



BEDA Accelerator Monthly Working Call

Nora Wang, Mark Borkum, Kevin Keene



PNNL is operated by Battelle for the U.S. Department of Energy





Pacific
Northwest
NATIONAL LABORATORY

Overview

- UBID Data Flow
- UBID FAQ

Common Question

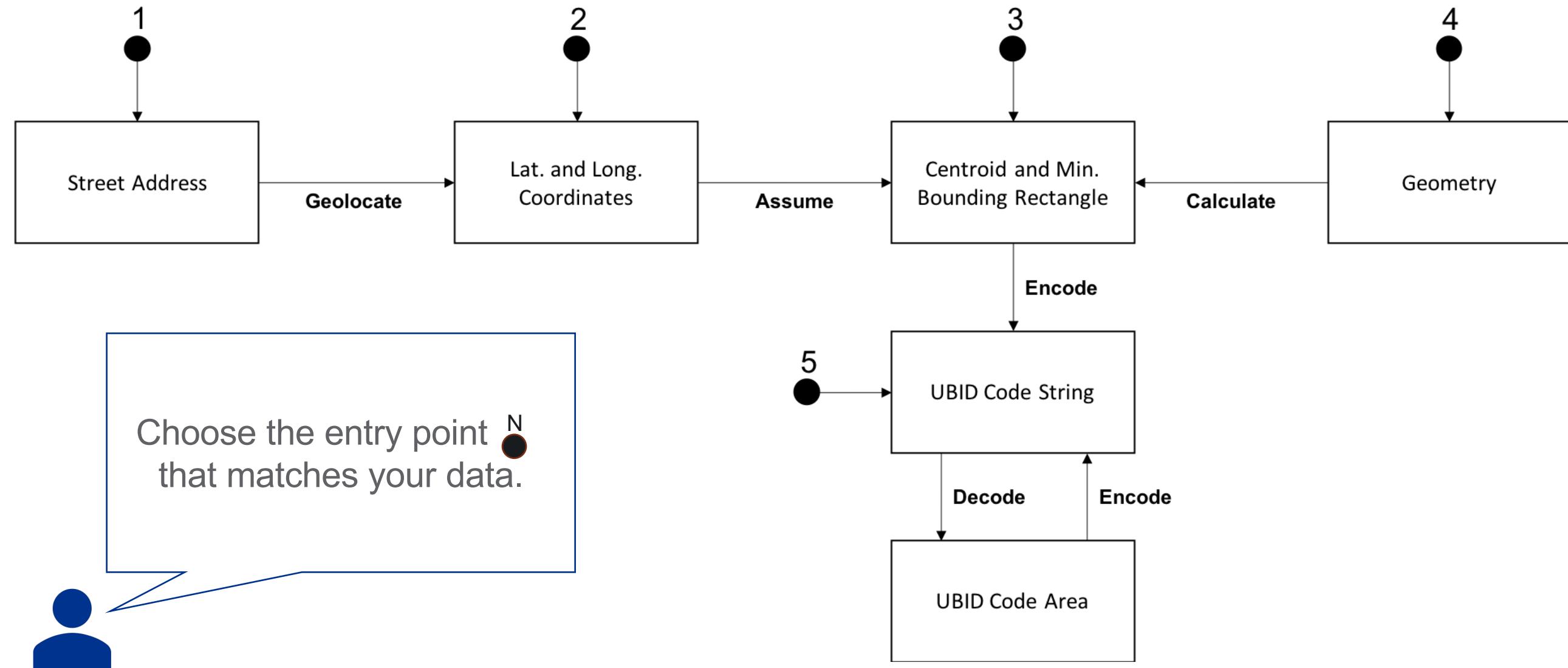
How do I assign UBIDs
to my data?



The answer depends upon
the type of the data.



