ESTABLISHING CLASSIFICATION AND HIERARCHY IN POPULATED PLACE LABELING FOR MULTISCALE MAPPING FOR *THE NATIONAL MAP*

W. J. Stroh, S. J. Butzler and C. A. Brewer

Department of Geography, The Pennsylvania State University, University Park, PA 16802, USA -
(wjstroh, sjb204, cbrewer)@psu.edu

KEY WORDS: *The National Map*, place names, GNIS, dynamic labeling, multiscale mapping

ABSTRACT:

Current federal mapping standards call for use of the Geographic Names Information System (GNIS) point layer for placement of United States populated place labels. However, this point layer contains limited classification information and no hierarchy, resulting in problems of map quality for database-driven, multiscale, reference mapping, such as maps served by *The National Map* Palanterra viewer from USGS. Database-driven mapping often relies simply on what labels fit best in the map frame. Our research investigates alternative sources for labeling populated places, including polygons defined by the U.S. Census Bureau, such as incorporated place, census designated place (CDP), and economic place. Within each of these polygon layers we investigate relevant attributes from the decennial and economic censuses, such as population for incorporated places and CDPs, and the number of firms, number of employees, and total revenue for economic places. The data selected are available for the entire country to serve national mapping requirements. This combination of data allows a more refined classification of populated places on maps that better represents relative importance. Visual importance on maps through scale should derive from more than simply residential population, but also economic importance. We differentiate a fourth category of GNIS populated place points, essentially "neighborhoods" and related features—which are not incorporated places, CDPs, nor economic places. Populated places in this fourth class do not have federally defined boundaries, necessitating an alternative method for determining hierarchy in label presentation in through scale.

1. INTRODUCTION

As digital maps continue to replace paper maps in settings ranging from driving directions downloaded to mobile devices to *The National Map*, cartographers continue to grapple with the production of these digital maps in a multi-scale, multi-representation environment. Database-driven cartography has created new pressures for rendering a meaningful map product. A cartographer can no longer simply consider a limited geographic area at a fixed scale when weighing which elements are essential to the map. Labeling map features is an especially burdensome problem, even with advanced label placement tools, such as Maplex (ArcGIS 9.3.1).

Labeling populated places provides an excellent case for considering the perceived importance of a given locale based on its placement and labeling in given geographic extents and at given scales. In its current form, *The National Map* Palanterra Viewer renders a somewhat complicated points layer when representing populated places (Figure 1). Our work considers the current state of labeling of populated place in *The National Map*, whether point-feature labeling of places is a suitable method for mapmaking, and whether richer attribute sets add value to labeling populated places.

2. BACKGROUND

2.1 Source for Geographic Names

The Geographic Names Information System (GNIS) is the official repository for domestic geographic names and is maintained by the U.S. Geological Survey (USGS) to support the efforts of the Board on Geographic Names (BGN) Domestic Names Committee (DNC), the body tasked with maintaining domestic geographic names. The DNC rarely initiates name changes or corrections (except in the case of derogatory names). State naming agencies/committees, local governments and constituencies as well as relevant Federal agencies are consulted as appropriate. Current Federal Geospatial Data Committee (FGDC) standards name GNIS as the official source for the naming of geographic features in terms of Federal maps. Categories of feature names within GNIS include natural features, populated places, civil divisions, areas and regions, and cultural features. The focus of our work is on the populated place names listed in GNIS.

Names contained within GNIS originate from multiple sources. Many were derived from the original USGS 1:24000 topographic map series, though names are also submitted by Federal government partners and by local constituents through a formal submission process. All geographic name records contain name and coordinate location attributes. Some also contain a variant or multiple variants which may represent additional spellings or historical variations. Some records contain a feature designation such as State Capitol or County Seat.

