Geographical Information Improvements

Carl Anderson, GISP

LED Partnership Workshop
Arlington, VA
March 9-10, 2011
Why Does Geography Matter?

Administrative record-based systems like OnTheMap use the universe of outcomes of interest measured directly at the geographical level of interest.*

* John M Abowd - APDU Newsletter Vol 33 No 2
Why Does Geography Matter?

- County Level Detail
Why Does Geography Matter?

- More detailed Geography, Allows for much more useful detail
Why Does Geography Matter?

Local Employment Dynamics

Local = Geography

Addresses = Geography

Worker/ Establishment = Geography
Challenges in Addressing

- Jurisdictions that have no addresses
- Unusual addressing schemes
- Establishments located by ZIP code only
- Establishments located without ZIP code
- Changes in ZIP code vs the longitudinal data
- Timing of Changes – How and when do changes appear in administrative records
Causes of Addressing Challenges

- Interpretation of data
- Ambiguous locations
- Missing data
Missing data

Longitudinal nature of the records

<table>
<thead>
<tr>
<th>Year Q</th>
<th>Company</th>
<th>Address</th>
<th>City</th>
<th>ZIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006Q2</td>
<td>ABC Ltd</td>
<td>451 Maple St</td>
<td>Anytown</td>
<td>20789</td>
</tr>
<tr>
<td>2006Q3</td>
<td>ABC Ltd</td>
<td>451 Maple St</td>
<td></td>
<td>20789</td>
</tr>
<tr>
<td>2006Q4</td>
<td>ABC Ltd</td>
<td>451 Maple St</td>
<td>Anytownn</td>
<td>OH  20789</td>
</tr>
<tr>
<td>2007Q1</td>
<td>ABC Ltd</td>
<td>451 Maple St</td>
<td>Anytown</td>
<td>OH  20789</td>
</tr>
<tr>
<td>2007Q2</td>
<td>ABC Ltd</td>
<td>451 Maple St</td>
<td>Anitown</td>
<td>20789</td>
</tr>
<tr>
<td>2007Q3</td>
<td>ABC Ltd</td>
<td>451 Maple St</td>
<td>Anytown</td>
<td>OH</td>
</tr>
<tr>
<td>2007Q4</td>
<td>ABC Ltd</td>
<td>St Rte 65 &amp; 451 Maple St</td>
<td>Anytown</td>
<td>MD  21789</td>
</tr>
<tr>
<td>2008Q1</td>
<td>ABC Ltd</td>
<td>451 Maple St</td>
<td></td>
<td>MD  21789</td>
</tr>
<tr>
<td>2008Q2</td>
<td>ABC Ltd</td>
<td>451 Maple St</td>
<td>Anytown</td>
<td>MD  21789</td>
</tr>
</tbody>
</table>

* Construed data, for use as an example only*
Interpretation of Addresses

- Postal vs Colloquial Addressing
- Develop strategies for specific unusual Address forms

Example:

40 Miles N of Sand Creek, Barrow Ak
US 4 & Rt 401 / 407 Westhill Road

* Construed data, for use as an example only
Interpretation of Addresses Cont’d

- Requires Developing Careful Strategies

Example:

403-22A Main Street NW
West 403 Main Street North Apt 22A
N403W22 Main Street Apt A
Ambiguous Addresses

- Location may be indistinct unless all address elements are perfect

- Uncertain geography

Example:
2\textsuperscript{nd} Ave SE, Atlanta GA, 30317
2\textsuperscript{nd} Ave NE, Atlanta GA, 30030
2\textsuperscript{nd} Ave, Atlanta GA, 30032
2\textsuperscript{nd} St, Atlanta GA, 30360
2\textsuperscript{nd} St NE, Atlanta GA, 30317
2\textsuperscript{nd} St NW, Atlanta GA, 30318
Ambiguous Addresses

Any one of 3 Counties and 4 Cities
What is LEHD doing to improve Geography & Addressing?

1. Redesigning Address Geocoding Process:
   - Improve interpretation of varying address styles (formats, special cases etc.)
   - Improve handling of missing data
   - Improve spatial accuracy
   - Utilize longitudinal address review
   - Utilize supplemental reference address data

2. Building improved tools to monitor the quality of Addresses & Geography within LED
Geocoding Process Improvements

**Existing**

- ACS POW
- AHS
- ES-202
- SSEL
- MAF

- Mix Address Data from All Sources
- Match, Standardize, Geocode, and De-Duplicate
- Link Datasets using Matched Addresses
- Build Cross-Walk Tables
- Add Census Geographic Codes based on links into Census Bureau Master Address File
-Approximate Census Geographic Codes using range searches into Geocoded Addresses
- Extrapolate Census Geographic Codes From other data (Tiger)
- Finalize Geographic codes with additional attributes

**Redesigned**

- ACS POW
- AHS
- ES-202 + Lat/Long
- SSEL
- MAF
- BR + Lat/Long
- CPR

- Collect Addresses By entity Key
- Build longitudinal “Best Observation” table for entities
- Longitudinal Edits To infill missing observations and partial observations
- Match, Standardize, Geocode, and De-Duplicate
- Update entity Keys
- Link Datasets using Matched Addresses

- Add Census Geographic Codes based on links into Census Bureau Master Address File
- Add Preliminary Census Geographic Codes
- Augment Data, Adding Missing Census Geographic Codes
- Validate assigned Census Geographic Codes using information in Data Sources
- Approximate Census Geographic Codes using range searches into Geocoded Addresses
- Extrapolate Census Geographic Codes From other data (Tiger, TPN, NSDI)
- Finalize Geographic codes with additional attributes
Quality Improvements & Visualization Cont’d

- Improved Quality Reports

**EQUI trends for (selected states)**

- 100%
- 90%
- 80%
- 70%
- 60%
- 50%
- 40%
- 30%
- 20%
- 10%
- 0%

- Improve Quality Reports

**Address Resolution by type**

- A
- B
- C
- D
- E
- F
- G

**Map of the USA**

- Improve Quality Reports

**Graph**

- Improve Quality Reports
How Can State LMI offices Help?

- Assist in understanding History of data submissions
- Participate in Data Quality Feedback Loops
- Act as Liaisons to other State offices who collect and record address data
Questions?

CES.Local.Employment.Dynamics@census.gov