

Title: Cartographic Database Production and Maintenance Using Open-Source Geospatial Data

Author: Barry Bitters, Ph.D., GISP

Affiliation: Research Scientist & Technical Fellow, Leidos, Inc. (formerly SAIC)

Contact Info: 6731 Avenida de Galvez, Navarre, FL 32566; phone [850] 684-3052

Abstract: Cartographic database production and maintenance is expensive and time consuming. Open-source and voluntarily generated (crowd-sourced) data is a relatively un-tapped Web resource containing cultural feature content that could be used in cartographic production efforts. Often, open-source and crowd-sourced data are more current and sometimes more “authoritative” than more traditional legacy feature data sources. However, these publicly available resources are produced by a variety of methods and techniques and are made available in varying degrees of currency, reliability, accuracy and precision. Even “authoritative” sources – data provided by open-government initiatives – will often provide geospatial datasets which, for various reasons are often incomplete, outdated and contain erroneous data. Numerous forms of “mash-up” and Extraction/Transformation/Loading (ETL) software are currently employed to reuse publicly available data. Many times, these techniques merely propagate legacy omissions, duplications and errors from their parent datasets. This paper reports on the development of repeatable procedures and processes to repurposing publicly-available geospatial data into verified and validated inventories of geospatial objects. This approach performs a modified form of logical union of open-source data to arrive at a more comprehensive inventory of cultural landmark features. The primary dilemmas addressed are: how best to locate, capture, merge and finally validate and verify open-source and crowd-sourced geospatial data with an emphasis on insuring the integrity, precision and accuracy of the resulting information. This approach has been applied to a variety of cultural feature types – both features that are easily recognizable on aerial imagery (airfields and golf courses) and features that are difficult to discern on imagery (fixed maritime lights, post offices, police stations, hospitals and medical clinics). During this research, over 5,000 cultural objects have been processed to test these validation and verification procedures. With this approach significant increases in feature counts have been realized and more accurate geo-positions have resulted.

Keywords: geospatial data, open-source, crowd-sourced, cartographic production, spatial data mining