

National Address Database Census Evaluation

FGDC Address Theme Subcommittee Meeting

February 14, 2018

Presented by

Matt Zimolzak, U.S. Census Bureau



Topics

DOT NAD Provider Submission Review - Summary Statistics

Census NAD Evaluation

Geocoding Comparison

Address Count Comparison and Analysis

NAD DOT State Submission Review: Quality Checks

Suite of Quality Checks for NAD Submissions

- Data Completeness Assessment
 - Required and supported NAD data variables - count and percentage populated
- Data Quality Assessment
 - Duplicate records
 - Missing required data variables
 - Missing required metadata
 - Invalid geometry
- Attribute Domain Quality Assessment
 - Invalid values
 - Non-documented values, proposed expansion

NAD DOT State Submission Review: Results Report

- Submission Review Results Report
 - Summary information, including accepted and flagged (rejected) address records
 - Detailed results accounting
 - Shared with provider
- QC process determines which records are included in the NAD
- Report is accounting of the QC process results

NAD DOT State Submission Review: Results Report Example

NAD Submission QA/QC Review

Colorado

DATA COMPLETENESS

* required

Flagged items are highlighted in

red

Flagged items not required to be populated at the state level are highlighted in

yellow

Total Feature Count		2,190,243		
SAUID	99.92%	2,188,597 populated,	1,646 empty/NULL	
PlaceID	68.80%	1,506,956 populated,	683,287 empty/NULL	
Longitude*	98.21%	2,151,068 populated,	39,175 empty/NULL	
Latitude*	98.21%	2,151,068 populated,	39,175 empty/NULL	
AddrNum*	99.02%	2,168,697 populated,	21,546 empty/NULL	
NumSuf	0.43%	9,385 populated,	2,180,858 empty/NULL	
PreDir	36.75%	804,813 populated,	1,385,430 empty/NULL	
PreType	2.21%	48,353 populated,	2,141,890 empty/NULL	
StreetName*	98.65%	2,160,783 populated,	29,460 empty/NULL	
PostType	87.69%	1,920,655 populated,	269,588 empty/NULL	
PostDir	0.64%	14,006 populated,	2,176,237 empty/NULL	
UnitType	9.21%	201,718 populated,	1,988,525 empty/NULL	
UnitNumber	13.56%	297,094 populated,	1,893,149 empty/NULL	
AddrFull	99.09%	2,170,296 populated,	19,947 empty/NULL	
PlaceName	67.59%	1,480,330 populated,	709,913 empty/NULL	
Zipcode*	93.52%	2,048,246 populated,	141,997 empty/NULL	
IsCAI	0.00%	0 populated,	2,190,243 empty/NULL	
ParcelID	6.63%	145,118 populated,	2,045,125 empty/NULL	
MOD_DATE*	80.08%	1,754,049 populated,	436,194 empty/NULL	
ACT_STAT	80.08%	1,754,049 populated,	436,194 empty/NULL	
PROC_STAT	80.08%	1,754,049 populated,	436,194 empty/NULL	
County*	100.00%	2,190,243 populated,	0 empty/NULL	

NAD DOT State Submission Review Summary Statistics

State Name	Records Added to NAD	Flagged Records	Accepted Pct.	Duplicates	Zero or Missing Address Number	Empty/Null Street Name	Invalid Geometry	Missing ZIP Code
Arkansas	1,473,345	21,204	98.6%	14,857	6,297	68	0	0
Colorado	2,054,701	135,542	93.8%	104,128	26,158	0	2,776	141,997 *
District of Columbia	367,082	7,197	98.1%	0	7,197	0	0	18
Indiana	2,987,387	46	100.0%	0	21	0	0	0
Massachusetts	3,426,884	17,482	99.5%	10,166	58	407	0	5,848
Missouri (11 Counties)	1,081,249	211,354	83.6%	0	211,354	100	0	12,698
Montana	515,518	6,437	98.8%	0	3,108	2,011	0	3,630
New Mexico	914,189	13,568	98.5%	0	2	1,080	0	0
New York #	5,241,010	1	100.0%	0	0	1	0	0
North Carolina	4,884,860	55,052	98.9%	0	22,445	23,289	0	4,648
Tennessee	3,355,566	9,038	99.7%	0	0	0	0	0
Utah	1,187,669	1	100.0%	0	0	1	1	0
Virginia	3,602,249	238	100.0%	0	0	0	0	3

* Records not rejected from NAD, as in other states

Excludes New York City Addresses

- Selected results.
- All summary statistics are for Beta NAD versions.
- Based on report versions shared with Census to date.

