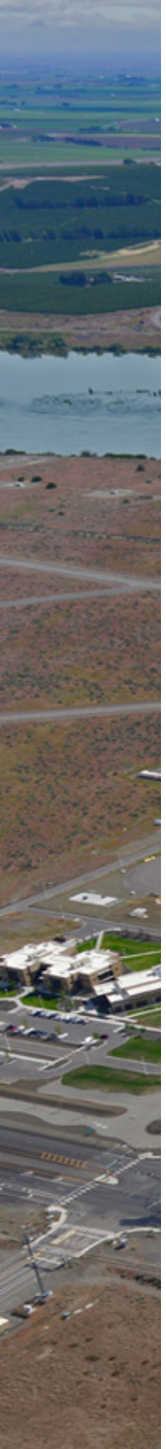
Pacific Northwest
NATIONAL LABORATORY

Miami-Dade Microsoft Cross Reference Analysis

Kevin Keene

U.S. DEPARTMENT OF ENERGY BATTELLE

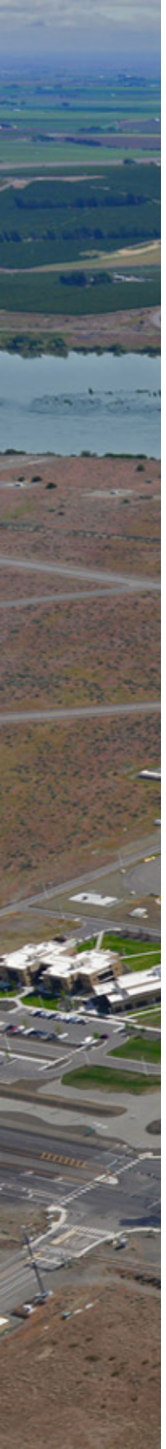
PNNL is operated by Battelle for the U.S. Department of Energy



Pacific Northwest
NATIONAL LABORATORY

Objectives

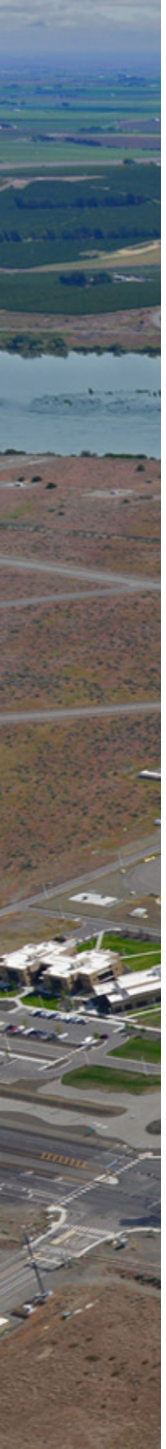
1. Compare Microsoft footprints to MDC footprints in GIS to see how similar the two datasets are
2. Investigate UBID one-to-one matching between two building footprint datasets (MS and MDC) and compare to GIS matching



Pacific Northwest
NATIONAL LABORATORY

Dataset Background

- Miami-Dade Open Footprints
  - Planimetric layer updated in 2012
  - LiDAR features updated in 2015
  - https://gis-mdc.opendata.arcgis.com/datasets/d511e9ebc5aa4f49a23ff5fa2fb99786_0
- Microsoft Footprints
  - Nation-wide open source building footprints from satellite data with geometric screening algorithms
  - From 2017
  - https://www.arcgis.com/home/item.html?id=f40326b0dea54330ae3958401280712 6
  - https://github.com/Microsoft/USBuildingFootprints

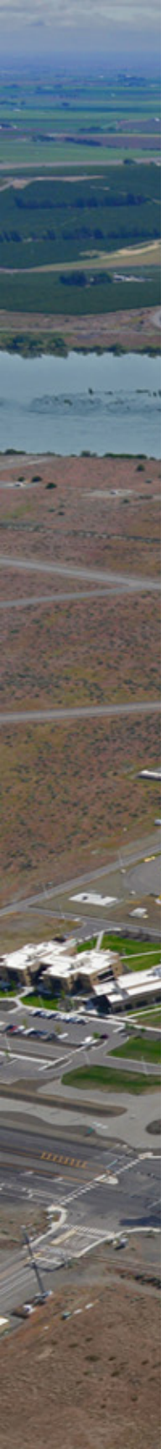


Pacific Northwest
NATIONAL LABORATORY

Comparing in GIS

→ Purpose: see the spatial relationship between the footprints in the two datasets

