Establishing Classification and Hierarchy in Populated Place Labeling for Multiscale Mapping for The National Map

Wesley Stroh, Stephen Butzler and Cynthia Brewer

Gould Center, Geography
Pennsylvania State University
Portion of USGS ScaleMaster work (24K, 50K, 100K...)
289K: With our basic hierarchy...

...and with no place hierarchy
578K: With our basic hierarchy... ...and with no place hierarchy
Background – Point vs Polygon

*The National Map* – data, portal and mapping tool

GNIS **point** data is names theme
- Limited for label positioning
- Limits use of attributes for hierarchy

An alternative – U.S. Census Bureau **polygons**
- GNIS place points (many) come from census geography
- Distinguish between places and *economic* places
- Class census attribute data for additional hierarchy
Background – GNIS

Geographic Names Information Service (GNIS)
Maintained by USGS Board on Geographic Names (BGN)

GNIS Feature Class Codes (broadest definitions of place)

• **Populated Place**—subjective definition of place
• **Civil**—incorporated places (administrative)
• **Census**—census-defined places (statistical)
• **Locale**—historic human activity (not active)
Census Data Types

2009 TIGER/Line Shapefiles (geography)
- Incorporated Places
- Census Designated Places (CDPs)
- Minor Civil Divisions (MCDs), no CCDs used
- Economic Places
  (NOTE: 2007 Economic Census definitions of Economic Place include all places with EITHER 5000 residents OR 5000 employees)

2002 Survey of Business Owners (attributes)
- part of Economic Census (every 5 years)
- provides data on employment numbers (used to create our classifications)
- place-level data for the 2007 SBO was not yet available at time of analysis
Cases and Methods

Pilot tests: Colorado (Wes) and Pennsylvania (Steve)
- Distinct geographies of “place”
  - MCDs prevalent in Pennsylvania (eastern U.S.)
  - PA denser, CO less dense

Initial (coarse) hierarchy of TIGER/Line polygons ranked by FIPS 55
- 1\textsuperscript{st} C1-6 – Incorporated municipalities
- 2\textsuperscript{nd} U1/U2 – CDPs
- 3\textsuperscript{rd} T1 – MCDs (eastern US only; e.g., townships)
Additional Hierarchy: Build Pittsburgh...

1st: Incorporated Place (C5) and Economic Place
1st: Incorporated Place (C5) and Economic Place

2nd: Incorporated Place (C5)
1st: Incorporated Place (C5) and Economic Place

2nd: Incorporated Place (C5)

3nd: CDP (U1/2) / MCD (T) and Economic Place

Pittsburgh
1st: Incorporated Place (C5) and Economic Place

2nd: Incorporated Place (C5)

3rd: CDP (U1/2) / MCD (T) and Economic Place

4th: MCD (T) and Economic Place
1st: Incorporated Place (C5) and Economic Place
2nd: Incorporated Place (C5)
3rd: CDP (U1/2) / MCD (T) and Economic Place
4th: MCD (T) and Economic Place
5th: CDP (U1/2)/MCD (T)
1\textsuperscript{st}: Incorporated Place (C5) and Economic Place

2\textsuperscript{nd}: Incorporated Place (C5)

3\textsuperscript{nd}: CDP (U1/2) / MCD (T) and Economic Place

4\textsuperscript{th}: MCD (T) and Economic Place

5\textsuperscript{th}: CDP (U1/2)/MCD (T)

6\textsuperscript{th}: MCD (T)
Census Class Codes (FIPS 55)

Class codes further define census polygons:

Class **C** — incorporated places
Class **M** — military places
Class **P** — pop. place = incorp. place
Class **T** — MCD (e.g. township)
Class **U** — CDP

Note: We ranked MCDs coextensive with CDPs (T/U) higher than sole MCDs (T).
Why TIGER/Line Shapefiles?

Polygon label placement more dynamic
- labels reposition in parts of polygons with zoom and pan

GNIS points provide single centroid
- Additional place point for adjacent topo quad, but limited use in seamless context
- However, GNIS points do provide an FID which could link back to census data...