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
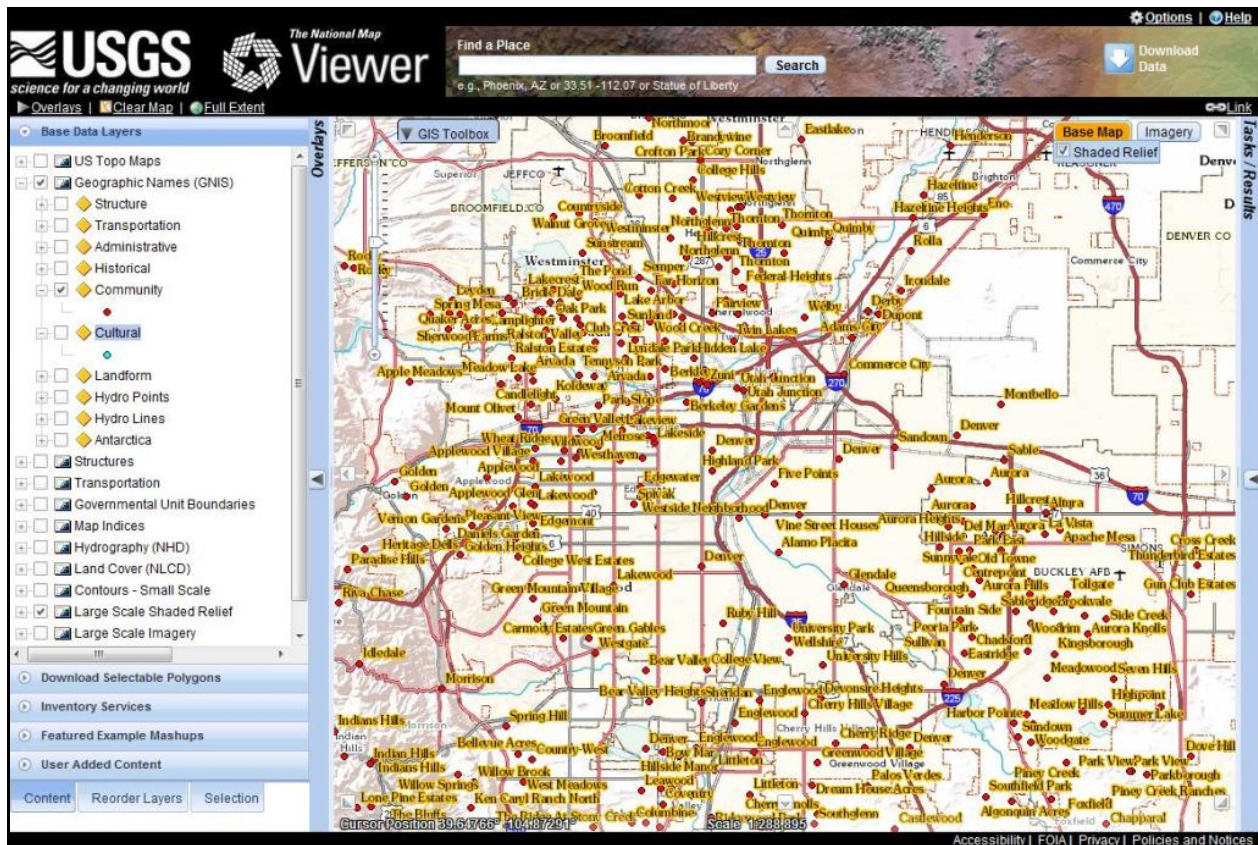November 15-19, 2010 Orlando, Florida

Figure 1. GNIS points and labels shown for populated places in a portion of Denver, Colorado, as seen in *The National Map* Palanterra Viewer. The place layer is red points with black/golden haloed labels laid over a basemap with additional place labels. (USGS, 2010a, http://viewer.nationalmap.gov/viewer/)

2.2 Defining Populated Place

In its broadest definition, the term populated place encompasses the villages, towns and cities that comprise the settlement landscape in the United States. However, a multitude of other types of organized settlement comprises this broad category including townships and boroughs. The

Within GNIS there is a Feature Class called "Populated Place". GNIS also provides distinctive classes for "Civil" and "Census" places. The USGS (2010a) defines these three feature classes as quoted below:

Census: A statistical area delineated locally specifically for the tabulation of Census Bureau data (census designated place, census county division, unorganized territory, various types of American Indian/Alaska Native statistical areas). Distinct from Civil and Populated Place.

Civil: A political division formed for administrative purposes (borough, county, incorporated place, municipio, parish, town, township). Distinct from Census and Populated Place.

latter are types of civil divisions that are incorporated, but exist primarily in the Eastern half of the US and serve as sub-county divisions. And populated place is represented in neighborhoods—which do not possess formal boundaries, except perhaps at a very local level. Neighborhoods are often cultural or historic designations which may possess limited value beyond the local constituency.