	Miami-Dade	Microsoft	Note
A. Buildings with 0 intersections	75,730	7,089	• Includes intersections with slight overlaps (5%) • Miami has many extra buildings that don't match to any in the MS dataset
B. Buildings with one-to-one	468,716	468,716	• Ideally, all buildings are in this category • Doesn't mean they are equivalent necessarily
C. Buildings with one-to-many or many-to-one	21,555	12,714	• These likely have multiple buildings in one dataset to represent on building in the other
Total	**566,001**	**488,519**	



Pacific Northwest
NATIONAL LABORATORY

ArcMap GIS Matching

GOAL:

- Create the “ground truth” (or as best possible) of what buildings are considered “equivalent” by setting an intersection threshold

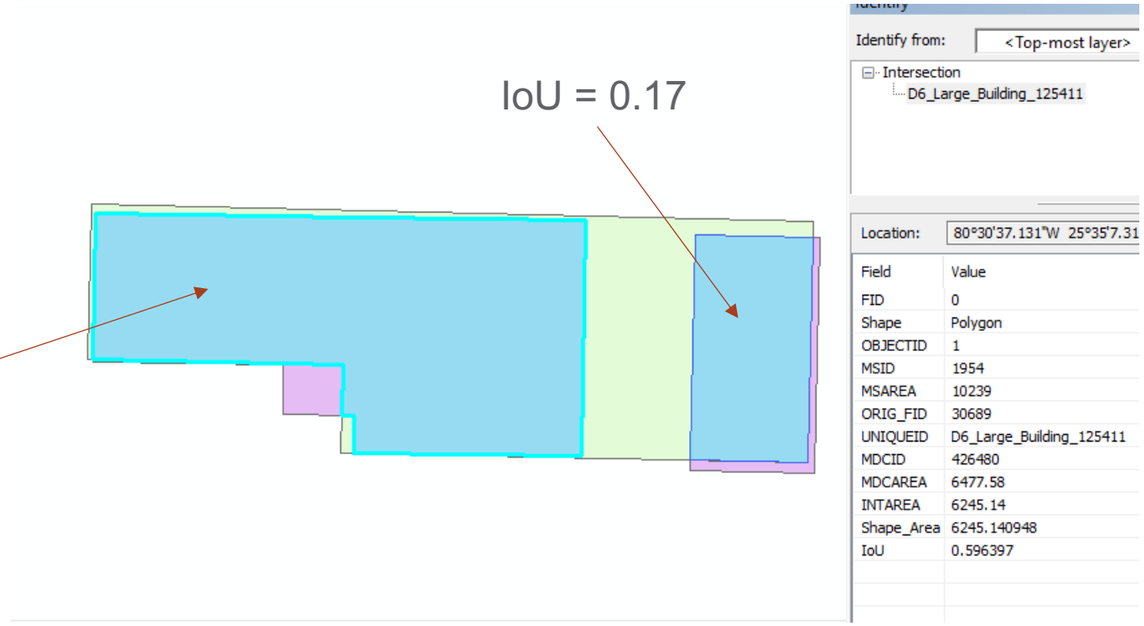
Process

1. Find the intersections between the two datasets and calculate Intersection over Union (IoU)
   - IoU = Intersect Area / (Footprint Area 1 + Footprint Area 2 – Intersect Area)
2. Group by matches with same ID and delete multiple matches to so that only one-to-one matches remain (with highest IoU)
3. Only keep over certain threshold of IoU (see next slides)

Investigating IoU GIS Threshold

What should be considered equivalent buildings?

