Building Energy Data Analysis (BEDA) Accelerator

January 8, 2019
Starting at 2:05 ET
Agenda

- Introductions & Who’s on the line
- Status updates from the field
  - DOE Updates: Miami-Dade County Data Analysis & Overview
  - Around the phone updates
- Product Road Map
- UBID Look-Up Site Preview
- Time Permitting: UBID Implementation Process
DOE Data Sharing Updates
Analysis of Miami-Dade County
UBID for Miami-Dade County

Mark Borkum
Computer Scientist
Data Source

- ESRI Shapefile
  - Polygons
  - WGS-84 projection
Methodology

• Open source:
  ▪ UBID assignment

• Closed source:
  ▪ Screening
  ▪ Duplicate detection
  ▪ Statistical analyses

Download data → Convert from ESRI Shapefile to CSV file

Screen for invalid records → Assign UBIDs to valid records

Detect duplicate records → Calculate UBID-enabled statistics
Screening

• “OBJECTID” values for invalid records:
  - 111697
  - 111735
  - 114083
  - 113512
  - 124290
  - 373308
• “OBJECTID” field = No duplicates
• “UNIQUE_ID” field = No duplicates

• UBID = 136 duplicates of 68 strings
  ▪ 2 duplicates per string

• 0.0122% of dataset is duplicated data

Duplicate Detection

• UBID detects duplicates that elude other methods
Statistical Analyses

Up to 99th %tile shown

- Chebyshev distance
- Manhattan distance
- Normalized Euclidean distance
- Rectangularity
- Percentage area increase
- Rotation
### Statistical Analyses (cont.)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>25(^{th}) %tile</th>
<th>50(^{th}) %tile</th>
<th>75(^{th}) %tile</th>
<th>99(^{th}) %tile</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chebyshev distance (OLCs)</td>
<td>0.939696</td>
<td>0.637992</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td>Manhattan distance (OLCs)</td>
<td>1.253446</td>
<td>0.954960</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>Normalized Euclidean distance (%)</td>
<td>0.037889</td>
<td>0.022865</td>
<td>0.000055</td>
<td>0.021009</td>
<td>0.034132</td>
<td>0.050323</td>
<td>0.109327</td>
<td>0.345063</td>
</tr>
<tr>
<td>Rectangularity (%)</td>
<td>0.547195</td>
<td>0.117476</td>
<td>0.020155</td>
<td>0.472921</td>
<td>0.557375</td>
<td>0.627712</td>
<td>0.800137</td>
<td>1.142382</td>
</tr>
<tr>
<td>Percentage area increase (%)</td>
<td>0.934695</td>
<td>0.546434</td>
<td>-0.124636</td>
<td>0.593087</td>
<td>0.794124</td>
<td>1.114519</td>
<td>2.942340</td>
<td>48.616482</td>
</tr>
<tr>
<td>Rotation (deg.)</td>
<td>74.022433</td>
<td>28.302354</td>
<td>0</td>
<td>84.037135</td>
<td>87.845227</td>
<td>88.572023</td>
<td>89.923000</td>
<td>89.999915</td>
</tr>
<tr>
<td>Area (sq. km)</td>
<td>0.350906</td>
<td>1.265453</td>
<td>0.001573</td>
<td>0.133946</td>
<td>0.213840</td>
<td>0.317833</td>
<td>3.106893</td>
<td>182.375183</td>
</tr>
<tr>
<td>OLC Area (OLCs)</td>
<td>81.719943</td>
<td>240.960888</td>
<td>2</td>
<td>42</td>
<td>56</td>
<td>80</td>
<td>550</td>
<td>48380</td>
</tr>
</tbody>
</table>

(Anomalies highlighted in red color)
• Cluster at (7, 7)
  ▪ Area of 350–450 m²
• Extreme outliers
Extreme Outliers

- Lengthiest building footprint (north/south extent) is Aventura Mall
- Widest building footprint (east/west extent) is Homestead Miami Speedway
“Bad” and “Ugly” Building Footprints

76QXXR66+QHP-9-8-11-10 (incorrect polygon)

76QXQV54+FPX-77-12-12-56 (non-contiguous polygon)

76QXRRV6+433-11-13-4-1 (self-intersecting polygon)

76QXRM76+WC3-27-41-92-51 (!?!?!)
Conclusions

• Successfully processed dataset for Miami-Dade County
  ▪ Found 6 invalid records
  ▪ Found 68 duplicate records

• Further statistical analysis needed
  ▪ Unusually high Chebyshev and Manhattan distances
  ▪ Unusually low rectangularity
  ▪ Unusually high percentage area increase
  ▪ Unusually high area
Thank you
Data Sharing Updates

Partner Round-Robin
What do we need?

- Technical Leads – who are the folks programming and supporting your database infrastructure?
- Two+ databases – where do you want to see UBIDs incorporated and matched to each other?

Process:

- Mark will Skype/WebEx/etc. in with your technical team to understand your database architecture
- Using the tooling developed at PNNL, UBIDs can be added into your existing systems. In the process, Mark can develop a replicable process for use by additional stakeholders in your organization.
Next Call: February 12th, 2019

Preview:
- Authoritative UBIDs – how do you produce a “UBID of Record”
- DC’s Newseum – condos and the museum in a shared space. How do UBIDs help clear up this relationship?
Thank You

Questions?
Harry.Bergmann@ee.doe.gov
Nora.Wang@pnnl.gov
Mark.Borkum@pnnl.gov