Geocoding 1.5 Million Registered Voters in Maricopa County, Arizona

Tim Johnson
Geographic Information Systems Manager
Maricopa County Recorder/Elections Department

Abstract

This paper presents the results of a Geographic Information Systems (GIS) project undertaken by the Maricopa County Recorder/Elections Department to accurately geocode all of the registered voters in the county. The purpose of the project was to locate each voter's address with a high degree of certainty and repeatability. The geocoded data is used for population and district analysis within the department and as part of a larger plan to employ GIS technology in real time voter district assignments. All together, there are over 1.5 million voters to geocode. Meeting a goal of a 100% match rate required a variety of data sources, data manipulation, and geocoding techniques.

Background

Geographically large and growing fast, Maricopa County, Arizona had the greatest numerical population increase of any U.S. county between 2000 and 2001. With more than 3.2 million residents, it is the 4th most populous U.S. County and is the 14th largest in area. Phoenix, the state's capitol, is located in Maricopa County, along with Mesa, Glendale, Chandler, Gilbert (the fastest growing city in the US), and 20 other cities and towns. The Maricopa County Recorder/Elections Department administers elections for the 1.5 million registered voters in the County.

Problem Statement

The Maricopa County Recorder/Elections Department began formally considering GIS technology as early as 1994. From the beginning, GIS was seen as a logical solution for managing several election tasks such as district mapping, reprecincting, redistricting, election planning and execution. It was also hoped that the technology could be integrated into the Voter Registration and Administration System, the enterprise system that manages the voter file and elections. In this way, GIS would serve as an engine for determining a voter's precincts and districts based on their address. A survey of the technology and data available at the time, as well as the cost to implement such an engine, revealed that the project was quite ambitious but could theoretically be accomplished. By far, the biggest obstacle was the need to have reliable and accurate geocoding results for the vast number of addresses in the voter file.

The decision was made that GIS in the department would be phased in over several years. The initial effort would be to create maps and build related datasets to handle the mapping needs of the department. Then, specialized procedures and applications were developed to handle planning, redistricting, and other projects. Finally, voter address geocoding would be tackled, and the results would be integrated into the department's systems and processes.

Voter Geocoding

Serious efforts to geocode voters began in late 1999. Now, five years later, the project has achieved a 99.88% match rate to an acceptable GIS reference layer and the remaining 1,800 or so voters' locations can be predicted using an address grid. The resulting data has proven invaluable to the department and is allowing the effort to spatially enable the Voter

Registration System to move forward after a decade-long delay. The methodology used in the project has several key features including: selection of reference data and standards, enhancements to the GIS software, targeted corrections, and incremental benefits. Each of these features will be discussed below.

Selection of Reference Data and Standards

Selection of reference data and standards provided the foundation for the geocoding project. In Maricopa County, there were a wide variety of reference data sources available, including both public and commercial entities. These included at least two countywide street networks maintained by regional governments, three commercially available street networks, two commercial geocoding services, half a dozen citywide street networks, and the Census Bureau's TIGER/Line street network. Additionally, the County Assessor was working to convert its one million plus parcel database to a GIS format that would include, eventually, the situs address.

The department established a goal of meeting, at a minimum, the National Map Data Accuracy Standards of plus or minus 40 feet at a 1:24,000 scale for the voter geocoding as well as its other GIS data. This immediately ruled out TIGER and many of the other data sources which were based on it. Another requirement was that the department needed immediate access to the data for corrections and additions. Maricopa County's explosive growth and the absolute necessity of being able to quickly fix geocoding misses drove this need. Finally, the department wanted enough ownership of the data to be able to share it with other county departments, government entities, and individuals. Evaluations of the data available at the time showed that no street network had overwhelming advantages over any other as far as completeness or match rates were concerned.

The countywide street network maintained by the Maricopa County Department of Transportation (MCDoT) was the best candidate to begin the geocoding process, and their GIS staff was most interested in assisting in the project. However, the initial lack of zip codes in the data and the volume of edits that needed to be done made this dataset a starting point for the Elections Department. As soon as the department began bulk edits of the data it started to evolve into another version of it. Care was taken to allow MCDoT access to our edits and additions so both departments could benefit from the work that was being done.

The first attempts to match the voter file to this network yielded a match rate of about 80%. This was a good start, but tens of thousands of addresses remained unmatched.

Enhancements to the GIS Software

The initial geocoding attempts produced good results, but nowhere near the 100% match rate needed by the project. Since over 200,000 voters were unmatched at this point, it did not seem efficient to jump into the data and begin correcting it on a personal level yet. Instead, the department sought ways to match voters *en masse*. Two methods were tried with great success: modifying the ArcView geocoding class files ("STNAME.CLS" and "US_ADDR.CLS" in our case), and adding additional reference data sources to the geocoding process.

In ArcView 3.x, the software used for the bulk of this project, geocoding class files contain the rules that govern how the geocoding indexes are built and how addresses are parsed into their individual address components. For example, the first line from "STNAME.CLS" reads like this:

BYU BAYOU A

This tells ArcView that whenever it encounters “BYU” in an address, expand it into “BAYOU”. A quick review of the Maricopa County unmatched voters and the street network revealed tens of thousands of addresses that should geocode – their address components match perfectly – but they fail to match. When the class files were examined, it was apparent that certain streets where being parsed incorrectly for our local data. For example, “Avenida” would be standardized to “Ave” as a street type. This would cause all of the addresses on that street to fail to match. The solution was simply to comment the line by adding a semicolon (“ ; ”) to the beginning of the line that controlled “Avenida”:

```
;AVENIDA          AVE          T
```

Then, rebuild the geocoding indexes and try again. The addresses would match. Thousands of Maricopa County’s addresses were matched by altering these files.