IoU = 0.59



Microsoft

Miami

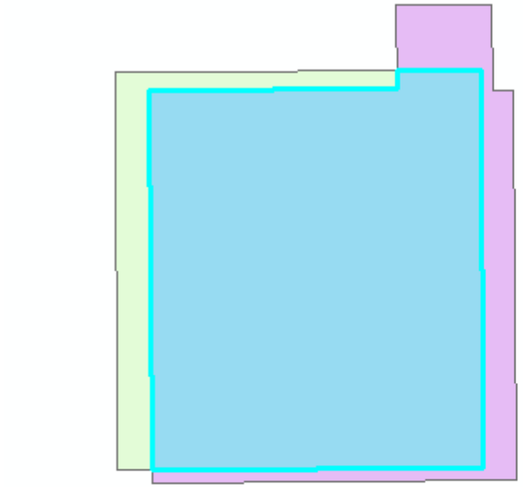
Intersection

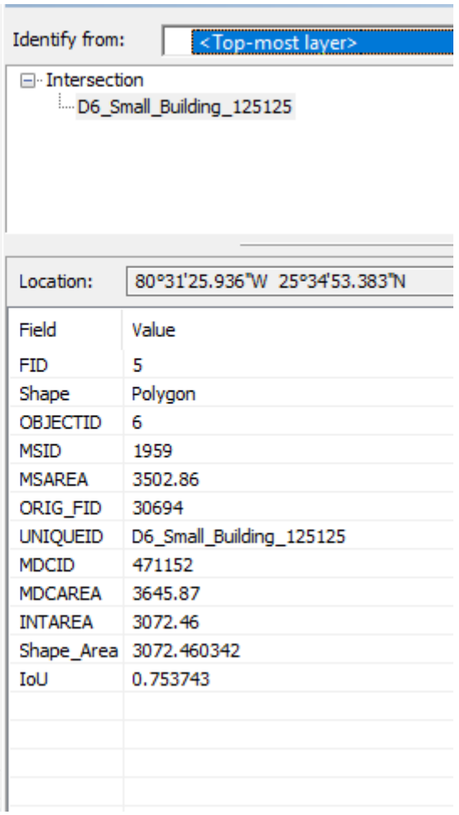
19355 Southwest 197th Avenue, Miami, FL

Investigating IoU GIS Threshold

What should be considered equivalent buildings?

IoU = 0.75





Field	Value
FID	5
Shape	Polygon
OBJECTID	6
MSID	1959
MSAREA	3502.86
ORIG_FID	30694
UNIQUEID	D6_Small_Building_125125
MDCID	471152
MDCAREA	3645.87
INTAREA	3072.46
Shape_Area	3072.460342
IoU	0.753743

Microsoft

Miami

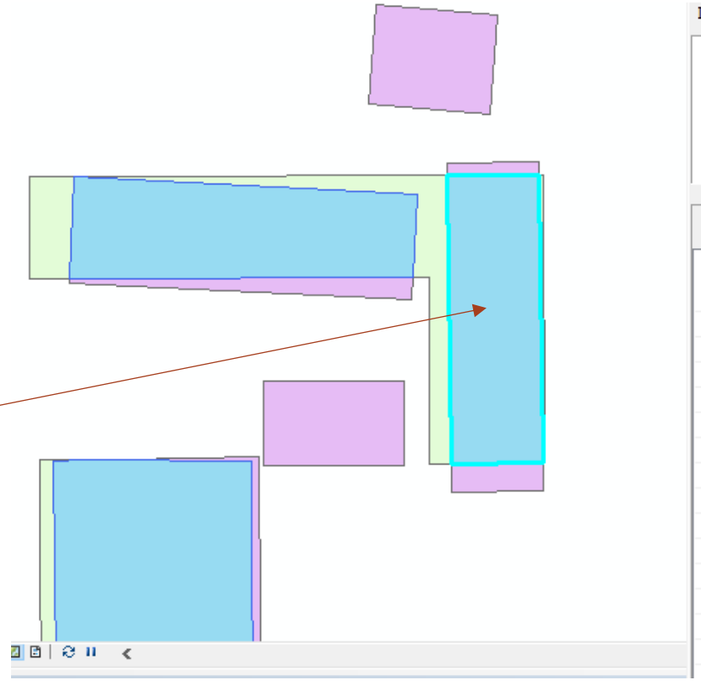
Intersection

20441 Southwest 198th Street, Miami, FL

Investigating IoU GIS Threshold

What should be considered equivalent buildings?

IoU = 0.34



Identify from: <Top-most layer>

Intersection
D6_MDC_Building_13355

Location: 80°29'49.488"W 25°28'35.298"N

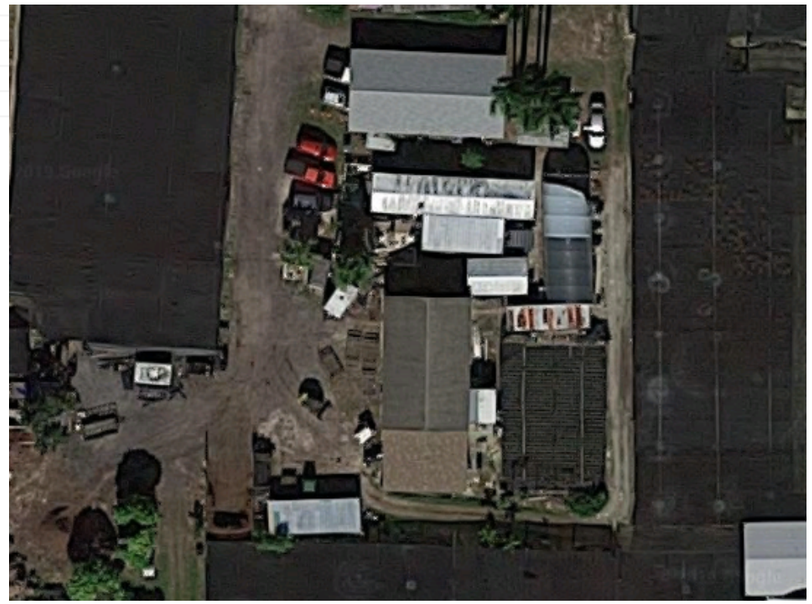
Field	Value
FID	865
Shape	Polygon
OBJECTID	866
MSID	70157
MSAREA	1579.45
ORIG_FID	962690
UNIQUEID	D6_MDC_Building_13355
MDCID	432977
MDCAREA	645.253
INTAREA	566.211
Shape_Area	566.211495
IoU	0.341401