Populated Place: Place or area with clustered or scattered buildings and a permanent human population (city, settlement, town, village). A populated place is usually not incorporated and by definition has no legal boundaries. However, a populated place may have a corresponding "civil" record, the legal boundaries of which may or may not coincide with the perceived populated place. Distinct from Census and Civil classes.

Additional related classes in GNIS are (USGS, 2010a):

Locale: Place at which there is or was human activity; it does not include populated places, mines, and dams but does include battlefield, crossroad, camp, farm, ghost town, landing, railroad siding, ranch, ruins, site, station, windmill.

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

Military: Place or facility used for various aspects of or relating to military activity.

We have chosen to disregard locale in this early study. We consider military places when they coincide with census features only.

Individual records are searched through a GNIS web interface, however, the BGN also provides the contents of the GNIS database via various digital file gazetteers. The records can be derived topically, such as "populated place", containing only the feature class populated place, or "concise", containing all large features appropriate for mapping at a scale of 1:250,000. The records are also accessible as national or state files containing all feature classes for a given geography. The third type of gazetteer file expands the records to add FIPS (Federal Information Processing) 55 Census ID and class codes. The legacy FIPS ID codes have been replaced with American National Standards Institute (ANSI) based GNIS IDs. Yet, the FIPS55 class codes will persist for the near term and are extremely useful in further categorizing populated places. Of interest to our work are (USGS, 2010b) quoted below:

Class C—Incorporated Places

C1: An active incorporated place that does not serve as a county subdivision equivalent.
C2: An active incorporated place legally coextensive with a county subdivision but treated as independent of any county subdivision.
C3: A consolidated city.
C4: An active incorporated place with an alternate official common name.
C5: An active incorporated place that is independent of any county subdivision and serves as a county subdivision equivalent.
C6: An active incorporated place that partially is independent of any county subdivision and serves as a county subdivision equivalent or partially coextensive with a county subdivision but treated as independent of any county subdivision.
C7: An incorporated place that is independent of any county.
C8: The balance of a consolidated city excluding the separately incorporated place(s) within that consolidated government.

Class M—Federal Facilities (including military bases)

Class P—Populated Places

P1: Populated Place that is also an incorporated place with the same name and the same Census Code

Class T—Active Minor Civil Divisions

Class U—Populated (Community) Place

U1: A census designated place with an official federally recognized name.
U2: A census designated place without an official federally recognized name.
U4: Official common name for a populated location within an incorporated place

U5: Populated Place that is also a census designated place with the same name
U6: A populated place that is not a census designated or incorporated place having an official federally recognized name

2.3 Representing Place as a Point or an Area?

GNIS features are all identified by a single XY coordinate. However, linear and areal features that spread out over multiple 1:24,000 topographic maps possess secondary points to ensure that the labeling occurs even when the primary point is on a different map sheet. At larger scales, these locator coordinates may not provide value in label placement if they fall off the area being viewed. Rather, the actual areal feature may provide better placement options in GIS-based mapping environments.

3. OBJECTIVES

Our goals in the initial pilot project included:
- determining point or area label placement for populated places
- evaluating the relationship between populated place points, civil points and census points
- evaluating the use of the Census class codes in categorizing places
- determining attributes which could provide an ordered classification and thereby further label hierarchy

4. METHOD

For our pilot project, we chose the GNIS points for the state of Colorado. Western states have comparatively simpler hierarchy because there are no Minor Civil Divisions (MCDs) such as townships or boroughs. Sub-county place labels comprise incorporated municipalities, Census Designated Places (CDPs) or U4/U6 populated places—neighborhoods. In many Eastern states, the MCD category complicates labeling as MCDs can be coextensive with municipalities though they are not always. Future work takes on the case of MCDs in the state of Pennsylvania, adding this layer of complexity.

Our first consideration was what type of feature representation to use for label placement. Despite the secondary coordinate contained in most GNIS records, we determined that area label placement was best suited for a GIS mapping environment of dynamic labeling that responds to zooming and panning. Hence, we acquired polygon features (Census TIGER/Line Shapefiles) in all cases where a GNIS point records had a corresponding polygon.