UBID Data Flow



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

1. Street Address

24 Willie Mays Plaza
San Francisco, CA 94107

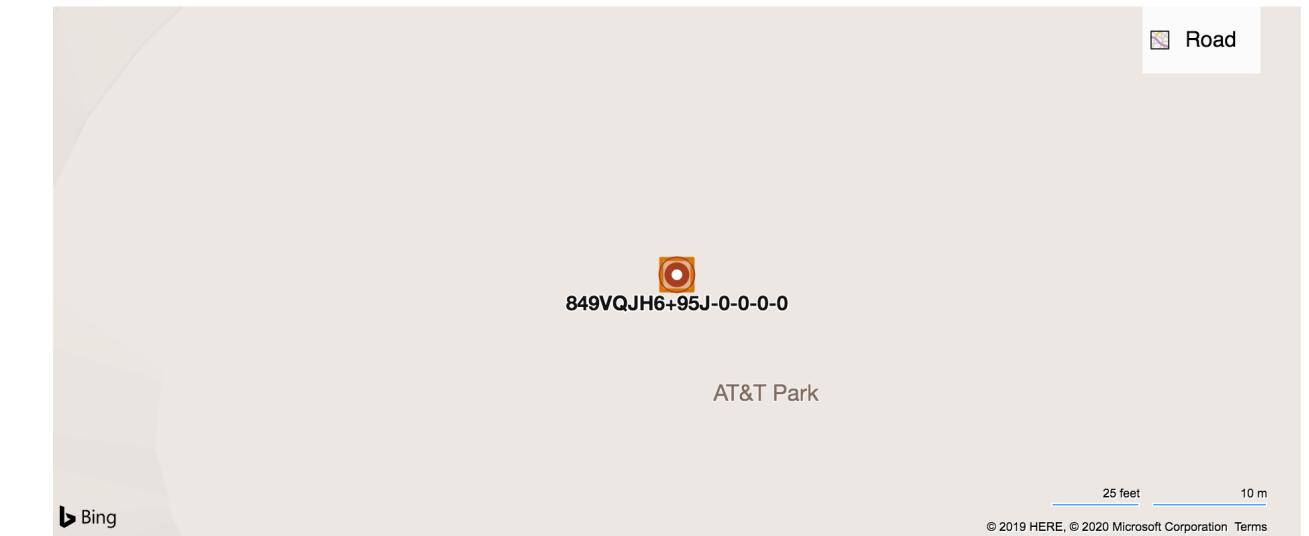
Geolocate

37.7784744216144 N,
122.3896056846471 W

Encode

Encode Example

The following example takes six coordinates and encodes them as a code string, decodes the code string into a shape using the [GeoJSON Module](#), then adds it to the map.



Bounding Box

Min. Latitude

37.7784744216144

Leave blank to use Centroid Latitude.

Min. Longitude

-122.3896056846471

Leave blank to use Centroid Longitude.

Max. Latitude

37.7784744216144

Leave blank to use Centroid Latitude.

Max. Longitude

-122.3896056846471

Leave blank to use Centroid Longitude.

Centroid

Latitude

37.7784744216144

Longitude

-122.3896056846471

Code

Code Length

11



2. Lat/Long Coords (Centroid only)

24 Willie Mays Plaza
San Francisco, CA 94107

Geolocate

37.7784744216144 N,
122.3896056846471 W

Encode

UBID Home Apps Examples

Encode Example

The following example takes six coordinates and encodes them as a code string, decodes the code string into a shape using the [GeoJSON Module](#), then adds it to the map.

A screenshot of the UBID Encode Example interface. At the top, there's a navigation bar with links for UBID, Home, Apps, and Examples. Below the navigation is a section titled "Encode Example" with a descriptive text about encoding coordinates and decoding them into a shape using the GeoJSON Module. A map shows the location of AT&T Park in San Francisco. A red marker is placed on the map, and its coordinates are displayed as "849VQJH6+95J-0-0-0-0". The map includes a legend for "Road", a scale bar from "25 feet" to "10 m", and a copyright notice for "© 2019 HERE, © 2020 Microsoft Corporation Terms".

Bounding Box

Min. Latitude

37.7784744216144

Min. Longitude

-122.3896056846471

Leave blank to use Centroid Latitude.

Max. Latitude

37.7784744216144

Max. Longitude

-122.3896056846471

Leave blank to use Centroid Latitude.