National Address Database: Census Evaluation

- Evaluation conducted on states from two different versions of NAD.
 - Pilot
 - Beta
- Used the most current version of a state available.
- Version based on what Census acquired from DOT to date.
- State NAD version is indicated in the geocoding analysis.
- State NAD version is omitted in Address Count Comparison and Analysis.

Geocoding Comparison

NAD County vs. Census County

Method

- NAD Minimum Content Guidelines include County designation.
 - County name is populated in accepted NAD records.
- Geocode NAD records to Census TIGER/Line files to determine Census County.
- Compare NAD County record counts to Census County record counts.
 - Red rows indicate Pilot NAD, Blue rows indicate Beta NAD.
 - Positive differences indicate more records in Census county.
 - Negative differences indicate more records in NAD county.
- Analyze and explain relevant differences.

Geocoding Comparison

NAD County vs. Census County Selected Results

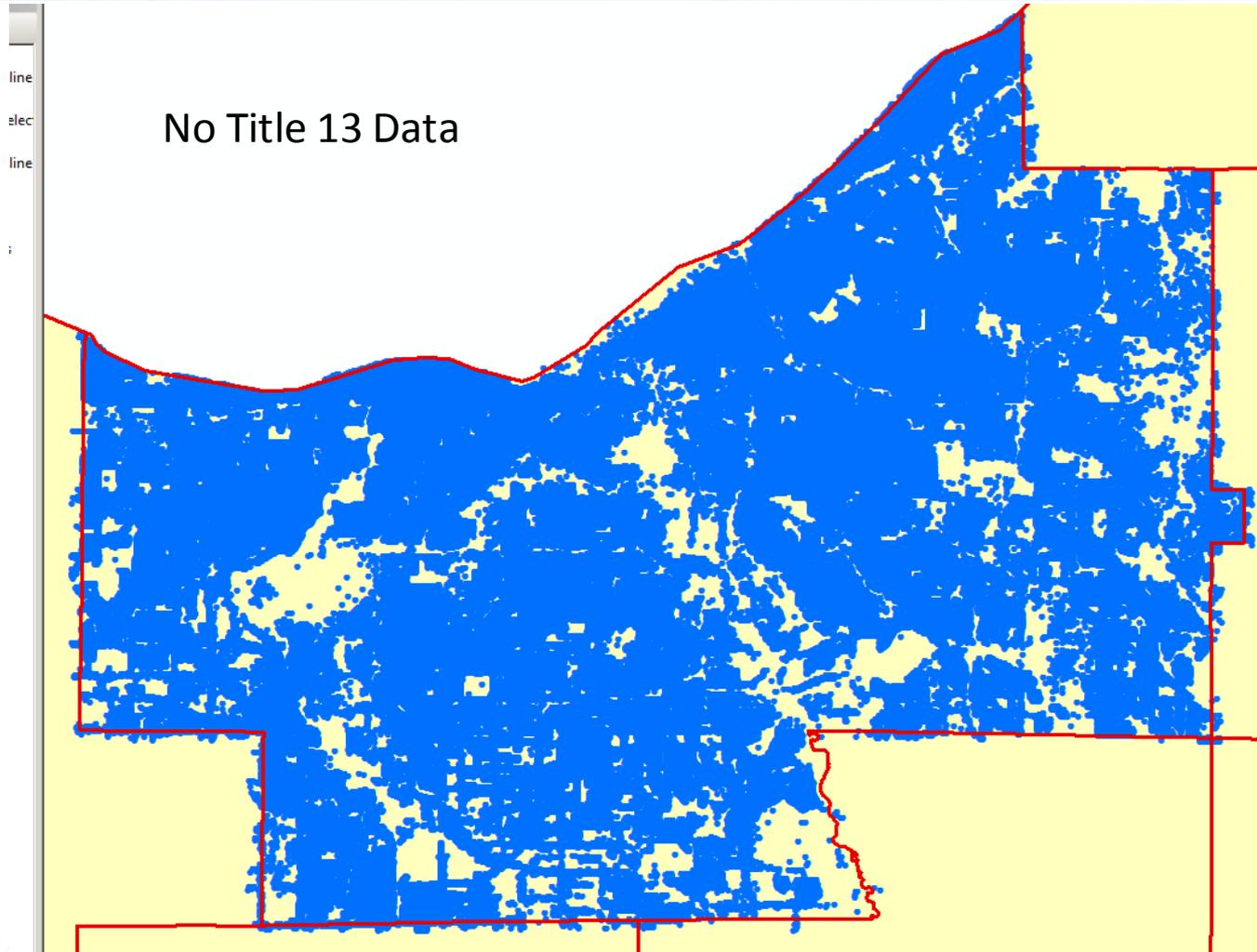
State	Counties	Count Based On TIGER/Line Geocoding	Count Based On NAD Attribute Geocode	Difference
Arizona	Yavapai	147722	147546	176
Arkansas	Monroe	16751	16772	-21
Arkansas	Arkansas	5705	5688	17
Arkansas	Benton	110243	110226	17
Indiana	Madison	58716	59531	-815
Indiana	Delaware	55612	54856	756
Indiana	Wells	12799	12338	461
Indiana	Allen	167598	168052	-454
Indiana	Benton	6001	5709	292
Indiana	Tippecanoe	61710	62002	-292
Missouri	Christian	37529	38255	-726
Missouri	Camden	97061	97606	-545
Missouri	Stone	30830	30373	457
Missouri	Cole	43491	43778	-287
Montana	Lake	16552	16544	8
Montana	Missoula	46060	46067	-7
New Jersey	Morris	202230	202160	70
New Jersey	Essex	223306	223316	-10

