Distribution of Quality Control to the Point of Data Collection in the Field: Impacts on Cost, User Experience, and Security

Andrea Johnson (U.S. Census Bureau)
Jonathan Krentel (Gunnison Consulting Group, Inc.)
Why Collect Housing Unit Locations?

• Accurate housing unit location matters
2010 Address Canvassing Map Spot Collection

1) Locate Housing Unit

2) Locate GPS “you are here” indicator on map

3) Tap on map

4) Confirm location of map spot (from tap)
Map Spot Collection Results

• 106,000,000 Map Spots collected

• For every map spot collected, two coordinate pairs were returned to the Census:
  – Manual
  – GPS
  – Only one point is “preferred”

• Three primary outcomes:
  – GPS = Manual
  – GPS location correct but manual location wrong
  – Manual location correct, GPS is wrong
Map Spot Collection Issues that Required Resolution

- **Curbstone Clusters**
  ~ Workload: 210,000

- **Map Spot Reconciliation**
  1,500,000

- **Feature Name Reconciliation**
  1,150,000
Curbstone Clusters Resolution

A) After Collection
B) If GPS Accepted and No Manual Review
A) After Manual Review
High Cost of Manual Resolution

<table>
<thead>
<tr>
<th>Task</th>
<th>People</th>
<th>Job Code</th>
<th>Total Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop Procedures</td>
<td>1</td>
<td>Geographer</td>
<td>120</td>
</tr>
<tr>
<td>Conduct Training</td>
<td>4</td>
<td>Geographers</td>
<td>128</td>
</tr>
<tr>
<td>Monitor Analysis</td>
<td>4</td>
<td>Geographers</td>
<td>128</td>
</tr>
<tr>
<td>Resolve Referrals</td>
<td>3</td>
<td>Geographer</td>
<td>384</td>
</tr>
<tr>
<td>Training</td>
<td>27</td>
<td>Clerical</td>
<td>864</td>
</tr>
<tr>
<td>Perform Analysis</td>
<td>27</td>
<td>Clerical</td>
<td>8208</td>
</tr>
<tr>
<td>Conduct Quality Control</td>
<td>4</td>
<td>Supervisory Clerical</td>
<td>384</td>
</tr>
<tr>
<td>Misc IT (Upload, Systems Dev)</td>
<td>3</td>
<td>IT Specialists</td>
<td>180</td>
</tr>
</tbody>
</table>

**Unique People**  ~34

**Total** 10,396
LiMA: Structure Type
LiMA: Address Status

Please select the Address Status for this structure.

- HOUSING UNIT
- UNINHABITABLE
- NONRESIDENTIAL
- GROUP QUARTERS
- DOES NOT EXIST
- OTHER LIVING QUARTERS
- UNDER CONSTRUCTION
- DEMOLISHED
- UNABLE TO LOCATE
- DUPLICATE
LiMA: Location Address
LiMA: Map-spotting
LiMA: Long Strand
Mapping Considerations

Locally rendered vector-based maps vs. tile-based maps
Benefits of vector-based maps (1)

Quality = Spatial-awareness
Benefits of vector-based maps (2)

Quality = Real-time detection of clustering
Benefits of vector-based maps (3)

Quality = Real-time long strand detection
Benefits of vector-based mapping

(4)

Quality = Maintain topological integrity when adding or updating features
“Thick Client” Technical Approach

• Benefits:
  – Rich data on the device
    • All objects on the map are programmatically addressable
    • Allows for locally executed Quality Control
    •Potentially eliminates the need for expensive post-processing operations
    • Allows for QC to take place by the person in the best position to make corrections
  – Enables native user experience
  – Mitigates connectivity risk

• Risks:
  – Multiplication of code bases = BYOD risk?
  – Data on the device = Security risk?
  – Heavy processing load on the device = Performance risk?