Microsoft

Miami

Intersection

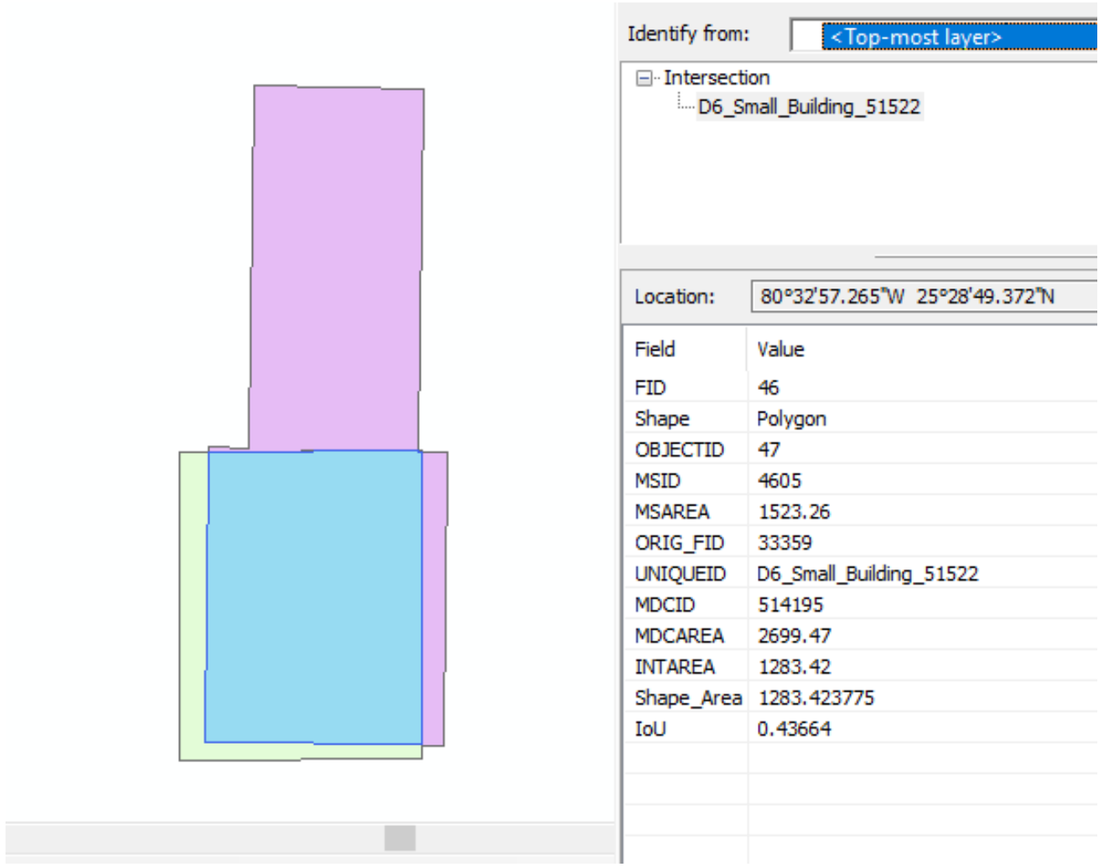
1708 Northwest 8th Street,
Homestead, Miami-Dade, FL



Investigating IoU GIS Threshold

What should be considered equivalent buildings?

IoU = 0.43



Microsoft

Miami

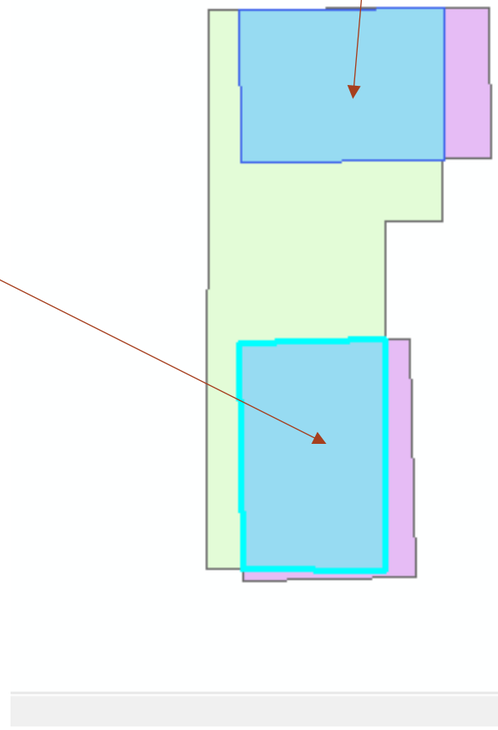
Intersection

Investigating IoU GIS Threshold

What should be considered equivalent buildings?