Centroid

Latitude

37.7784744216144

Longitude

-122.3896056846471

Code

Code Length

11

<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

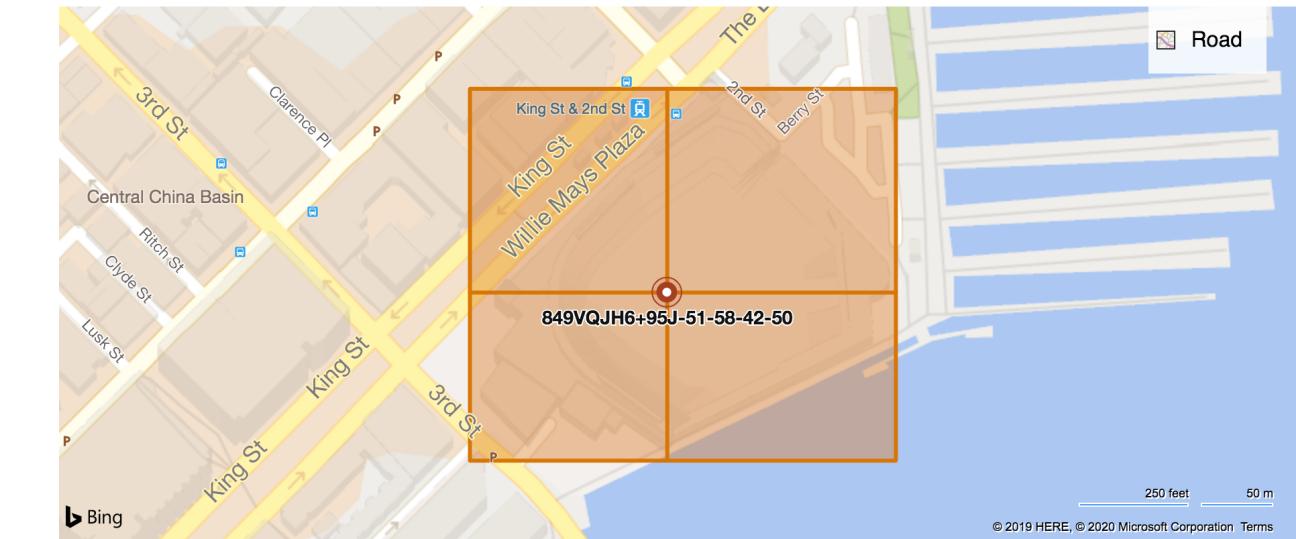
Encode

Encode

Encode

Encode Example

The following example takes six coordinates and encodes them as a code string, decodes the code string into a shape using the [GeoJSON Module](#), then adds it to the map.



Bounding Box

Min. Latitude

37.777404202489

Leave blank to use Centroid Latitude.

Min. Longitude

-122.391163268176

Leave blank to use Centroid Longitude.

Max. Latitude

37.779726076137

Leave blank to use Centroid Latitude.

Max. Longitude

-122.387805614265

Leave blank to use Centroid Longitude.

