Combining Varied Federal Data Sources for Multiscale Map Labeling of Populated Places and Airports for *The National Map* of the United States

Cynthia A. Brewer <cbrewer@psu.edu> James E. Thatcher <jethatcher@gmail.com>, Stephen J. Butzler <sjb204@yahoo.com>; Department of Geography, Pennsylvania State University, 302 Walker Building, University Park, PA 16802

Abstract:
Current U.S. federal mapping standards call for the use of the Geographic Names Information System (GNIS) point layer for placement of all United States populated place labels. This point layer contains both limited and duplicate classification information and therefore complicates the balanced labeling of place with congestion at smaller scales. In turn, this affects map quality for database-driven, multiscale, reference mapping, such as maps served on *The National Map Viewer* from the U.S. Geological Survey (USGS; a U.S. national mapping agency (NMA)). Using heterogeneous and multiple data sources, our research investigates alternatives for the automatic generalization and creation of label hierarchy within NMA reference maps in order to accurately reflect the relative importance of point features such as populated places and airports.

We examine several sets of populated place polygon categorizations from multiple data sources in order to present specific solutions to this problem, each with specific benefits and problems. All of the data selected was constrained to consistent nation-wide compilations in order to serve national mapping requirements. In addition to GNIS point locations and names, we augment hierarchy and category decisions with U.S. Census Bureau data for places (Figure 1a) and with Federal Aviation Administration (FAA) listings of airport characteristics (number of flights per airport) for our case study feature types.

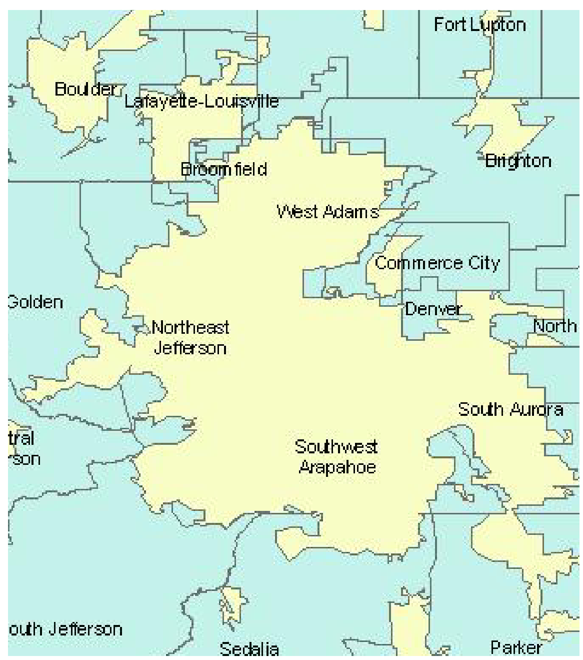
We are able to create an automated procedure for designating GNIS points as within urban or rural contexts using the urban boundary definition polygon coupled with incorporated place and census designated place (CDP) polygons, all from the Census Bureau (Figure 1b). This definition automatically propagates label rule sets through which rural and urban contexts may generate differing hierarchies. However, this definitional schema does not provide enough attributes for hierarchy within rural and urban settings.

We also examine a subset of U.S. Census Bureau data created by the Local Employment Dynamics (LED) program of the Economic Census (economic data collected every five years by the Census Bureau). Local Employment Dynamics data are compiled at the census block level and contain the number of jobs, age of workers, and industry types for both residents and commuters within a given census-defined block. When aggregated to the incorporated place and census designated place polygon level, this information may be used to create an automatically generated hierarchy of economic importance of place. This provided more complete hierarchies than the Survey of Business Owners (SBO) data which are limited to only places that meet thresholds for designation as Economic Places (residential population of 5000 or 5000 employees), as shown in Figure 2.

The overall goal of our work is to couple population and employment information with GNIS names to create a nested hierarchy within the Urban/Rural distinction and through scale. The results are presented using ScaleMaster tools to organize the changes in label and symbol choices and elimination decisions through a wide and continuous range of map scales (1:20,000 to 1:1,000,000). This sounds simple perhaps, but the particular names and polygons vary among all of these federal programs, and the same names appear in multiple categories, complicating the processing goal of generating data for a clear and readable map. Work we presented at the AutoCarto 2010 conference on this theme examined Pennsylvania and Colorado with economic and decennial census attributes and a simpler model that only used decennial census attributes. This work extends placename work to the entire United States and compares multiple attempts at database-driven, automatically propagated, multiscale placename and airport hierarchies from GIS software.

We are proposing a demonstration format for our participation in the Workshop because the realities of meeting our NMA goal (USGS is funding this project) with multiple federal sources that are not typically used together will be useful to other NMA participants (and will be quite dull if it degenerates to enumerating the attributes of each dataset in a talk). The realities of this effort (Figure 3) will perhaps also be enlightening to researchers who do not usually work with a messy overabundance of federal data through scale.