NationalMapping.us

AutoCarto, November 2010
Other Possible Attributes

Additional hierarchy with census data

• In pilot, used SBO “**Number of Employees**” but there are other indicators …

• For places:
  – Total population

• For economic places:
  – Total number of firms
  – Total revenue
Processing for three layers

- Repeating geometry
- Unique geometries
- Slivers ★
• Sub-county polygons most comprehensive (MCD layer)
• Collapse place and Economic Place polygons to points
• Identify which places do not have a polygon at sub-county level
• Strip out identified unique-geometry place points
• **Spatial Join** the remaining place and Economic Place point attributes to sub-county polygons
• Create separate polygon layer of unique-geometry places
• Then...establish hierarchy of places
<table>
<thead>
<tr>
<th>PA Place Hierarchy</th>
<th>CO Place Hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pittsburgh</td>
<td>Denver</td>
</tr>
<tr>
<td>West Mifflin</td>
<td>Aurora</td>
</tr>
<tr>
<td>Bethel Park</td>
<td>Englewood</td>
</tr>
<tr>
<td>Mount Oliver</td>
<td>BowMar</td>
</tr>
<tr>
<td>Penn Hills</td>
<td>Columbines</td>
</tr>
<tr>
<td>Cranberry</td>
<td></td>
</tr>
<tr>
<td>Reserve</td>
<td></td>
</tr>
<tr>
<td>Mount Pleasant</td>
<td></td>
</tr>
<tr>
<td>Gastonville</td>
<td></td>
</tr>
<tr>
<td>Shadyside</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>T</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorp. Place</td>
<td>Economic Place</td>
<td>MCD (e.g. township)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CDP</td>
</tr>
<tr>
<td>Incorporated place</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incorporated place</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic place</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic place</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic place, CDP and MCD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic place and MCD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCD and CDP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(dashed outline)</td>
<td></td>
<td>CDP</td>
</tr>
<tr>
<td>U4 and U6 GNIS points (no polygons)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*CDP: Census Designated Place*
PA Best design 2 mill-ish
Top 4 of 5 classes in R1
2.3M
<table>
<thead>
<tr>
<th>PA Place Hierarchy</th>
<th>CO Place Hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pittsburgh</td>
<td>Denver</td>
</tr>
<tr>
<td>West Mifflin</td>
<td>Aurora</td>
</tr>
<tr>
<td>Bethel Park</td>
<td>Englewood</td>
</tr>
<tr>
<td>Mount Oliver</td>
<td>BowMar</td>
</tr>
<tr>
<td>Penn Hills</td>
<td>Columbine</td>
</tr>
<tr>
<td>Cranberry</td>
<td>Roxborough Park</td>
</tr>
<tr>
<td>Reserve</td>
<td></td>
</tr>
<tr>
<td>Mount Pleasant</td>
<td></td>
</tr>
<tr>
<td>Gastonville</td>
<td></td>
</tr>
<tr>
<td>Shadyside</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>T</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorpor-</td>
<td>Economic</td>
<td>Place</td>
</tr>
<tr>
<td>orated</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

U4 and U6 GNIS points (no polygons)
Places with No Census Attributes

At large scales add GNIS point data of
• U4/U6 – Populated Places
Evaluating Our Results

1. Are labels placed on maps?
2. Does the additional label hierarchy provide value?

We compared our results to *The National Map Viewer* and *Google maps* at three scales ...
Pittsburgh at 1:144,448 using our place classing with The National Map Viewer and Google maps comparisons
Yes... different sets of other features labeled and different label sizes among compared maps interfere with precise comparisons—this is pilot work.

Pittsburgh at 1:288,895 using our place classing with The National Map Viewer and Google maps comparisons.
Pittsburgh at 1:577,791 using our place classing with The National Map Viewer and Google maps comparisons.
Google/Bing scales for cached map tiles in WGS 1984 Web Mercator projection