Centroid

Latitude

37.7784744216144

Longitude

-122.3896056846471

Code

Code Length

11

4. Geometry

Geometry



Calculate

37.777404202489 N,
122.391163268176 W

Encode

37.779726076137 N,
122.387805614265 W

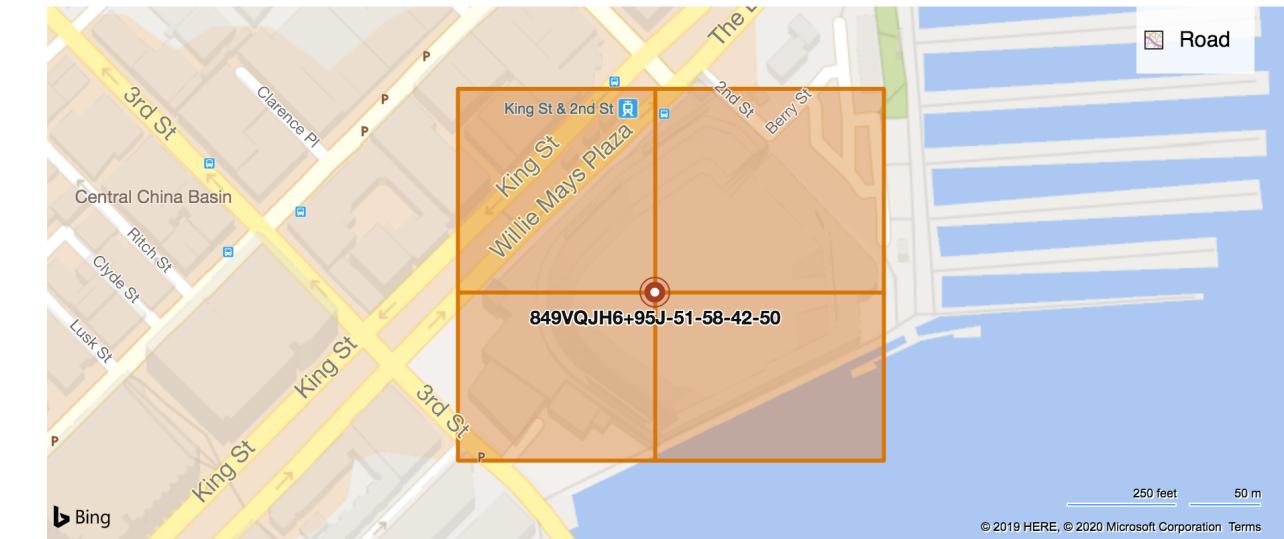
Encode

37.7784744216144 N,
122.3896056846471 W

Encode

Encode Example

The following example takes six coordinates and encodes them as a code string, decodes the code string into a shape using the [GeoJSON Module](#), then adds it to the map.



Bounding Box

Min. Latitude

37.777404202489

Min. Longitude

-122.391163268176

Leave blank to use Centroid Latitude.

Max. Latitude

37.779726076137

Max. Longitude

-122.387805614265

Leave blank to use Centroid Latitude.

Centroid

Latitude

37.7784744216144

Longitude

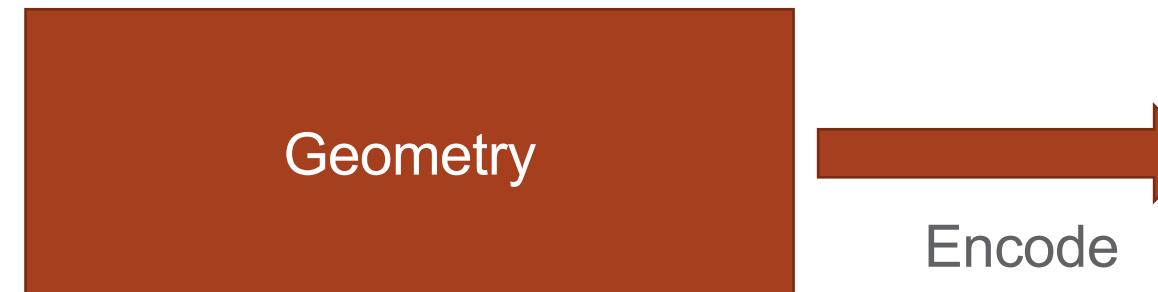
-122.3896056846471

Code

Code Length

11

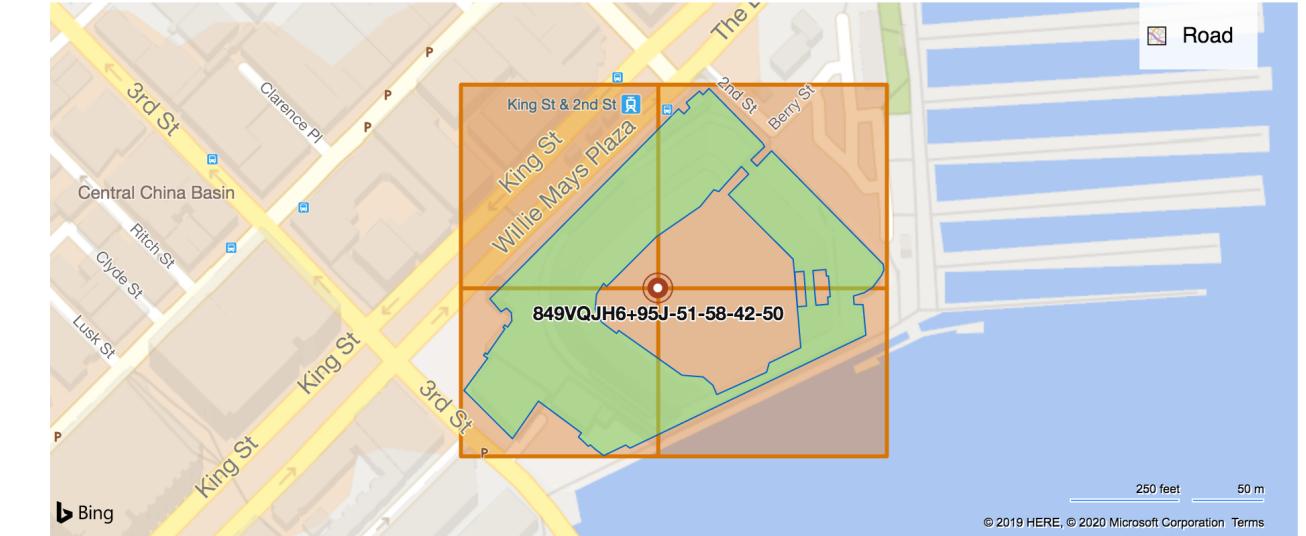
4. Geometry (cont.)