This choice allowed us to refine the GNIS populated place point records into point-only records (those neighborhood points possessing no formally-defined boundaries) and populated point records which where duplicates of civil points (incorporated places) and census points (CDPs).
The initial coarse ranking of places:

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

1. incorporated municipalities
2. CDPs
3. populated places which are neither incorporated nor CDPs

allows for an initial prioritization of labels. Categories 1 and 2 could be further classified internally by ranking via population. However, residential population may not be the only attribute relevant to perceptions of place.

To create a more holistic representation of place, we turned to the Census Bureau's Economic Census, which is conducted every five years, most recently in 2007. The Economic Census creates a subset of places (historically, only incorporated places) which meet a given population threshold, varying from census to census. These subsets of places are called "Economic Places". The 2007 Economic Census changed the criteria to include both incorporated places and CDPs and the threshold is met either by residential population of 5000 or 5000 employees in the place. Adding economic place to our regime allowed us to add economic attribute data to our categorization.

When evaluating the Economic Census attribute data, we looked first at composite sector data. Sectors are broad measures of industry within the economy such as mining, agriculture, retail, manufacturing, and services. Hence a composite of all sectors characterizes the overall economy of a specific geography. However, the Economic Census does not aggregate data coarser than two-digit sector codes at place-level geography, and efforts to aggregate the multiple sector data manually proved cumbersome. We turned to the Survey of Business Owners (SBO) data which provides aggregate data on sales and receipts, number of firms, and number of employees at the place level. During our initial pilot the 2007 SBO data had not yet been released at the economic place level, so we reverted back to the 2002 SBO attribute data for number of employees. However, the population thresholds defining an economic place had changed from the 2002 SBO to the 2007 SBO. This did create problems due to our use of 2009 place and economic place TIGER/Line Shapefiles, as some places met the new thresholds in 2007 that had not in 2002. Hence, we did have cases of 2007 defined economic places with no 2002 SBO data, but as the 2007 definition is more robust, we chose to overlook this mismatch until later this year when 2007 SBO data are released and we can join those attributes to our most current TIGER/Line Shapefiles.

The economic place polygons are, essentially, a subset of place polygons, based on 2007 economic place definitions, and hence we can land all economic places directly on top of an identically named place. Having already joined the 2002 SBO data to the economic place shapefile, we collapsed all relevant polygons into points and used a Spatial Join (in ArcGIS 9.3.1) to the place to merge the economic attributes. At this point, we were able to create separate layers using definition queries and exporting the selections for:

1. C1 (incorporated places) which are economic places
2. U1/U2 (CDPs) which are also economic places
3. C1 (incorporated places) which are not economic place
4. U1/U2 (CDPs) which are not economic places

Incorporated places that are also economic places were then classified using the attribute "Number of Employees". Other economic attributes are available, but employees—the number of people working in a place—provides a logical complement to residential population, another attribute that is easily accessible for place.

The final step was to showcase the place hierarchy with the appropriate use of a labeling hierarchy. Places were labeled with Maplex dynamic labels, not converted to annotation using ArcGIS 9.3.1. Incorporated places (C1) with the largest number of workers receive the most prominent label while CDPs receive the least prominent label (Figure 2). Denver, Arvada, Littleton and Sheridan are incorporated places coextensive with economic places ranked according to number of workers. Frederick is an incorporated place coextensive with economic place in 2007 but not in 2002. Applewood is a CDP coextensive with economic place in 2007. Bow Mar and Genesee are an incorporated place and a CDP respectively, neither of which meets the threshold for economic place in 2007.
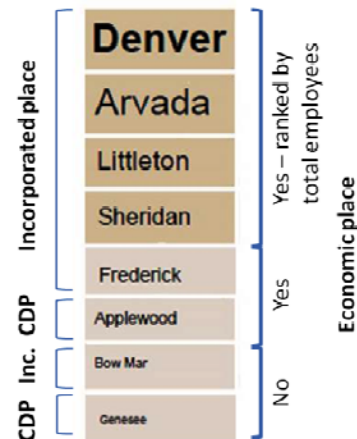

