing facility at 1737 Cambridge St. this folder is located on the shared T:/ArcGIS Media Kit 9 for ArcGIS/DM\_StreetMap\_2006(E)/ folder, in the HSPH microlab at 677 Huntington Ave. the folder is located on the G:/Shared/GIS network location. The data is also available for download for Harvard affiliates at <http://gis.harvard.edu> after logging in, click GIS Software and Download Software and scroll down to the ESRI Data & Maps and StreetMap USA topic.
3. With the StreetMapUSA Locator highlighted, click OK. In the Geocode Address window, browse to the table to be geocoded. Valid formats include .xls, .dbf, and comma or tab delimited text files. Make sure the Street or Intersection, City, State, and Zip fields correspond to their respective matches in the input table. Specify an output shapefile, and click OK (see below) The geocoder will run.



A result displaying how many addresses found matches will appear. It is then possible to interactively geocode the unmatched addresses, or change the settings of the address locator to refine the results. The ESRI ArcGIS geocoding tutorial is a more in depth tutorial, and is available for download for Harvard affiliates by going to <http://gis.harvard.edu> logging in, and clicking the Technical Support link, and scrolling to the bottom left.

## Geocoding address data in Google Earth tutorial link

Google Earth is free software that can be downloaded and installed at <http://earth.google.com>. Using the new spreadsheet mapper functionality within Google Earth (GE) allows for geocoding up to 400 addresses, and includes the option to include additional information or images specific to each address for display in GE. Addresses must be in one field in the table, as follows:

ADDRESS
1171 PIEDMONT AVE NE, Atlanta, GA
MAIN ST & 41ST AVE, Atlanta, GA

Google has a wonderful tutorial complete with a video explaining how to use spreadsheet mapper at:

[http://earth.google.com/outreach/tutorial\\_mapper.html](http://earth.google.com/outreach/tutorial_mapper.html)

## How to geocode non-address data

The ability to geocode data is not limited to address data by any means. If your data has a geographic identifier that is not an address, for example a zip code, census tract, state, etc., it is still possible to easily geocode the data. Simply put, if a GIS layer exists with an attribute naming the different features, it can be used for geocoding input data. Batchgeocode and Google Earth Pro can be used if the geographic identifier is a city, zip code, international postcode, or some other element of an address. Try it out, and see what match rate you get using these methods. Most likely creating a custom geocoder using ArcGIS will be necessary to geocode non-address data. As with address data, ensuring that the input data is free of misspellings, special characters (such as “ ? # ‘ \ %), and abbreviations is critical to success. The tutorial below explains how to create a custom geocoder for U.S. census tract data below. **This tutorial can be applied to any GIS layer one wishes to georeference input data to.**

## Creating and using a custom (census tract) geocoder in ArcGIS

Imagine you’ve interviewed people for a research project in different census tracts and want to make a map of this, perhaps displaying the results geographically by score, and then eventually doing analysis with the census variables contained in the GIS layer. Your input table looks like this:

