517 – Georeferencing Catalog Data – Status Report
September 4, 2013

The following summarizes our accomplishments since May 2013;

Our goal was to geocode more HOLLIS data in order to provide a better representative sample of HOLLIS data in our user interface (GeoHOLLIS Beta). To that end we have extracted approximately 1.1 million HOLLIS records from the Library Cloud data set (we are running a local copy of LC). We put these records through our geotagging process. The result was a database of 5.8 million geotagged records. Otherwise stated, the software came up with 5.8 million place references (records) from the 1.1 million HOLLIS records. The next step is to load these data into a Solr index and have the GeoHOLLIS Beta interface point at it. This should show the functionality and its potential more clearly than the first set of ~150,000 HOLLIS records.

The GeoHOLLIS Beta user interface is now integrated with the GPU technology mentioned in the last progress report. This technology allows us to render large amounts of data extremely fast and overlay it with our maps. Users can now create on-the-fly heat maps using millions of records as well as selectively map/render millions of points. **However, since the GPU is not running on a Harvard server it is not always available**. Before having the GPU developer ingest the new data set into his GPU we hope to make sure that there's better reliability and up time of our data. GeoHOLLIS is still functional without the GPU mapping component but is more limited in what it can present to the user. Also, when the GPU is down the map interface is not visually pleasing as there are 2 overlays that appear pink and the user must turn them off.

We tested another open source geotagger called CLAVIN. It has excellent potential to eventually replace the proprietary software we are currently using from MetaCarta. However, results from MetaCarta are still more accurate and complete. Simply stated, CLAVIN is not ready to replace MetaCarta at this time.

In the upcoming weeks we will ingest the new data set into Solr and the GPU. Then, make it available in the GeoHOLLIS Beta web site. From there we hope to talk to the Scaling and Interoperability Committee and explore hanging GeoHOLLIS off the Library Portal in order to get more feedback and accurately gauge the system's potential. We would also like to make the geocoded HOLLIS records available for download by others. Also, we'll be exploring a method to push the geotagged records back to Library Cloud.