UBID Home Apps Examples

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.



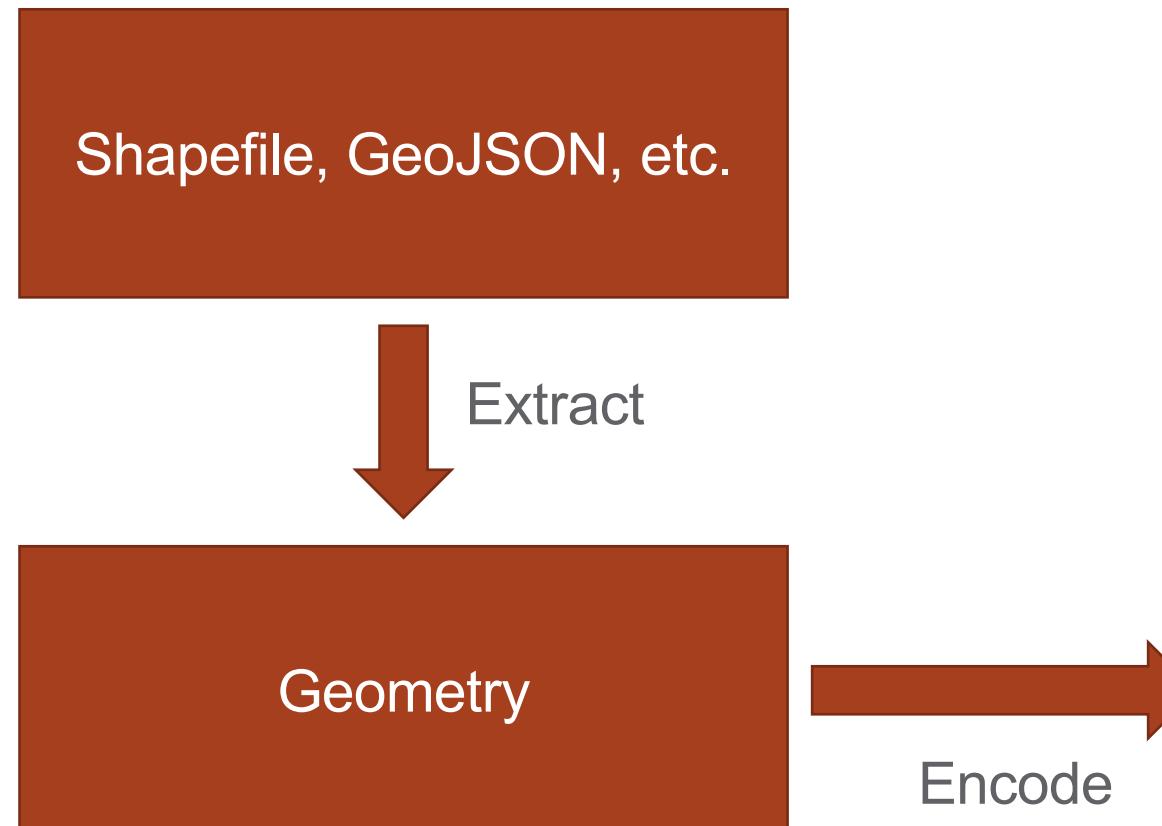
Well Known Text

```
MULTIPOLYGON (((-122.38881976231 37.77923523529, -122.388804119691 37.779248288341, -122.388804089115 37.779249576693, -122.388736336901 37.779305686423, -122.388735345943 37.779305673294, -122.388703640352 37.779332128723, -122.387810661362 37.778614548526, -122.387814609477 37.778585892065, -122.387811140111 37.778585064958, -122.387805638074 37.778578430108, -122.387805614265 37.778578395859, -122.387816682916 37.77854710155, -122.387824621151 37.778543969508, -122.387829917765 37.778545142646, -122.387843084163 37.778523260839, -122.387896335497 37.778468172867,
```