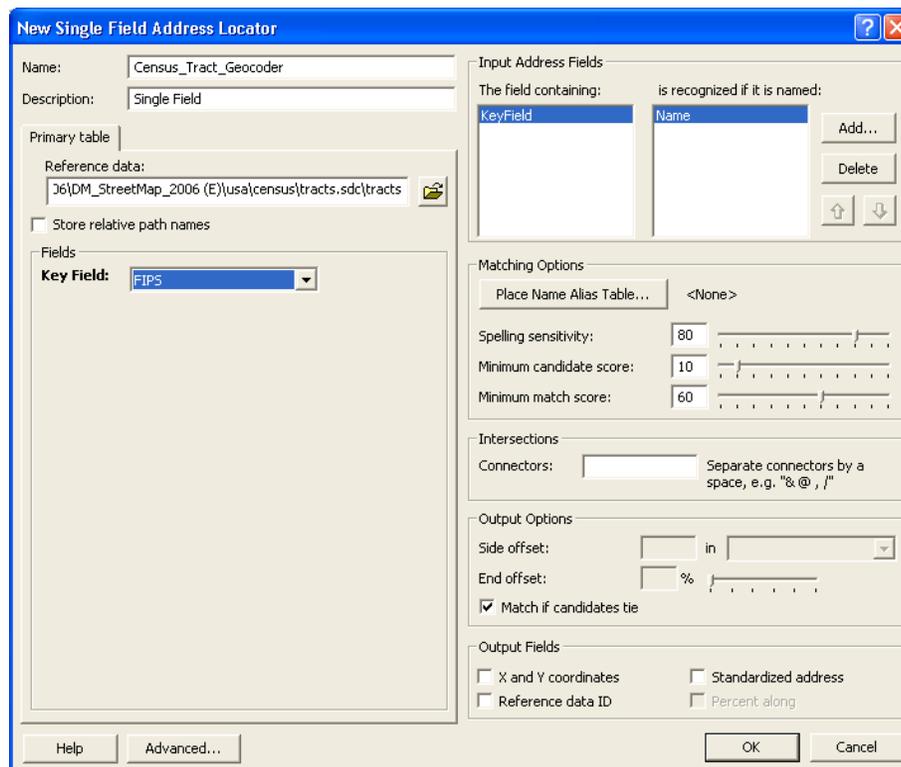
STATE_CNTY_FIPS	SUBJECT_NAME	INTERVIEW_SCORE	GENDER
01001020500	John Johnson	345	M
01003010701	Sarah Brown	432	F

In order to geocode this table, creating a custom census tract geocoder will be necessary.

1. Click Start > All Programs > ArcGIS > ArcCatalog
2. In the ArcCatalog directory tree on the left side of the window, right click C:\ and choose New > Folder. Name the folder Census\_Geocoder.
3. Right click the Census\_Geocoder folder, and choose New > Address Locator. Click Single Field and then OK.

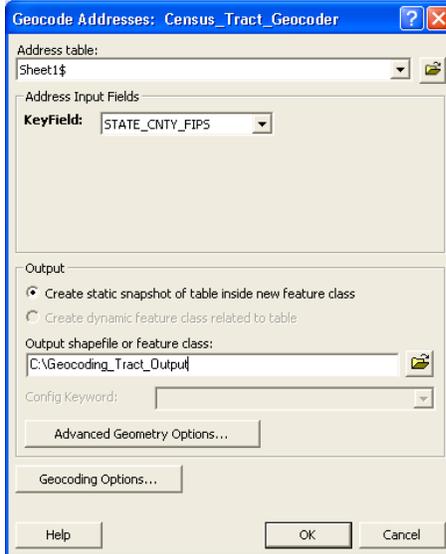


- In the dialog window, name the Address Locator `Census_Tract_Geocoder`. For the Primary Table Reference data, browse to the `/usa/census/tracts.sdc/tracts` GIS layer. For the Key Field, choose `FIPS`. This is a combination of the State, County, and Tract FIPS number which matches your input dataset's `STATE_CNTY_FIPS` field. Click `OK`, and the geocoder will be created. This may take a while, as it has to create a reference index for all 66,175 census tracts.



- Once the Census locator is created, start ArcMap by clicking `Start > All Programs > ArcGIS > ArcMap`. Click `OK` to open a new map.
- Click `Tools > Geocoding > Geocode Addresses`. In the `Add Address Locator` window, click the `Add` button, and browse to the `C:/Census_Geocoder` folder. Choose the

7. Census\_Tract\_Geocoder locator and click Add. With the Census\_Tract\_Geocoder highlighted, click OK.
8. In the Geocode Address window, browse to the input table to be geocoded. Valid formats include .xls, .dbf, and comma or tab delimited text files. Make sure the KeyField corresponds to the STATE\_CNTY\_FIPS field in the input table. Specify an output shapefile, and click OK (see below). The rows in the input table will be geocoded to the centroids of the corresponding census tracts.



For an electronic version of this document, or answers to geocoding or any GIS questions, please contact the CGA at [contact@cga.harvard.edu](mailto:contact@cga.harvard.edu)