

Georeferencing Catalog Data – Status Report May, 2013

Activity was slow since the last report. However, during the last weeks of May we solved numerous technical issues that will allow us to work more steadily.

We obtained a copy of the Library Cloud data set, imported it to our own Solr core and have been testing our geotagging process with it. Our copy has a different output format than the original. That meant modifying our automation scripts. The developer at Berkman changed his code to accommodate the new format as well. We've now automated most of the geotagging process and started batch work last week. We processed an initial set of 50,000 HOLLIS records and are evaluating the results for correctness. Using the Library Cloud data set as a base, we have come up with a new data schema that can be used by others interested in creating similar systems. The new schema also holds additional data that supports server-side point data clustering and other geospatial mapping methods.

We teamed-up and have an informal agreement with a developer from MIT to use our server-side clustering and geocoded data points in a system he created that's capable of drawing vast amounts of data extremely fast (using GPU technology). The system also creates on-the-fly heat maps, which is a very useful tool for data visualization. By the next progress report we hope to have these tools working (creating heat maps) in the GeoHOLLIS Beta application for demonstrative purposes.