

## **Geocoding**

Jason M. Smith

### **Introduction**

According to the ArcGIS 9.1 Help Manual, geocoding is the “process of creating geometric representations for descriptions of locations.” In simpler terms, geocoding provides a spatial reference to data that does not have one. It is also commonly referred to as “address matching.” Geocoding an address allows the address point to be linked to other geospatial databases and the creation of contextual datasets. Point distributions can also be analyzed spatially using any of a variety of geostatistical and exploratory spatial data analysis techniques.

### **What sorts of data might be Geocoded?**

Assuming certain elements of an address field exist it is possible to geocode many different types of data, data on individuals as well as on institutions such as schools, clinics, crime events, day care centers, toxic waste sites, or virtually anything else that can be linked to a particular spot on the earth. Once we have geocoded addresses we can create contextual data sets. For example, we can link individual addresses to demographic data from the US Census, education data from School District Data Book, or health data from Area Resource Files.

The most common way for creating this linkage, at least when using data in the United States, is through address matching to a street file. This process is carried out through a Geographic Information System (GIS), such as ArcGIS 9.1, or ArcView 3.X, both ESRI products.

### **What is needed for Geocoding?**

First, one must have a data file for which they want to provide a spatial reference. This is called the “Address Table.” Second, one needs a “Street File,” or “Reference File,” which contains the streets to which one wants to link the Address Table. StreetCD is a software package through which one can obtain all of the streets in any county in the United States. (Other sources for such reference files exist, though this one is most often used in the GIA Core for this task.) The idea is to match the address listed in the Address Table to a street segment in the Reference File. (One must consider the time period for which each file was created: some addresses may be new, or outdated, and thus not reflected in the reference file.) The process looks for street names in the Reference File that resemble the street name in the Address Table. The process then attempts to locate the street number from the Address Table to a particular street segment in the Reference File. (Each segment has an address range for either side of the street, and the Address Table value will be placed in a segment with the appropriate range.) While this process is conceptually rather straightforward, conducting it in a GIS environment can be complex.

When using ArcView 3.X, the user must “prepare the data for geocoding” through the Theme menu (Theme → Properties, then choose “Geocoding” from the buttons on the left.) Once the appropriate parameters are set, this will “build the geocoding index” ArcView needs to conduct the matching process.

In ArcGIS 9.1, this same task involves the use of the ArcCatalog application to “Create a new Geocoding Service.” This is done through the tree diagram in the left frame of the ArcCatalog

window. Again, the parameters one wishes to use for conducting the Geocoding process must be set, and then saved in the Geocoding Service. This “Service” is then imported into the ArcMap application through the Tools menu. (Tools → Geocoding → Geocode Addresses).

(The setting of the parameters in either case is crucial to the process. The GIA Core conducts a workshop on Geocoding, which explains all of these steps and parameter choices, and leads participants through a hands-on exercise. Materials from this workshop are available from the GIA Core.)

### **Getting through the Geocoding Process**

It is likely that many addresses in your Address Table will not match after the initial (automated) attempt. The table below gives some of the most common reasons why:

<b>Errors in Address Record</b>	<b>Example from Allegheny County Schools (workshop example)</b>
Incomplete Address	Forest Rd -- no number
Error in Street Number	0 Porters Hollow Rd -- need to verify correct number
Error in Street Name	845 McLain St should be 845 Mc Lain St
Error in Street Type	501 Butler Rd should be Ave.
Error in Zone	we did not use a Zone field
<b>Errors in Street Segment Record</b>	
Missing Range	710 Dorseyville Rd -- best segment has no range for right side of street
Error in Range	2401 Rochester Rd -- range starts at 2407 -- need to verify
Error in Street Name	1595 Brodhead Rd -- street file has "Broadhead"
Error in Street Type	1595 Brodhead Rd -- street file has "St"
Error in Zone	we did not use a Zone field
<b>Other Errors</b>	
No Street with Name Given	392 William Penn Rd. -- need to find where this is (if correct)
No Segment with Number Specified	No example, but often a result of Street Name/Highway Route switches
Multiple possibilities	360 School Ln -- need to inspect each possibility

Adapted from Cromley, E.K. 2002. GIS and Public Health. NY: The Guilford Press.

The user must manually go through the unmatched records and match them to appropriate street segments. Sometimes this is an easy process, for example, when a simple misspelling of the street name exists either in the Address Table or the Reference File. Other times, a street may change names in the Reference File; for example, from North Atherton St., to US 322. Multiple options may exist for matching, such as a road with a name used in more than one location within the Reference File (i.e., multiple “Main St.” options.) In all of these situations, the user must determine where the correct location is, and manually select the match from a list of candidates the GIS will provide.

### **References**

The best place for help on Geocoding in ArcGIS 9.1 is the ESRI Manual *Using ArcCatalog*. Chapter 14 is devoted entirely to Geocoding Addresses. Also, see *Getting to Know ArcGIS* for further coverage of the topic. The aforementioned workshop materials, prepared by GIA Core staff, lead the user through an example in this GIS package. For help when using ArcView 3.X, refer to ESRI’s *Introduction to ArcView GIS*. Chapter 7 covers the process in this GIS package. ESRI’s *ArcView GIS Exercise Book Ch 10* guides the user through an example, with additional discussion of the process. All of these materials are available in the GIA Core, 8<sup>th</sup> floor of Oswald Tower. The GIA Core’s website is ([www.pop.psu.edu/gia-core/](http://www.pop.psu.edu/gia-core/))

Another online resource is ESRI’s website ([www.esri.com](http://www.esri.com)); search the website for “geocoding” to get numerous links to additional information.