Figure 2. Labels with initial groupings and classifications

Since CDPs could not be economic places in 2007, we reverted CDPs to their 2002 status when creating our label classes (Figure 3) since we currently are using 2002 SBO. Once 2007 SBO data can be added to the 2009 polygons, the new economic place CDPs can be classified using the employment figures in the same manner as incorporated places.
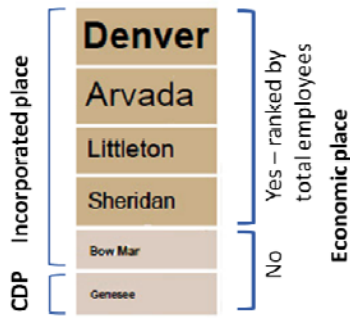
A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

Figure 3. Labels with groupings and classifications amended for the 2002 SBO attribute data.

5. PRELIMINARY RESULTS

Overall, we are pleased with initial tests of incorporating attributes to further create category and hierarchy in labeling of the populated places in the GNIS database. Figures 4 through 9, which have been reduced by 50% from originals for inclusion in this paper, show the results of the labeling scheme through multiple scales. We did not alter the regime for particular scale intervals, but rather evaluated the results using the same method across scales.

Figure 4. Labels captured at 1:1,000,000

Figure 4 shows many of the most prominent cities on the Colorado Front Range in relative hierarchical importance. However, noticeable cities such as Greeley and Loveland in Northern Colorado and Castle Rock to the south of Denver are not labeled at the expense of smaller, less important cities such as Johnstown and Berthoud.
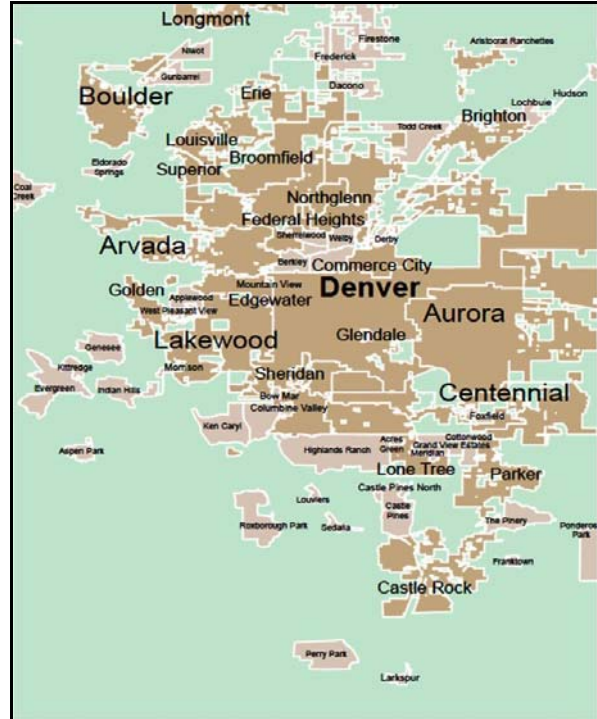

Figure 5. Labels captured at 1:500,000

Figure 5 shows the Denver metropolitan area. There is balance to the overall feel of this scheme at this scale as well, though again there are obvious omissions, including Thornton and Westminster. Centennial feels too prominent in comparison to Boulder and Aurora.

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
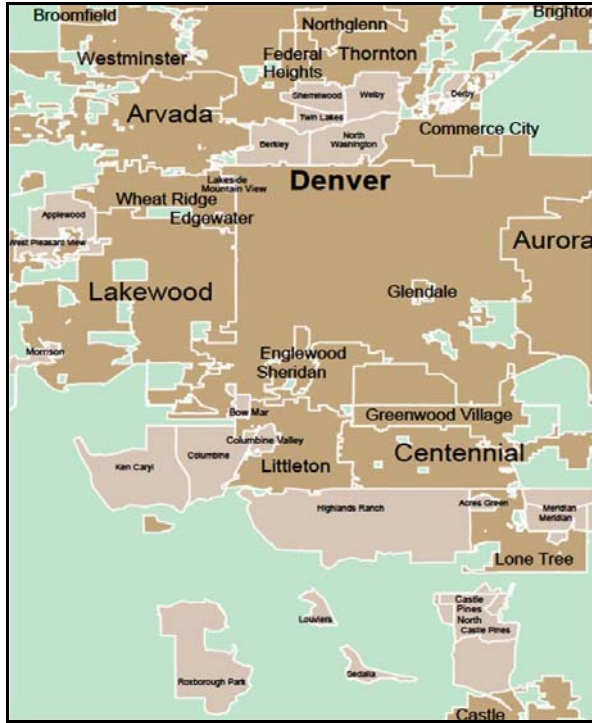November 15-19, 2010 Orlando, Florida