Geocoding Comparison

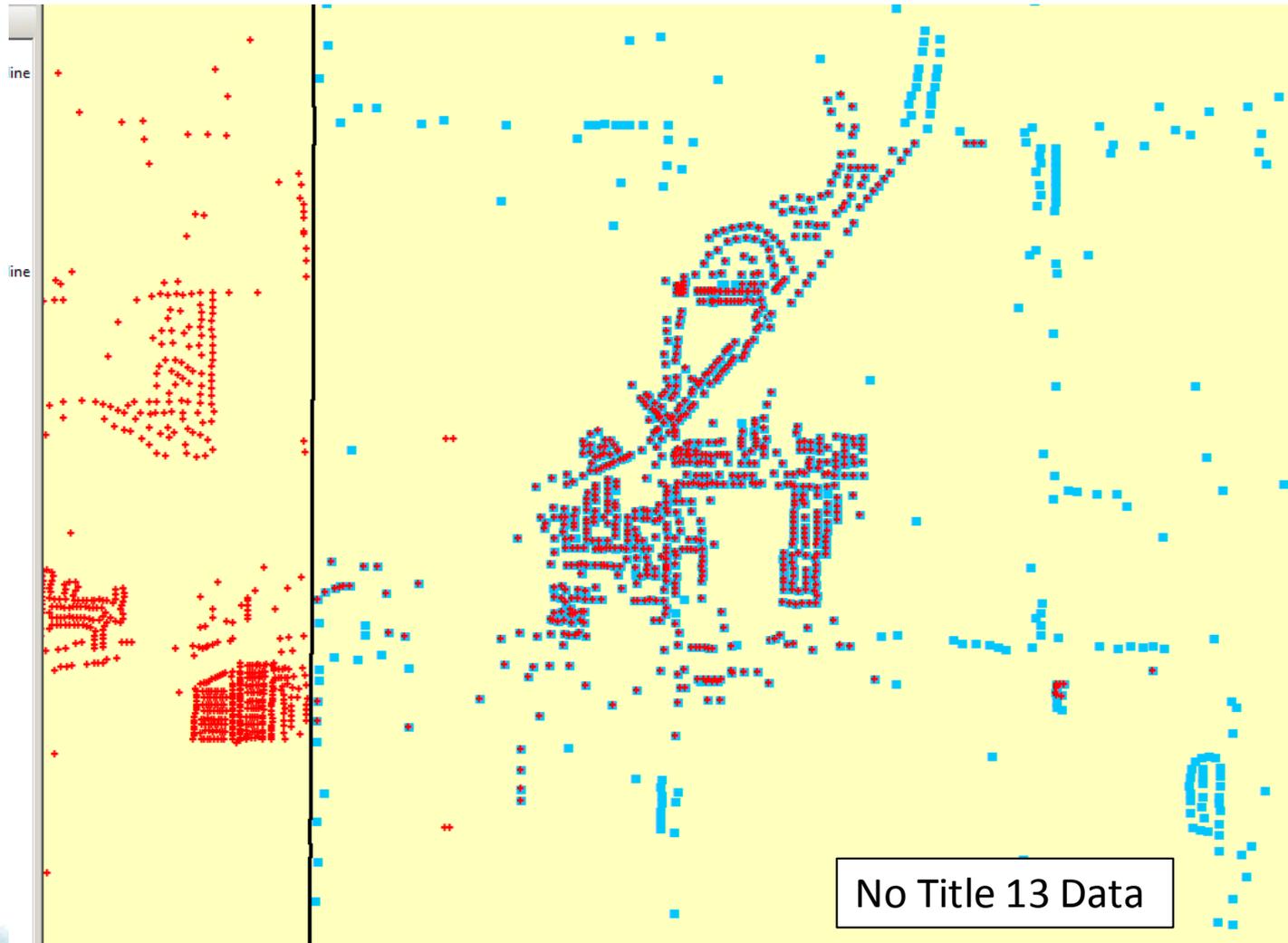
NAD County vs. Census County Selected Results (Cont.)