IoU = 0.26

IoU = 0.27

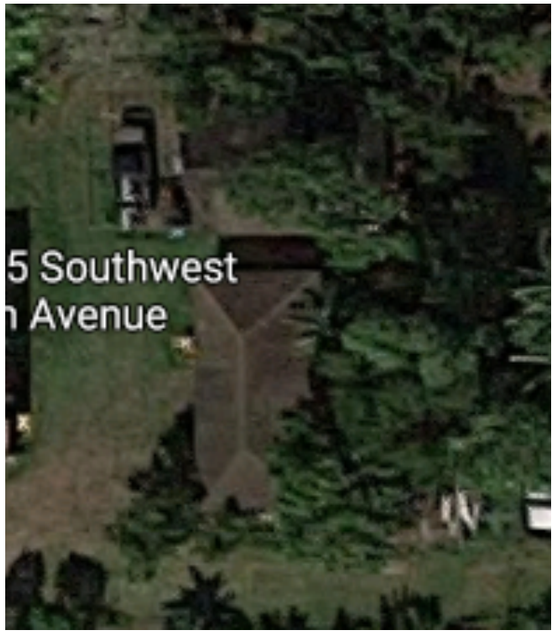


Identify from: <Top-most layer>

- Intersection
  - D6_Small_Building_96232

Location: 80°31'35.031"W 25°35'4.13"N

Field	Value
FID	731
Shape	Polygon
OBJECTID	732
MSID	57985
MSAREA	1839.19
ORIG_FID	827613
UNIQUEID	D6_Small_Building_96232
MDCID	561841
MDCAREA	673.245
INTAREA	543.625
Shape_Area	543.625259
IoU	0.276119



Microsoft

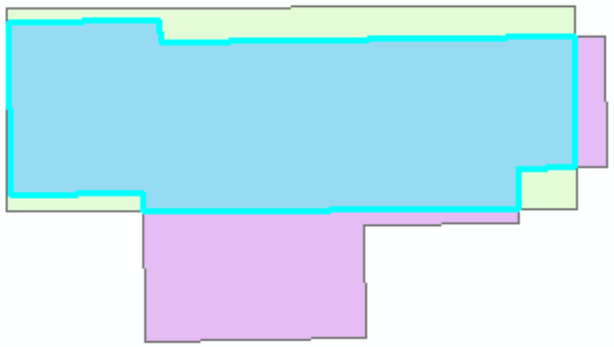
Miami

Intersection

Investigating IoU GIS Threshold

What should be considered equivalent buildings?

IoU = 0.63



Location: 80°30'26.302"W 25°28'49.29"N

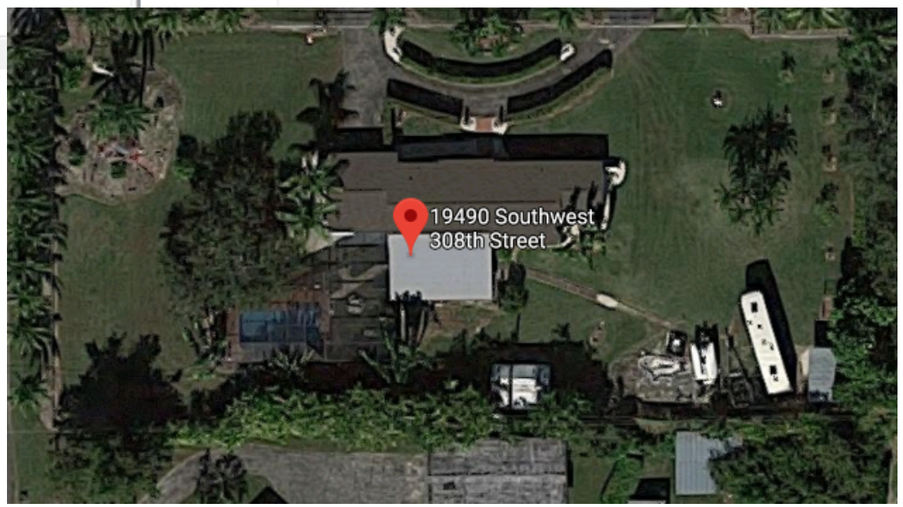
Field	Value
FID	218
Shape	Polygon
OBJECTID	219
MSID	18533
MSAREA	3465.72
ORIG_FID	232752
UNIQUEID	D6_Small_Building_49829
MDCID	512383
MDCAREA	3861.48
INTAREA	2836.01
Shape_Area	2836.011328
IoU	0.631461