Figure 6. Labels captured at 1:250,000

At the larger scales, the scheme holds up rather well as there are simply fewer labels competing. There may be a natural break between 1:250,000 and 1:500,000 at which point some smaller features are not labeled.
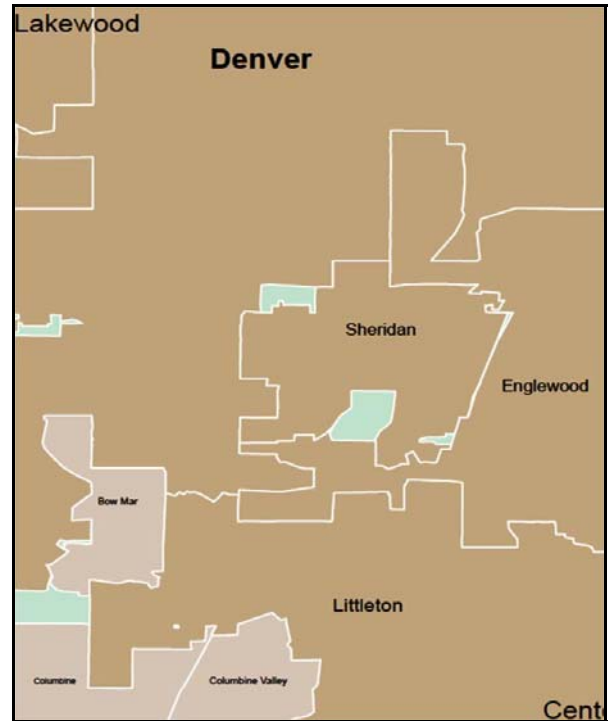

