BEDA Accelerator Monthly Working Call

Nora Wang, Mark Borkum, Kevin Keene
Overview

• UBID Data Flow
• UBID FAQ
Common Question

How do I assign UBIDs to my data?

The answer depends upon the type of the data.
UBID Data Flow

Choose the entry point that matches your data.

https://github.com/pnnl/buildingid/blob/master/DATA_FLOW.md
1. Street Address

24 Willie Mays Plaza
San Francisco, CA 94107

37.7784744216144 N, 122.3896056846471 W

[Image of the process: Geolocate -> Encode]

https://buildingid.github.io/examples/encode-example.html
2. Lat/Long Coords (Centroid only)

24 Willie Mays Plaza
San Francisco, CA 94107

Geolocate

37.7784744216144 N, 122.3896056846471 W

Encode

https://buildingid.github.io/examples/encode-example.html
3. Lat/Long Coords (Centroid+Box)

- **37.777404202489 N, 122.391163268176 W**
- **37.779726076137 N, 122.387805614265 W**
- **37.7784744216144 N, 122.3896056846471 W**

![Encode Example](https://buildingid.github.io/examples/encode-example.html)
4. Geometry

Calculate

Encode

Encode

Encode

37.77404202489 N, 122.391163288176 W

37.779726076137 N, 122.38705614265 W

37.778474216144 N, 122.389605614265 W

https://buildingid.github.io/examples/encode-example.html
4. Geometry (cont.)

Geometry

Encode

https://buildingid.github.io/examples/well-known-text-read-example.html
4. Geometry (cont.)

Shapefile, GeoJSON, etc.

Geometry

Extract

Encode

Well Known Text Read Example

The following code example takes a Well Known Text string, reads it into a shape, and encodes it into a code string using the Well Known Text Module, then adds it to the map.


Code Length

11

https://buildingid.github.io/examples/well-known-text-read-example.html
5. UBID Code String

849VQJH6+95J-51-58-42-50

Decode

https://buildingid.github.io/examples/decode-example.html
• What is a UBID?
• What are the UBIDs for a given spatial entity?
• How should I choose an OLC grid resolution?
• How are UBIDs used for cross-reference?
• How are UBIDs used for duplicate detection?

https://github.com/pnnl/buildingid/blob/master/FAQ.md
What is a UBID?

• Transparent string (no hidden information)

• Encoding of 6x latitude and longitude coordinates:
  ▪ Centroid
  ▪ Axis-aligned, minimum bounding box

• Format:
  ▪ C-N-E-S-W
    • “C” is the Open Location Code (OLC) for the centroid
    • “N”, “E”, “S”, “W” are the cardinal extents of the
      axis-aligned, minimum bounding box

UBIDs identify regions of the surface of the Earth that are occupied by a spatial entity (not the spatial entity itself).
What are the UBIDs for a given spatial entity?

- There is 1 UBID per OLC grid resolution level
- Higher resolution levels identify surface regions with greater accuracy

<table>
<thead>
<tr>
<th>OLC Resolution</th>
<th>OLC Length</th>
<th>OLC Width</th>
<th>UBID</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>180°</td>
<td>360°</td>
<td>00000000+0-0-0-0</td>
</tr>
<tr>
<td>2</td>
<td>20°</td>
<td>20°</td>
<td>84000000+0-0-0-0</td>
</tr>
<tr>
<td>4</td>
<td>1°</td>
<td>1°</td>
<td>849V0000+0-0-0-0</td>
</tr>
<tr>
<td>6</td>
<td>0.05°</td>
<td>0.05°</td>
<td>849VQJ00+0-0-0-0</td>
</tr>
<tr>
<td>8</td>
<td>0.0025°</td>
<td>0.0025°</td>
<td>849VQJH6+0-0-1-1</td>
</tr>
<tr>
<td>10</td>
<td>0.000125°</td>
<td>0.000125°</td>
<td>849VQJH6+97-10-12-8-15</td>
</tr>
<tr>
<td>11</td>
<td>0.000025°</td>
<td>0.00003125°</td>
<td>849VQJH6+97C-52-50-41-58</td>
</tr>
<tr>
<td>12</td>
<td>0.000005°</td>
<td>0.000078125°</td>
<td>849VQJH6+97CV-256-199-209-230</td>
</tr>
<tr>
<td>13</td>
<td>0.000001°</td>
<td>0.00001953125°</td>
<td>849VQJH6+97CVG-1279-797-1043-922</td>
</tr>
</tbody>
</table>
How should I select an OLC grid resolution?

- Consider the spatial entities whose footprints are being identified
- Select the highest OLC grid resolution where a single grid cell is meaningful

- Example – Identify space occupied by cars
  - Average size = 2m x 5m
  - OLC grid resolution 11 = 2.5m x 3.5m

When the correct OLC grid resolution is selected, small perturbations in the source geometry do not significantly affect the encoded UBID.
How are UBIDs used for cross-reference?

• Start with 2 UBID-assigned datasets (called “left” and “right”)
• For each pair of “left” and “right” records, test the UBIDs:
  ▪ If the test passes, then the pair is a match
  ▪ Otherwise, continue

• Tests include:
  ▪ String equality
  ▪ Distance between centroids
  ▪ Intersection over union (“IoU”) of bounding boxes

\[
\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}
\]
How are UBIDs used for cross-reference search?

• Search is cross-reference!

• The “left” dataset contains exactly 1 record (the needle)
• The “right” dataset contains any number of records (the haystack)

If your software system implements cross-reference, then you get search capabilities for free!
How are UBIDs used for duplicate detection?

- Suppose that we have assigned UBIDs to every record in a given dataset.

- 2 records are duplicates if:
  - They have identical UBIDs
  - They have similar UBIDs (e.g., significant IoU score)

- Hence, duplicate detection is cross-referencing a dataset with itself.
Questions?

• Please contact us with your questions
• With your permission, we will add the answers to the UBID FAQ website
Thank you