State	Counties	Count Based On TIGERLine Geocoding	Count Based On NAD Attribute Geocode	Difference
North Carolina	Robeson	61080	61100	-20
North Carolina	Cumberland	129943	129924	19
Ohio	Cuyahoga	483281	485189	-1908
Ohio	Lorain	133708	133099	609
Ohio	Butler	186530	187080	-550
Ohio	Fulton	21632	21258	374
Ohio	Williams	18570	18935	-365
Ohio	Sandusky	31228	30970	258
Utah	Emery	6325	6365	-40
Utah	Grand	4548	4510	38
Virginia	Henry	29624	36329	-6705
Virginia	Martinsville city	13178	6592	6586
Virginia	Franklin city	8210	4133	4077
Virginia	Southampton	9655	13728	-4073
Virginia	Galax city	6571	3584	2987
Virginia	Carroll	19460	20992	-1532
Virginia	Grayson	10986	12503	-1517

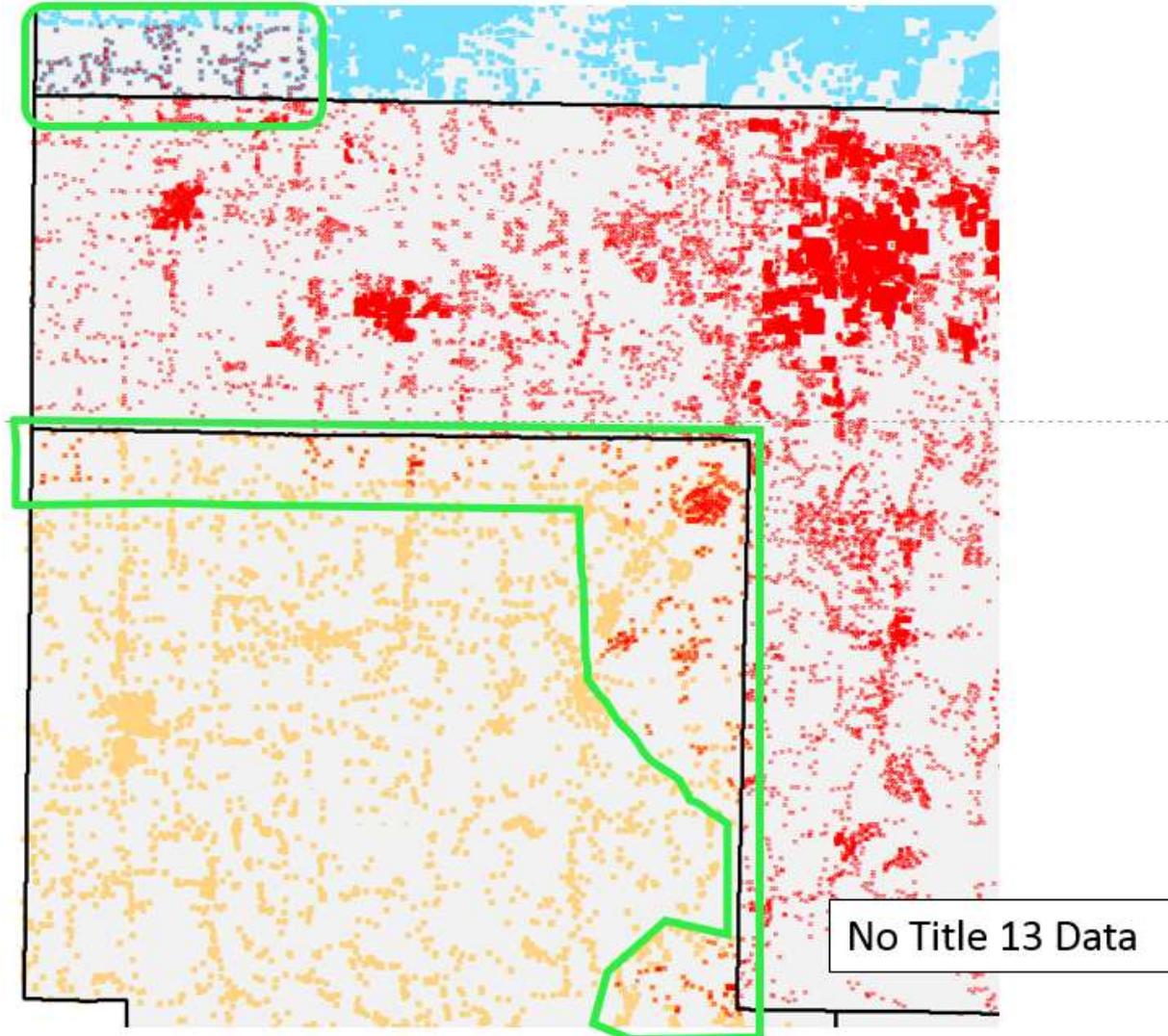
Analysis - Cuyahoga County, OH NAD Address Point Overflow



