

ABSTRACT

ACCURACY AND COMPLETENESS AS MEASURES OF THE QUALITY OF VOLUNTEERED POINT-FEATURE GEOSPATIAL DATA AND EVALUATION OF THE EFFECT OF DEMOGRAPHICS ON THAT QUALITY

William F. Mullen, Ph.D.

George Mason University, 2014

Dissertation Director: Dr. Peggy Agouris

Over the last several years there has been a tremendous growth in the generation of Volunteered Geographic Information (VGI) data. The assessment of the quality and accuracy of VGI contributions, and by extension the ultimate utility of VGI data has fostered much debate within the geographic community. The limited research on quality of VGI has predominately focused on linear feature elements, such as road segments or points resulting from intersections of such linear segments, has shown that error within the data is heterogeneously distributed. Additionally, several studies have suggested that the quality of VGI is impacted by demographic information such as population density or the socio-economic characteristics of an area.

Researchers have suggested that the quality of VGI is impacted by demographic characteristics, such as population density or socio-economic status; however, these have mainly been qualitative in nature. Research on crowd-sourced initiatives has shown that data produced by numerous contributors will result in a more accurate product than an

individual, although research on VGI is more infrequent. In addition researchers have begun to assess the impact of contributor motivation and bias that has been noted to significantly impact the quality of VGI data. Currently, it remains an open research question on how local demographic factors impact the quality of such datasets.

The focus this research is to investigate VGI data accuracy and completeness as representation of data quality and then to quantify VGI error in terms of demographic properties in an effort to quantify a potential causal relationship of demographics on the quality of the contributed data. This research specifically focused on the demographic characteristics of the crowdsourced area and its relation to the accuracy and completeness of the corresponding VGI contributions with respect to OSM.

This research effort developed a method for quantifying the completeness and accuracy of a select subset of infrastructure-associated point feature datasets of volunteered geographic data in two separate geographic areas: a portion of Metropolitan Denver, Colorado, and Fairfax County, Virginia using a national geospatial dataset as the reference benchmark with a national datasets from volunteers used as a comparative test dataset. In addition, in the Denver Study Area, a second, hybridized volunteered dataset, which had undergone a focused quality control process was also evaluated for accuracy and completeness in comparison with the OSM test data.

This research used standard count and comparative assessments of the accuracy and completeness and statistical and spatial analysis techniques to identify potential correlations among such data and the distribution of errors in the form of accuracy and completeness found in VGI data. The findings within this study area do not provide

statistically significant support for the earlier arguments that local demographic properties may affect the spatial accuracy or completeness of VGI. Significantly the results of this research illustrate the benefits of including quality control in the collection process for volunteered data. The research adds to the understanding of how demographics impact on the quality of VGI data and lays the foundation to further work.