Microsoft

Miami

Intersection

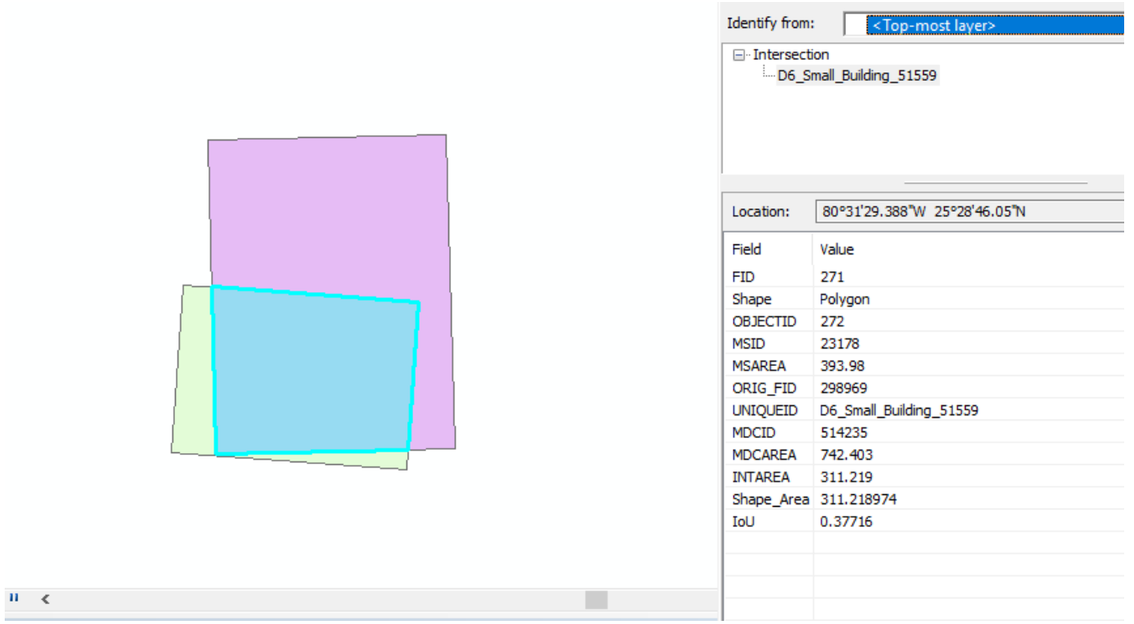
19490 Southwest 308th Street, Homestead, FL

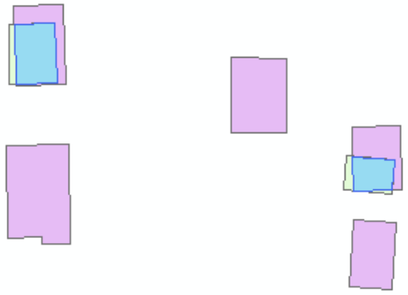


Investigating IoU GIS Threshold

What should be considered equivalent buildings?