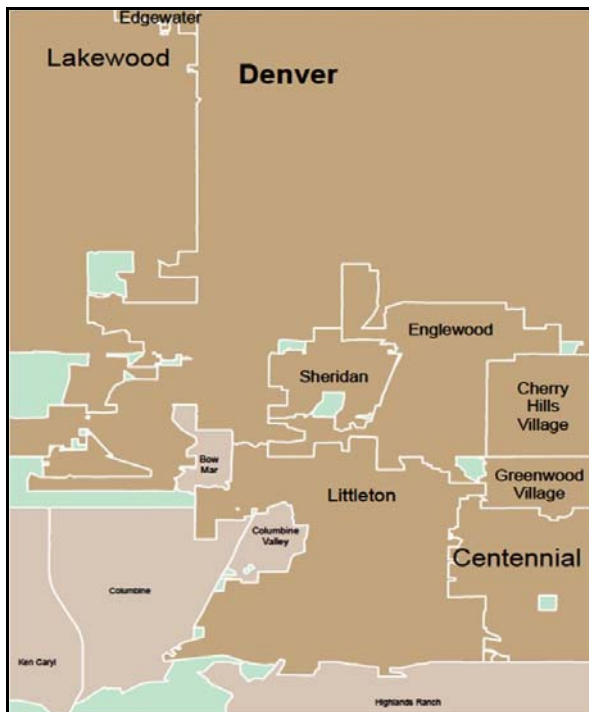Figure 8. Labels captured at 1:50,000

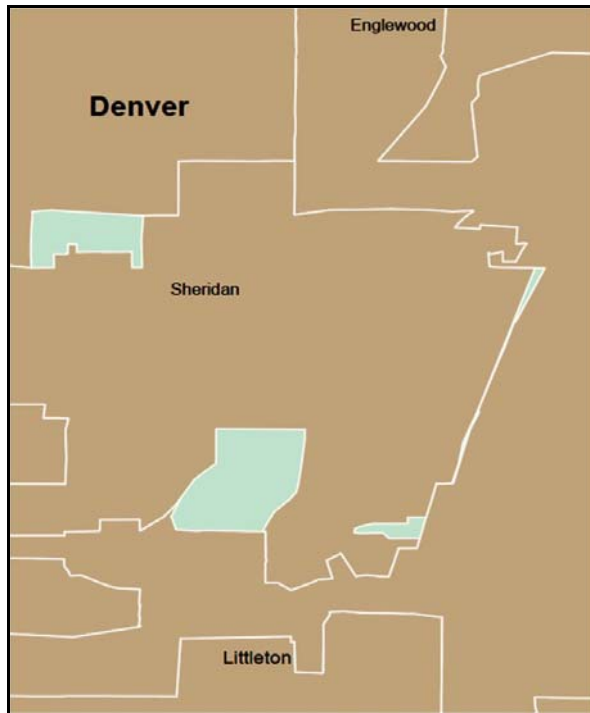Figure 7. Labels captured at 1:100,000

Figure 9. Labels captured at 1:24,000

At scales of 1:50,000 and larger there is ample space to bring in the U4/U6 features. We deem these valuable primarily at the local level, as they are representations of neighborhoods. There is no easily identifiable attribute for population or for economic activity. However, future work will take on this problem by locating the points within Census blocks/block groups/tracts and associating population and number of employees to these polygons, followed by a surface analysis. Points inside polygons with higher combined residential and employment populations will receive priority over those with lower populations. A methodology for ranking those is currently under consideration.

Additional issues which this initial pilot showed:

- Consider using both resident population and employees to further classify our four larger groupings. This could provide extra weighting in Maplex to ensure larger cities are labeled in place of smaller (such as the Greeley/Loveland and Thornton/Westminster examples above.)
- There are Military Bases with a CLASSFP of M1 that could be included in an aforementioned category or in its own category by extending this procedure

The somewhat arbitrary label font and style choices are meant to show relative importance only. We eventually hope to incorporate results into the overall redesign of the USGS topographic maps which would involve a thorough consideration of label styles in conjunction with many other map elements and features.

6. CONCLUSION

The next step is to incorporate the 2007 SBO data once available to add attributes to CDP labeling. Once complete, we will be able to run the labels again to determine where truly obvious omissions are occurring. At that point consideration should be given to determining scale breaks at which classification rules should change. We also plan to test using both worker and residential populations to ensure comparatively large places are always dynamically labeled on maps.

Place labels may well identify the most recognizable features on a map. Though perceptions of place may be somewhat personal to the map reader, the relative importance of place is often communal. By incorporating richer attributes and using the implicit categories, maps can better represent place to the reader. Furthermore, a methodology which incorporates more meaningful attributes, applied thoughtfully across scales will enhance overall map usability.

ACKNOWLEDGEMENTS

REFERENCES

Jordan, P., 2009. Some considerations on the function of place names on maps. *Proceedings of 24th International Cartographic Conference* (ICC2009), Santiago, Chile, November 15-21, 10 pp.

Kadmon, N., 1972. Automated selection of settlements in map generalization. *Cartographic Journal, The,* 9(2), pp. 93-98.

Michna, Izabela, 2009. Generalization of geographic names on atlas maps. *Proceedings of 24th International Cartographic Conference* (ICC2009), Santiago, Chile, November 15-21, 10 pp.

U.S. Board on Geographic Names, 2003. *Principles, Policies, and Procedures.* http://geonames.usgs.gov/domestic/policies.htm (accessed 15 Mar. 2010)

U.S. Census Bureau, 2010a. *2002 Survey of Business Owners: Economy Wide Estimates of Business Ownership: State of Colorado: Place.* http://factfinder.census.gov (accessed 15 Mar. 2010)

U.S. Census Bureau, 2010b. *2009 TIGER/Line Shapefiles.* http://www.census.gov/geo/www/tiger/tgrshp2009/tgrshp2009.html (accessed 15 Mar. 2010)

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

U.S. Geological Survey, 2010a. *The National Map* Viewer. http://viewer.nationalmap.gov/viewer/ (accessed 9 Sep. 2010)

U.S. Geological Survey, 2010b. Feature class definitions. http://geonames.usgs.gov/domestic/feature_class.htm (accessed 9 Sep. 2010)

U.S. Geological Survey, 2010c. Census class code definitions. http://geonames.usgs.gov/pls (accessed 9 Sep. 2010)

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida