APPLYING OPERATIONS RESEARCH PRINCIPLES TO IMPROVE GEO-DATABASE QUALITY

Lean Principles and Value-Stream Mapping as Tools for Improving Geocoding Success

2014 ESRI International User’s Conference
Data Compilation Case Studies: QA/QC
July 17, 2014
San Diego, CA

Anirudh Kannan Vinakayaram, MEME
Research Scientist
Geospatial and Population Studies
University of New Mexico

Ron Lumia, PhD
Professor, Mechanical Engineering
University of New Mexico

Ganesh Balaksharan, PhD
Assistant Professor, Electrical and Computer Engineering
Center for High Tech Materials, CHTM
University of New Mexico

Jack Baker, PhD
Sr. Research Scientist
Geospatial and Population Studies
University of New Mexico

Nathan Crouse, MS
Analyst Programmer
Geospatial and Population Studies
University of New Mexico
PRESENTATION OUTLINE

• Potential and Pitfalls in Geocoding for Small-Area Demographic Estimates: Project Goals
• Geocoding as a Manufacturing Process
• Lean Principles & Value-Stream Mapping
• Case Study & Simulation
• Results & Recommendations
GEOCODING AND SMALL-AREA DEMOGRAPHIC ESTIMATES

Geospatial and Population Studies (GPS) at the University of New Mexico
GEOSPATIAL AND POPULATION STUDIES
AT THE UNIVERSITY OF NEW MEXICO

- Makes official population estimates for State of New Mexico.

- Estimates:
  - Form basis of health surveillance and reporting.
  - Inform public sector decision-making (legislation)
  - Support private sector decision-making (real estate, business development, etc.)

- Input data used in GPS estimation process:
  - Informs the US Census Bureau’s Population Division in making their State, County, and Municipal-level estimates.
  - Directly fed into the updating of the US Census Bureau’s Master Address File through programs including the Local Update of Census Addresses and the 2010 Count Review Program.
POTENTIAL AND PITFALLS IN GEOCODING FOR DEMOGRAPHIC ESTIMATES

1. Geo code address data.
2. Summarize counts within geographic units.
3. Make demographic estimates.
PROJECT GOALS

- Improvements in geocoding mean more accurate estimates.
- Geocoding success rates vary:
  - 75 to 90% in urban areas
  - 10% to 40% in rural areas
- **Goal:** Improve geocoding in urban areas by 5%--moving from 90% to 95%.
- This improvement could result in a savings of approximately $\frac{1}{5}$ of the GPS operating budget for geocoding and improve the accuracy of small-area demographic estimates by as much as 10 percentage points.
LEAN PRINCIPLES AND VALUE-STREAM MAPPING
GEOCODING AS A PRODUCTION PROCESS
LEAN PRINCIPLES IN GEOCODING

- Eliminate waste
- Waste adds cost with no value
- Define “value”
- Goal is to minimize waste to maximize value.
CONCEPTS OF WASTE (MUDA)

7 types of waste defined by Toyota

1. Overproduction
2. Waiting
3. Transporting
4. **Process Defects (Quality Control)**
5. Excess Inventory
6. Unnecessary Motion
7. Inappropriate Processing
VALUE STREAM MAPPING AS A TOOL

Current State Map for GPS Geocoding Process

- Value Stream Mapping
  - Operationalizes Lean Principles
  - Represents a process graphically
  - Uses this graphic as a means of identifying waste
  - Suggests avenues for process improvement (eliminate waste)
- Graphic = “State Map”
  - Current
  - Future (lean principles applied)
CASE STUDY AND SIMULATION
CASE STUDY

• Case study is used to identify defects in standardization process.

• Provides information on the specific types of errors and their frequency.

<table>
<thead>
<tr>
<th>Error Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete addresses</td>
<td>11</td>
</tr>
<tr>
<td>Post Direction</td>
<td>22</td>
</tr>
<tr>
<td>Street Type</td>
<td>14</td>
</tr>
<tr>
<td>On road</td>
<td>1</td>
</tr>
<tr>
<td>Address Change</td>
<td>1</td>
</tr>
<tr>
<td>Misspelling</td>
<td>25</td>
</tr>
<tr>
<td>Incorrect Address</td>
<td>3</td>
</tr>
<tr>
<td>Zip Code</td>
<td>7</td>
</tr>
<tr>
<td>Excess</td>
<td>1</td>
</tr>
<tr>
<td>Post Direction and Street Type</td>
<td>24</td>
</tr>
<tr>
<td>Space between address</td>
<td>2</td>
</tr>
<tr>
<td>Name</td>
<td>2</td>
</tr>
<tr>
<td>Incomplete addresses and PD</td>
<td>1</td>
</tr>
<tr>
<td>Multiple addresses</td>
<td>2</td>
</tr>
</tbody>
</table>
RESULTS FROM CASE STUDY

Types of errors:
- Multiple addresses
- Incomplete addresses and PD
- Name
- Space between address
- Post Direction and Street Type
- Excess
- Zip Code
- Incorrect Address
- Misspelling
- Address Change
- On road
- Street Type
- Post Direction
- Incomplete addresses

Number of errors
SIMULATION STUDY

• Used to study the production system and mimic the error-generating process.

• An artificial city was built with 100 x 100 houses.

• 4 Zip Codes, 4 Post Directions and 2 Street Types were used.

• Addresses were corrupted according to frequency and type of error identified in case-study.

• **Correction algorithms were designed in light of results.**
FUTURE STATE MAP—SIMULATION TO ACTION
A NEW STATE MAP AND REMEDIATION PROCESS

1. Local Address Data
2. Standardization
3. Initial Address List
4. Geocode
   - Remediate through Re-standardization
   - Fall Out
   - Successful
5. Final Address List
RESULTS AND RECOMMENDATIONS
RESULTS

• New State Map/Process Results:
  • 95.79% of defective addresses were corrected.
    • House numbers (missing or out of range)
    • Incorrect Zip Code designations
    • PO Boxes
  • 79% of the remediated addresses were successfully re-geocoded.
  • In the final tally, geocoding increased from 78% to 95%--a 17 percentage point improvement.
## COST OFFSET

<table>
<thead>
<tr>
<th>Activity</th>
<th>Number of Records</th>
<th>Cost Per Record</th>
<th>Total Cost</th>
<th>Cumulative Cost</th>
<th>Percent of Data Capture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geo-coding</td>
<td>20,379</td>
<td>$0.50</td>
<td>$10,189.50</td>
<td>$10,189.50</td>
<td>78.00</td>
</tr>
<tr>
<td>Interactive Hand-Matching</td>
<td>2,874</td>
<td>$1.00</td>
<td>$2,874.00</td>
<td>$13,063.50</td>
<td>88.00</td>
</tr>
<tr>
<td>Aerial Imagery Review/Google Maps</td>
<td>1,725</td>
<td>$5.00</td>
<td>$8,625.00</td>
<td>$21,688.50</td>
<td>94.65</td>
</tr>
<tr>
<td>Fieldwork</td>
<td>1,149</td>
<td>$15.00</td>
<td>$17,235.00</td>
<td>$38,923.50</td>
<td>100.00</td>
</tr>
<tr>
<td>Geo-coding</td>
<td>20,379</td>
<td>$0.50</td>
<td>$10,189.50</td>
<td>$10,189.50</td>
<td>78.00</td>
</tr>
<tr>
<td>Proposed Process</td>
<td>4,540</td>
<td>$.23</td>
<td>$1,023.75</td>
<td>$11,213.25</td>
<td>95.28</td>
</tr>
<tr>
<td>Left Imagery/Fieldwork</td>
<td>1,208</td>
<td>$10.00</td>
<td>$12,080.00</td>
<td>$23,293.25</td>
<td>100.00</td>
</tr>
</tbody>
</table>
RECOMMENDATION: FUTURE STATE MAP
FUTURE IMPROVEMENTS

• Focus on methods for remediating errors in house numbers
  • Missing (Imputation)
  • Out of Range (Machine Learning Correction Algorithms)
  • PO Boxes (Linkages)

• Zip Codes
  • Error rates in assignment
  • Missing elements (incomplete digits)
  • Zip code changes over time
ACKNOWLEDGMENTS

• Dr. Adelamar Alcantara, Director of Geospatial and Population Studies

• Dr. Michael Sonksen, Assistant Professor of Statistics, University of New Mexico
CONTACT INFO

Anirudh Kannan Vinayakaram
Research Scientist
Geospatial and Population Studies
MSC06 3510
1 University of New Mexico
Albuquerque, NM 87131
(505) 277-1704
akannan@unm.edu