Code Length

<https://buildingid.github.io/examples/well-known-text-read-example.html>

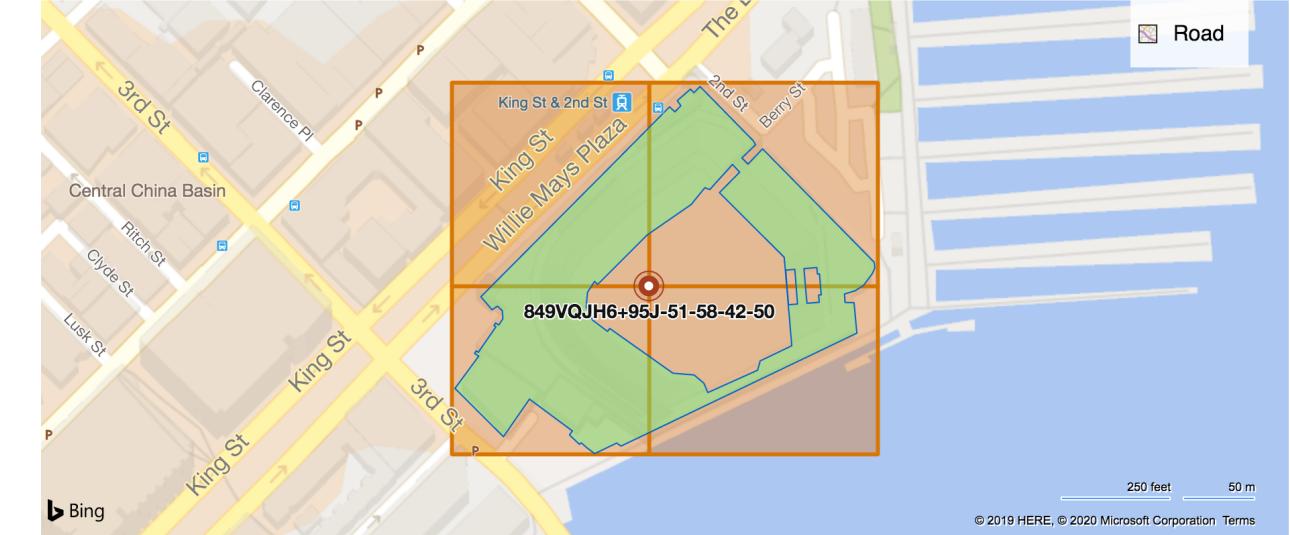
4. Geometry (cont.)



UBID Home Apps Examples

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.



Well Known Text

```
MULTIPOLYGON (((-122.38881976231 37.77923523529, -122.388804119691 37.779248288341, -122.388804089115 37.779249576693, -122.388736336901 37.779305686423, -122.388735345943 37.779305673294, -122.388703640352 37.779332128723, -122.387810661362 37.778614548526, -122.387814609477 37.778585892065, -122.387811140111 37.778585064958, -122.387805638074 37.778578430108, -122.387805614265 37.778578395859, -122.387816682916 37.77854710155, -122.387824621151 37.778543969508, -122.387829917765 37.778545142646, -122.387843084163 37.778523260839, -122.387896335497 37.778468172867,
```

Code Length

<https://buildingid.github.io/examples/well-known-text-read-example.html>

5. UBID Code String

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

Decode

UBID Home Apps Examples

Decode Example

The following example takes a code string and decodes it into a shape using the [GeoJSON Module](#), then adds it to the map.



Code

849VQJH6+95J-51-58-42-50 ✓

▼ Code Area

Bounding Box

Min. Latitude	Max. Latitude	Min. Longitude	Max. Longitude
37.777399999999965	37.77975000000004	-122.39118750000102	-122.38778124999878

Centroid

Min. Latitude	Max. Latitude	Min. Longitude	Max. Longitude
37.77845	37.778475	-122.389625	-122.3895937499997

<https://buildingid.github.io/examples/decode-example.html>

UBID FAQ

- 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

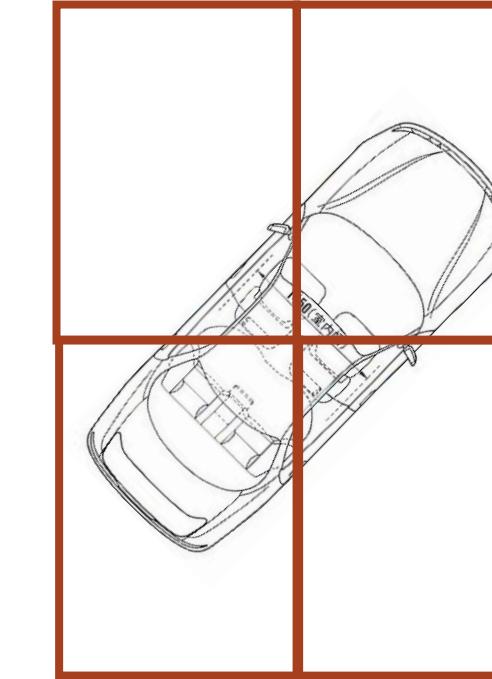
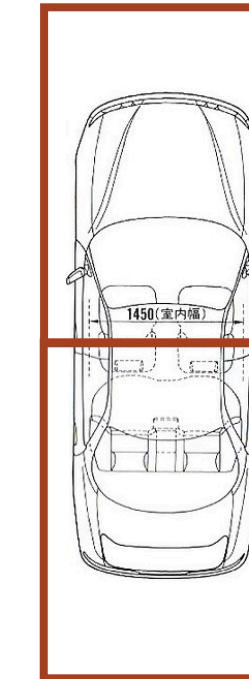
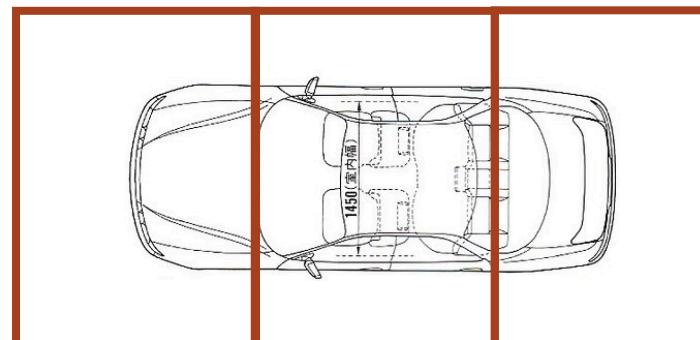


AT&T Park, San Francisco, CA

OLC Resolution	OLC Length	OLC Width	UBID
0	180°	360°	0000000+-0-0-0-0
2	20°	20°	84000000+-0-0-0-0
4	1°	1°	849V0000+-0-0-0-0
6	0.05°	0.05°	849VQJ00+-0-0-0-0
8	0.0025°	0.0025°	849VQJH6+-0-0-1-1
10	0.000125°	0.000125°	849VQJH6+97-10-12-8-15
11	0.000025°	0.00003125°	849VQJH6+97C-52-50-41-58
12	0.000005°	0.0000078125°	849VQJH6+97CV-256-199-209-230
13	0.000001°	0.000001953125°	849VQJH6+97CVG-1279-797-1043-922

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

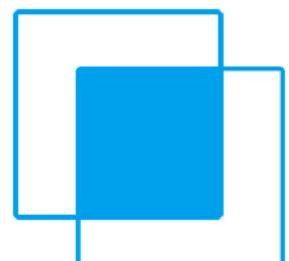


When the correct OLC grid resolution is selected, small perturbations in the source geometry do not significantly affect the encoded UBID.



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-referencing~~ 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



Pacific
Northwest
NATIONAL LABORATORY

Thank you