After the geocoding indexes are modified, the department began to look for ways to incorporate other geocoding reference layers into the process to tap their various strengths. In particular, the Assessor’s parcel situs address database was rapidly improving and was another excellent candidate for geocoding voters. Other sources such as data from local jurisdictions and TIGER could be tapped to provide enhancements to the match rate and as a spatial reference to make corrections to the street data. To this end, the department developed a “fall back” geocoding process. In this process, matchable data layers were arranged in a list in the order of preference to provide matches. High priority layers such as the parcels and streets were on top, lower priority layers such as other street networks of TIGER streets were towards the bottom. An Avenue script was written that would sift addresses through this list. A match would be attempted at the top of the list first, and the script would continue to try to find a match in subsequent layers until it did. The source of the match was written to the address table to provide tracking and analysis of success rates (see Figure 1). By rearranging, adding and removing layers the department was able to

X	Y	Ay_mtheme	Ay_status	Ay_score	Address
651261.7407	894611.9461	Streets	M	100	1010 N 2ND AVE
651261.7407	894611.9461	Streets	M	100	1010 N 2ND AVE
651261.7407	894611.9461	Streets	M	100	1010 N 2ND AVE
651330.6720	894583.8300	Address XY	M	100	1011 N 2ND AVE
651330.6720	894583.8300	Address XY	M	100	1011 N 2ND AVE
651312.8004	894650.5438	Streets	M	100	1013 N 2ND AVE
651330.7580	894603.3255	Address XY	M	100	1015 N 2ND AVE
651330.7580	894603.3255	Address XY	M	100	1015 N 2ND AVE

Figure 1: example results table showing match source

test and compare match sources. The results from one layer were also used to improve the quality of other layers. A final process was developed to predict a voter’s location based on a countywide address grid and the alignment of his or her address. This ensured a point could be placed on the map for each voter. Continued editing is aimed at replacing this predicted point with the voter’s true location.

Targeted Corrections

Batch corrections and processing techniques brought the match rate above 90%. The department knew that at some point a large scale editing effort would be required to advance the project. Targeted corrections were developed in order to increase matches for as many voters as possible very quickly.

This process involved summarizing unmatched addresses so that concentrations of voters could be fixed with a few edits rather than fixing each individual voter’s address. During the early editing stages, it was possible to match several hundred voters at a time by fixing

addresses for apartment buildings and other multiple-voter residences. As the editing effort matured, smaller and smaller groups of voters were encountered at individual addresses. Then, the unmatched addresses were summarized and edited by street and by "from – to" address range. As those numbers dwindled, other summarizing strategies were used: by zip code, by voting precinct, and others. The time did come where only few voters were at a given address. However, by this time the match rates had exceeded 98%. The final unmatched 2% of the voter addresses reflected a high degree of effort for the return because the addresses that were left are the most difficult to find and have just a few voters residing on them.

Incremental Benefits

As mentioned earlier, the ultimate goal of this geocoding effort was to achieve a 100% match rate for voter geocoding so that GIS technology could be used to assign precincts and districts to a voter based on his or her address. Fortunately, the success of this project is not measured by just reaching that goal. The department needs 100% accuracy to realize its ultimate plans for the data, but there have been numerous benefits derived from the project as it matured.

First, the data has been used to derive voter counts for new voting precincts or districts. Prior to geocoded voters, the department had to convert a contemplated precinct or district into address ranges and alignments that could be loaded into the System and processed in a nightly batch. The next day counts would be available. Now, any precinct or district can be drawn in the GIS software and voter counts aggregated on the fly. This basic improvement in election administration was first used for reprecincting in 2002.

Next, voter addresses can be visualized and queried through ESRI's ArcIMS on the department's intranet mapping application. Any department employee can access and view voter locations on an interactive map as well as select and query them. Each address also hyperlinks through the map to the voter lookup screens on the intranet. This provides a high level of detail for any address on the map.

Finally, the geocoded voter addresses have been extensively used to check and verify new and existing boundaries in the Voter Registration System. The current System relies on translations of districts into address ranges. These ranges have proven to be difficult to manage and update, and occasional errors have sometimes required costly and undesirable efforts to correct. The department has implemented an auditing procedure in which district codes from the geocoded data are compared to district codes within the system. Any discrepancies are reported and investigated. This provides a critical cross check especially when sweeping changes such as reprecincting are introduced into the system. Another benefit is that the procedure identifies errors in the GIS data as well as "proves" the data when no problems are identified.

Conclusion

Ten years ago, the Maricopa County Recorder / Elections Department identified a need for embedding GIS technology into its Voter Registration System. Geocoded voter data was a vital part of fully realizing the promise of GIS technology. Thanks to creative data gathering and use, focused but extensive editing, and a firm resolve to complete the project, the department has created a solid foundation to build GIS into its vital applications.