IoU = 0.37

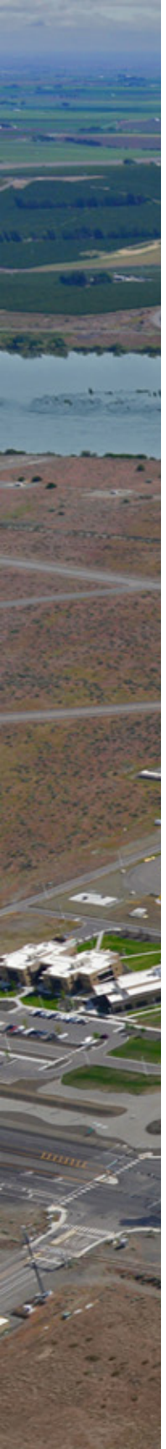




Microsoft

Miami

Intersection



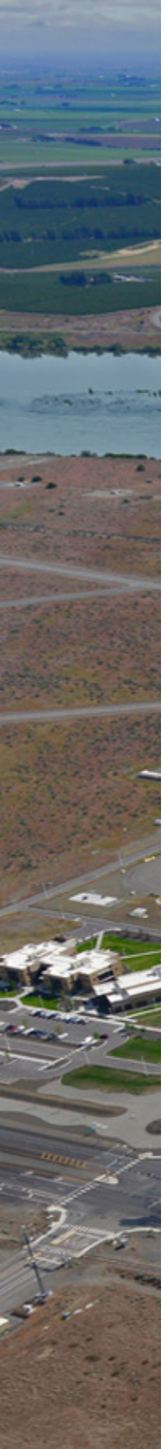
Pacific Northwest NATIONAL LABORATORY

Choosing GIS Threshold

- Somewhere in the 0.3 – 0.5 range seems logical
  - Only 8.1k buildings in this range
  - Choose **0.4** as GIS threshold for this analysis as threshold of what is considered equivalent between two buildings

GIS IoU Threshold	Number of Intersects		GIS IoU Threshold	Number of Intersects
0	478,791		**0.5**	468,706
0.1	478,626		**0.6**	458,106
0.2	477,971		**0.7**	415,488
0.3	476,833		**0.8**	241,015
0.4	474,650		**0.9**	26,178

- Number of buildings matched = 474,650
  - This is **82%** of the buildings between the two datasets, which is an indicator of how similar they are based on the matching threshold chosen



Pacific Northwest NATIONAL LABORATORY

UBID Cross Reference

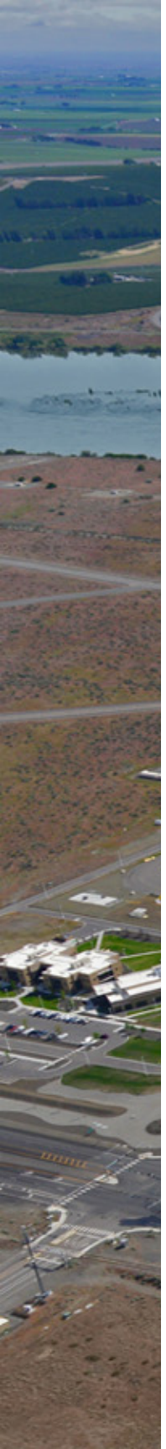
- Can find best match based on IoU, centroid distance, or other heuristics
- Success rate doesn't indicate what's correct, but how well UBID cross reference can replicate the results of GIS matching

Grouping based on IoU

UBID IoU Threshold	Same matches as GIS	Total matches found	Success Rate
0.0	473,786	480,060	98.7%
0.1	473,786	479,691	98.8%
0.2	473,778	479,188	98.9%
0.3	473,647	478,255	99.0%
0.4	**472,540**	**476,915**	**99.1%**
0.5	466,913	475,667	98.2%
0.6	448,775	475,003	94.5%
0.7	406,058	474,882	85.5%
0.8	303,254	474,802	63.9%
0.9	149,902	474,745	31.6%

Grouping based on Centroid distance

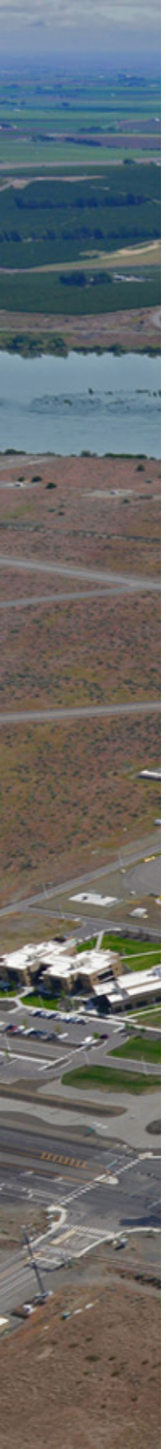
UBID IoU Threshold	Same matches as GIS	Total matches found	Success Rate
0.0	463,662	485,194	95.6%
0.1	463,662	480,871	96.4%
0.2	463,655	479,406	96.7%
0.3	463,545	478,160	96.9%
0.4	**462,630**	**476,730**	**97.0%**
0.5	457,715	475,505	96.3%
0.6	440,804	474,985	92.8%
0.7	400,081	474,871	84.3%
0.8	300,169	474,793	63.2%
0.9	149,654	474,742	31.5%



Pacific Northwest
NATIONAL LABORATORY

Conclusions

- UBID cross reference can achieve 99.1% correspondence to GIS cross reference using an IoU threshold of 0.4 for GIS polygons
  - If UBID cross reference can achieve similar results to GIS (which is the current best practice for spatial matching [w/o machine learning]), then UBID is a feasible mechanism for establishing equivalency between similar datasets
  - UBID has advantages like transcribability, natural key, universal coding/decoding, etc.
- Miami-Dade is mostly spread out since it includes the whole county – next step is to test methodology on dense urban centers