(1a)



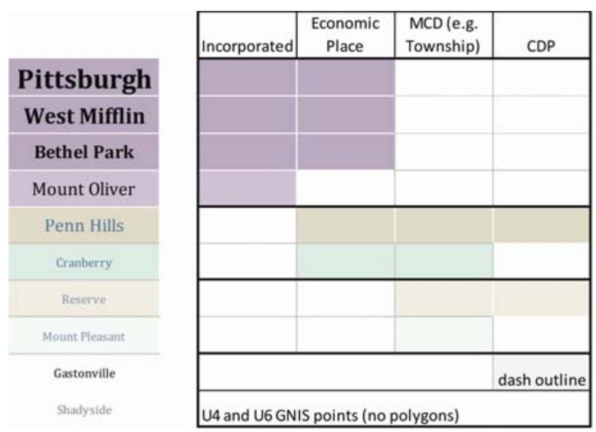
(1b)

Figure 1: Denver, Colorado USA, area showing labels based on economic place and incorporated place combinations (1a) compared to the Census Bureau defined urban boundaries (1b) (these are not administrative boundaries). See Figure 2d for a key to 1b colors; the dark purple and darker beige are the economic places for example. A generalized urban area is used to set different rules for placenames and other feature types (such as churches and airports, not shown here) for better mapping results in both urban and rural places through scale.

(2a)

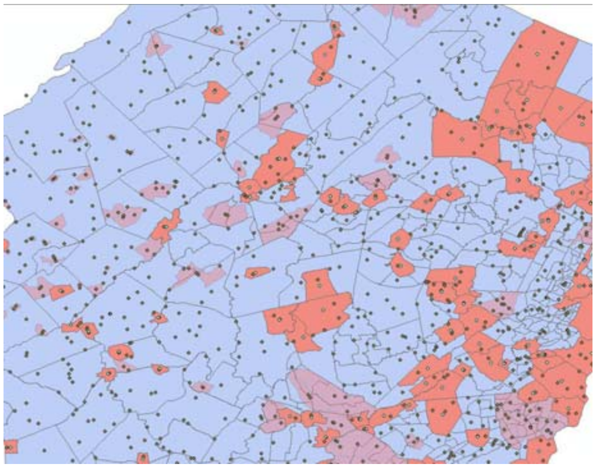
(2b)

(2c)

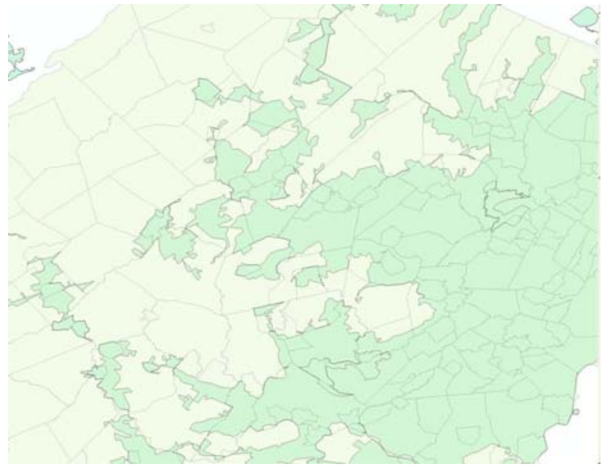


(2d)

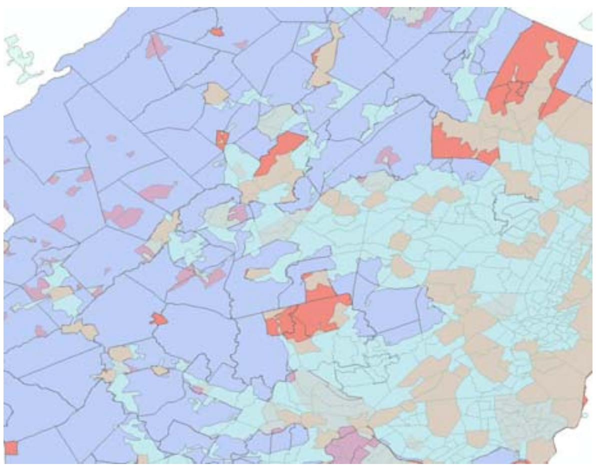
Figure 2: Pittsburgh, Pennsylvania USA, labeling shown at three scales 1:144K (2a), 1:289K (2b), and 1:588K (2c) (figures are captured on screen and reduced from the stated scale for this abstract page). A guide to the hierarchy of sizes and categories is shown in 2d, including incorporated place, economic place, minor civil division, and census designated place partitions. Notice that most places fall in multiple categories. Colors are used on these maps to show categories of place attributes but are not proposed for use on *The National Map.* Likewise, label styles are exaggerated in size and difference to clarify the use of the place attributes but are not intended for final NMA map appearance.



(3a)



(3b)



(3c)

Figure 3: A portion of raw data for New Jersey (USA) showing census subcounty units overlayed with GNIS points that carry required labels in attribute tables in 3a (not all points have associated areas with census attributes). Figure 3b shows the urban boundary for this area, and 3c shows the subcounty units overlayed by the urban boundary to show the variety of boundary relationships among these data.

___

The proposed demonstration fits most closely with the "Generalisation in Support of Data Integration" theme for the Workshop (particularly schema alignment) because it deals with multiple sources that augment each other to accomplish multiscale mapping for the 1:20K to 1:1M scale range.