

# MetroGIS Geocoder Project

## Outline for Coordinating Committee Review 6/27/2007

Project Participants: Dave Bitner (MAC), Nancy Read (MMCD), Mark Kotz (Met.Co.), Jim Maxwell (TLG), Gordy Chinander (MESB), Chris Cialek & Jim Dickerson (LMIC), Bob Basquez (St. Paul), Kent Treichel (MN Dept. of Revenue).

Focus of project:

1. Develop geocoding software that meets the following requirements:
  - Parse: take a given "initial address" character string and transform that into something that can be used to search against a database
  - Geocoding Engine: search a database (streets, parcels, or some other locational db) and return a list of lat/lon coordinates (point) of possible matches, and estimate of quality of match
  - Cascade: if Engine can't find a match in primary dataset, search next, etc. Priority and number of datasets searched should be configurable. Data returned on quality of match should indicate which dataset used for match.
  - Database "template" needs to match Geocoding Engine toolset; original data could be shapefile or PostGRE/GIS or some other data format.
2. Set up the above software on a host site with associated data and any supporting software such that geocoding can be provided as a web service for individual requests from other web applications.

Scope and Design issues:

1. Start with single requests, not batch.
  - a. Software could be used in-house by participants to do in-house batch geocoding against datasets they are already licensed to have.
  - b. a batch geocoding service (free OR charge) could be set up by a participant, depending on licensing issues.
2. Final product is web service that returns initial address string, parsed corrected address(es), lat/lon coordinates, and match quality info.
  - a. It is up to the developers of the web sites consuming this service to handle translation from lat/lon to other coordinate systems (including custom systems like King Map Book or systems like Military Grid), to handle match options and match quality display. If there are sufficient resources, code samples for doing these chores could be included, or may consider adding the most common conversion (UTM) to service.
  - b. Returned data format should reflect industry standards for geocoding services (e.g., standard schemas for XML transfer).
  - c. setting up a mapping site directly usable by the public is not within scope of this project.
3. The corrected addresses (text) returned could meet some national standard... [?]
4. Geocoder engine could use any dataset with US-style address. As part of project we plan to make data templates more specific to locally-available data: TLG streets, Metro Parcels, and eventually Occupiable Units. We plan to launch the web service using TLG streets and Metro Parcels.
5. Prefer that all parts of software are freely available/sharable, include comments in code, and documentation for anyone to install and use.
6. The complete process of submitting an initial address string, parsing, running geocoder engine, and returning list of matches should have a fast response time.
7. Software design should recognize potential future needs for enhancements, including intersection look-up and reverse geocoding (lat/lon to address).

**Total \$ Amount requested:** Not to exceed **\$14,000.**

Activity	
1. <b>define functional requirements</b> of a geocoding service for the MetroGIS community, scope of current project and develop RFP's	- to be done by team
2. <b>develop parsing code and geocoder engine</b> - evaluate existing geocoding code offered by MAC or available from other sources, assess changes needed to meet MetroGIS community needs, and use funding for programming to make those changes and/or develop new code as needed.	- RFP #1a - \$10,000 We expect to hire a consulting firm that can coordinate the evaluation of existing resources, with review by the group, and can perform or subcontract programming, possibly including code contributions from group members.
3. develop <b>documentation</b> for those planning to build applications that use the service or those wishing to use the geocoder code, either in open-source or ArcIMS environments	- RFP #1b - \$1000 (expect to be done with 1a)
4. define draft roles and responsibilities of “regional custodian” of service (the host organization) as well as source data providers (e.g. parcels & TLG)	- to be done by team and prospective host(s), as details of needs become clearer
5. find an organization willing to <b>host</b> the service and set up service on their server	- LMIC has offered to host. Probably no charge; will need to know what assumptions are made about host environment. Could also do as RFP #2, in which case would need another ca. \$1000. May also consider a multi-node setup, especially since some organizations may want to attach their own data to the address points for querying. This could also providing a means to load-balance.
6. <b>maintenance procedures</b> for TLG street data and other data used, such as translating to template form, rebuilding indexes, conforming to standards (Av vs Ave etc).	- Possibly RFP #3 - \$1000? Will need to determine with host and data providers. Some existing code from City of Saint Paul might be used.
7. add street <b>intersection</b> look-up 8. add <b>landmark</b> look-up	- add-on to RFP #1 - \$1000 Could start with existing intersection code for TLG dataset from City of St. Paul. Note that if code base is relatively generic, would make the end product much more valuable overall. Landmark lookup is one type of datasource, but there are many others. Not much work to increase the return on investment.

MMCD has agreed to serve as administrator as needed for handling funding.