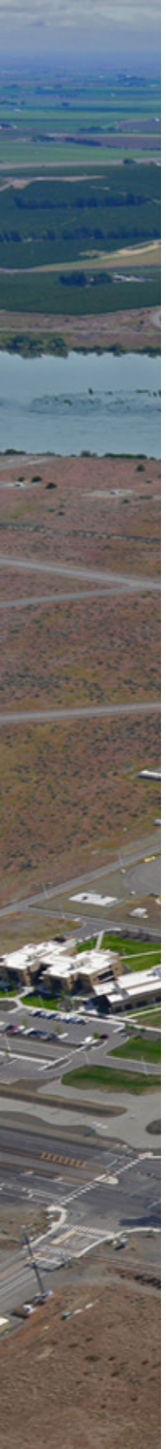
Pacific Northwest NATIONAL LABORATORY

Manhattan

- Microsoft and open city data do not align well
- Different definitions of what a building is

NYC Open Data Footprints

Microsoft Footprints

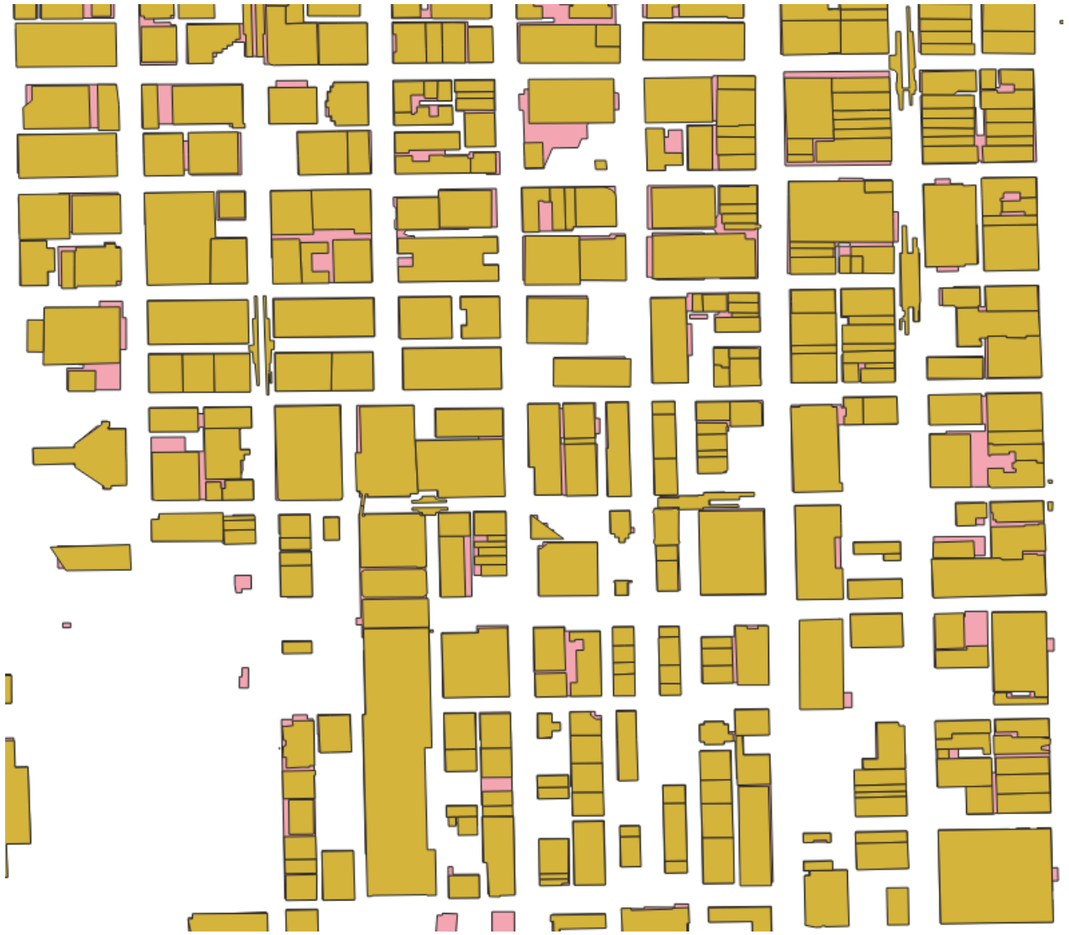


Pacific Northwest NATIONAL LABORATORY

Chicago

- Not as pronounced for downtown Chicago, but the difference between the datasets is still evident

Chicago Open Data Footprints



Microsoft Footprints

Pacific Northwest
NATIONAL LABORATORY

Thank you