<table>
<thead>
<tr>
<th>Scale</th>
<th>File Size</th>
<th>Scale</th>
<th>File Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1,128.497220</td>
<td>1.2K</td>
<td>1:1,155,581.153</td>
<td>1.2M</td>
</tr>
<tr>
<td>1:2,256.994440</td>
<td>2.3K</td>
<td>1:2,311,162.307</td>
<td>2.3M</td>
</tr>
<tr>
<td>1:4,513.988880</td>
<td>4.5K</td>
<td>1:4,622,324.614</td>
<td>4.6M</td>
</tr>
<tr>
<td>1:9,027.977761</td>
<td>9K</td>
<td>1:9,244,649.227</td>
<td>9.2M</td>
</tr>
<tr>
<td>1:18,055.95552</td>
<td>18K</td>
<td>1:18,489,298.45</td>
<td>18.5M</td>
</tr>
<tr>
<td>1:36,111.91104</td>
<td>36K</td>
<td>1:36,978,596.91</td>
<td>37M</td>
</tr>
<tr>
<td>1:72,223.82209</td>
<td>72K</td>
<td>1:73,957,193.82</td>
<td>74M</td>
</tr>
<tr>
<td>1:144,447.6442</td>
<td>144K</td>
<td>1:147,914,387.6</td>
<td>148M</td>
</tr>
<tr>
<td>1:288,895.2884</td>
<td>289K</td>
<td>1:295,828,775.3</td>
<td>296M</td>
</tr>
<tr>
<td>1:577,790.5767</td>
<td>578K</td>
<td>1:591,657,550.5</td>
<td>592M</td>
</tr>
</tbody>
</table>
Evaluation: Label Inventory Process

289K Placenames

Google map

TNM Viewer

PSU

6 pages of placenames
Comparison among PSU, Google, TNM maps—High category, ordered by number employees:

Dark cells indicate placed labels and light cells for unplaced labels
# Unplaced Label Performance

<table>
<thead>
<tr>
<th>Most significant cities, boroughs &amp; townships</th>
<th>Other boroughs</th>
<th>Other townships</th>
<th>Other populated places (U4 U6 points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSU</td>
<td>Google</td>
<td>TNM</td>
<td>Viewer</td>
</tr>
<tr>
<td>R1, R2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>144K number</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>percent</td>
<td>4</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>288K number</td>
<td>0</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>percent</td>
<td>0</td>
<td>17</td>
<td>41</td>
</tr>
<tr>
<td>577K number</td>
<td>13</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>percent</td>
<td>19</td>
<td>19</td>
<td>57</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unplaced</td>
<td>14</td>
<td>28</td>
<td>67</td>
</tr>
<tr>
<td>percent</td>
<td>9</td>
<td>19</td>
<td>45</td>
</tr>
</tbody>
</table>

Yellow cells highlight map with fewest unplaced labels, by scale and label group

Stroh, Butzler, Brewer – NationalMapping.us

AutoCarto, November 2010
Same labels on maps by PSU, Google and TNM Viewer at three scales

Incorporated and economic place
MCD, CDP and economic place
Incorporated place
Other MCD/CDP

0     5    10   15   20   — number of labels

144K
289K
578K

R1
R2
R3
R4/5/6

Stroh, Butzler, Brewer – NationalMapping.us

AutoCarto, November 2010
Labels compared to PSU map

- Incorporated and economic place
- MCD, CDP and economic place
- Incorporated place
- Other MCD/CDP

(on right) PSU map **did better** than Google and/or TNM Viewer

(on left) PSU map **did worse** than Google and/or TNM Viewer

- Incorporated places not on PSU map
- Many more township names on PSU map

Number of labels — 30 25 20 15 10 5 0 5 10 15 20 25 30 35 40 45
Labels **un**placed on all maps

- Incorporated and economic place
  - MCD, CDP and economic place
  - Incorporated place
  - Other MCD/CDP

- Labels **not on** maps by PSU, Google and TNM Viewer at three scales

- 144K
- 289K
- 589K

- Number of labels

Stroh, Butzler, Brewer – NationalMapping.us

AutoCarto, November 2010
Next steps ...

Address the U4/U6s

- Subjective places (e.g. neighborhood points)
- No attribute data
- Surface analysis of block group level data perhaps
  - Residential population
  - Number of employees (using LED)
  - Overlay U4/U6 points using surface hotspots to rank
  - Use ranks for label hierarchy or pruning decisions

And make maps
Acknowledgements

USGS Center of Excellence for Geospatial Information Science (CEGIS), 2007 to present

Resources
Project resources: ScaleMaster.org
Gould Center: NationalMapping.us
CEGIS-USGS: cegis.usgs.gov

Thanks:
Babs Buttenfield
CU-Boulder
Lynn Usery
Larry Stanislawski
USGS-CEGIS
Charlie Frye
ESRI

Stroh, Butzler, Brewer

AutoCarto, November 2010