Statewide Service Geocoding Requirements

Date: December 22, 2010   Version: 1.7

Prepared By: Mike Dolbow and Kent Treichel

The state needs a service which can translate a street address or street intersection into an (X, Y) coordinate. A Statewide Geocoding Service (SGS) would meet that demand under the following requirements. See other documents for more information defining the need and potential benefits.

GENERAL REQUIREMENTS

Full Coverage for the State of Minnesota

The service will return an (X, Y) coordinate for a street address or street intersection that is generally understood to be within the boundaries of the state of Minnesota. Ancillary data for non-Minnesota areas can be included to support the service if it is freely available, meaning some non-Minnesota requests may produce a valid return. However, requirements for supporting non-Minnesota addresses will not be considered within the scope of the current project.

Support for Addresses, Intersections, and Mile Markers

The service should be able to return a coordinate for both addresses and intersections of streets, with the expectation that intersection matches are inherently complex, will typically have multiple results, and as a result are generally less precise than address matches. Support for matching mile markers along trunk highways will only be included if request data and support data meet the impending Federal Address Data Standard.

Precision of (X, Y) Coordinates

The service will return, at the highest level of precision available, a point corresponding to the rooftop/parcel location of the address. Less precise searches, such as those utilizing only city, town or ZIP, will yield less precise points. A recommended "zoom level" will be returned for each matched record which appropriately reflects the request.

Address Validation and/or Standardization

Users of the service will be presented with information about the impending Federal Address Data Standard and the United States Postal Service (USPS) address standard and their benefits for address formatting. Addresses that meet one or both of these standards will produce better geocoding results from the service than addresses that do not meet a standard.

The service will include an option for the user to pre-process their inputs via USPS Delivery Point Validation (DPV). If included in this step, resolved addresses will return the

appropriate DPV code and Locatable Address Conversion System (LACS) code. Addresses listed as "not deliverable" will be standardized to USPS specifications.

If MnGeo becomes the host a non-proprietary service that meets these requirements, they should consider a variety of funding methods for the approximately $4,000 annual subscription to the USPS web API to run all elements of CASS. Potential options include annual maintenance fees to state agencies and per-record charges for batch runs.

Batch Processing

The service will deliver (X, Y) coordinates for address data submitted in a batch process. There is no hard limit on the number of addresses which can be submitted within one batch. However, reasonable limitations may be placed on batch processing based on demand for the service. One limitation option to include would be for the host to accept requests for large batch runs and then scheduling them for non-peak time periods.

Online/Interactive Processing

The service will deliver (X, Y) coordinates for address data submitted in an online/interactive process. State agencies and other public or non-profit organizations will be able to integrate the geocoding service into their workflow and applications. High availability will be necessary to meet this need.

No Internal Use or Product Use Restrictions

State agencies and public organizations must be able to employ the service on internal software, including internal desktop and web mapping applications. While the service may be supported by restricted use or proprietary data, there shall be no use restrictions on the data products returned from the geocoder.

No Cost / Low Cost for Use

The host agency should make all attempts to keep the geocoding service at no cost to the user or very low, competitive costs, and explore a variety of funding mechanisms to keep it that way. Cost structures should favor convenience over revenue generation in order to facilitate ease of use and fund transfers.

No Storage of Request Data

The host provider will not store information about requests from clients, including address requests, unique IDs passed inside batch requests, or any other information stored alongside requests. The host may, however, aggregate information regarding total numbers of requests, match rates, and other statistics that can be useful in determining the efficiency and effectiveness of the service.

SPECIFIC REQUIREMENTS

Cascading through Support Databases

The service should have the capability to "cascade" among support databases, in order to find the best available match from multiple sources. For example, if a match is not found within a support database of address points, the service should cascade to other support databases as available, such as parcel polygons, addressable street centerlines, city centroids and zip code boundaries. The order and conditions of the cascading should be configurable within the service.

Data Returned

The service will return data in a format which reflects industry standards, such as CSV, XML or JSON. Returned data will include:

- initial address string
- initial unique ID (if sent in a batch request)
- parsed corrected address components (street address, postal city, state, and ZIP code)
- DPV match code and LACS code (if the user opted for this preprocessing)
- the coordinates expressed as lat/long **and** UTM Zone 15N coordinates
- match quality information (see below)

Returning "point-in-polygon" or "proximity finder" information about the coordinate is out of scope for the geocoding service, since a prototype service is under development by MetroGIS and SharedGeo.

Match Quality Information

The service will return match quality information which provides the details of the results of the geocoding operation. This will include:

- the support database in which the address was found
- the portions of the address which were corrected/cleaned
- the precision of the returned geographic coordinates (a ranking of 1 to 5 is suggested, with 1 being the highest quality match and 5 being the lowest)
- the suggested "zoom level" of the first match, with "6" representing a full-state view, and "22" representing a rooftop-level view.

Customizable Address Dictionary

The service will include the ability to incorporate a customized address dictionary. This will allow for geocoding of commonly understood addresses which do not match USPS standards, parcel data or street centerline conventions. This could include locations such as the Mall of America (MOA), or buildings with street addresses which do not receive mail. This feature will reduce the amount of pre-processing required for problematic addresses and provide a mechanism for matching and precisely locating these addresses..

Customizable Application Calls

The service will allow for customizable calls from a variety of desktop, database, and web clients, including but not limited to; ArcGIS Desktop, ArcGIS Server, MapInfo Desktop, MapServer, and GeoMoose.

Response Time

The service must have an extremely fast response time for single, interactive requests originating from web applications, and a reasonably fast response time for batch runs numbering in the hundreds. For larger batch runs, the host agency should have the ability to schedule the request and return results within days.

Software Sharing

All parts of the geocoding service *software* should be freely available/sharable, including comments in the code, and documentation for anyone to install and use. This requirement does not apply to the support data used at the host agency, which may be either licensed or sharable.