Analysis - Address Points that Overlap in a Municipality



Address Points that Overlap in Neighboring Counties



Geocoding Comparison

NAD County vs. Census County - Analysis Continued

Other possible explanations:

- Minor differences in spatial accuracy between county boundaries and address points.
- Emergency Management Areas serve neighboring areas of adjacent counties.
- Access points originating in adjacent counties for structures with locations in the assigned county.

Census Block Address Count Comparison: NAD vs. MAF

Background

- Not a direct indicator of which resource is better, just a comparison of two different address resources.
- There are significant differences between the composition of the MAF and the NAD.
 - NAD contains addresses for structures of both human habitation and human activity.
 - Residential
 - Commercial
 - Government/Public
 - Other
 - NAD may contain other locations that are not structures of human habitation or activity.

NAD vs. MAF Continued

Background (Continued)

- MAF/NAD Differences Continued
 - The Census Bureau's mission is to collect all residential addresses in the MAF.
 - The MAF does contain commercial and other addresses, identified as “Non-residential.”
 - Census does not actively seek to add commercial addresses to the MAF, only identify these as “Non-residential” when there is a business reason.
 - Field work
 - Indication of trusted address resource
 - USPS Delivery Sequence File (DSF) indication

NAD vs. MAF Address Count Comparison

Method

- Create MAF address counts by Census 2010 Tabulation Block for each NAD state in the comparison.
- Utilize both residential and non-residential addresses from the MAF in the counts, to more closely match the address inventory in the NAD.
- MAF Count Specifics
 - Passes Decennial Census Filter / Eligible for Decennial Operations.
 - Above includes housing units, group quarters and transitory locations.
 - MAF non-residential, including those commercial and government/public addresses that exist in the MAF.
- This does not ensure an exact inventory comparison with the NAD.

NAD vs. MAF Address Count Comparison (Continued)

Method (Continued)

- Geocode NAD addresses to TIGER/Line File 2010 Census Tabulation Blocks.
- Tabulate NAD address counts by 2010 Census Tabulation Blocks.
- Import MAF and NAD counts by 2010 Census Tabulation Blocks into a database application.
- Create calculated fields comparing MAF and NAD count differences.
 - Absolute difference - MAF address count minus NAD address count
 - Percentage difference compared to MAF - NAD count divided by MAF count, excluding MAF zero count blocks

NAD vs. MAF Address Count Comparison (Continued)

Method (Continued)

- Sort the MAF / NAD absolute count difference descending, identifying census blocks with the largest positive MAF count difference.
- Identifies blocks for further investigation, where the MAF may be more complete than the NAD in a given census block.
- Investigate further with additional resources to analyze and determine reason(s) for count difference.
 - Compare NAD address inventory in the census block to MAF inventory.
 - View imagery in census block to determine structure inventory and type.
 - What address inventory and physical conditions are observed? Is there significant evidence to support a conclusion?

Typical Explanation of Count Difference High Rise Apartments with BSAs Only, No WUIDs



Typical Explanation of Count Difference Low Rise Apartments with BSAs Only, No WUIDs

No Title 13 Data



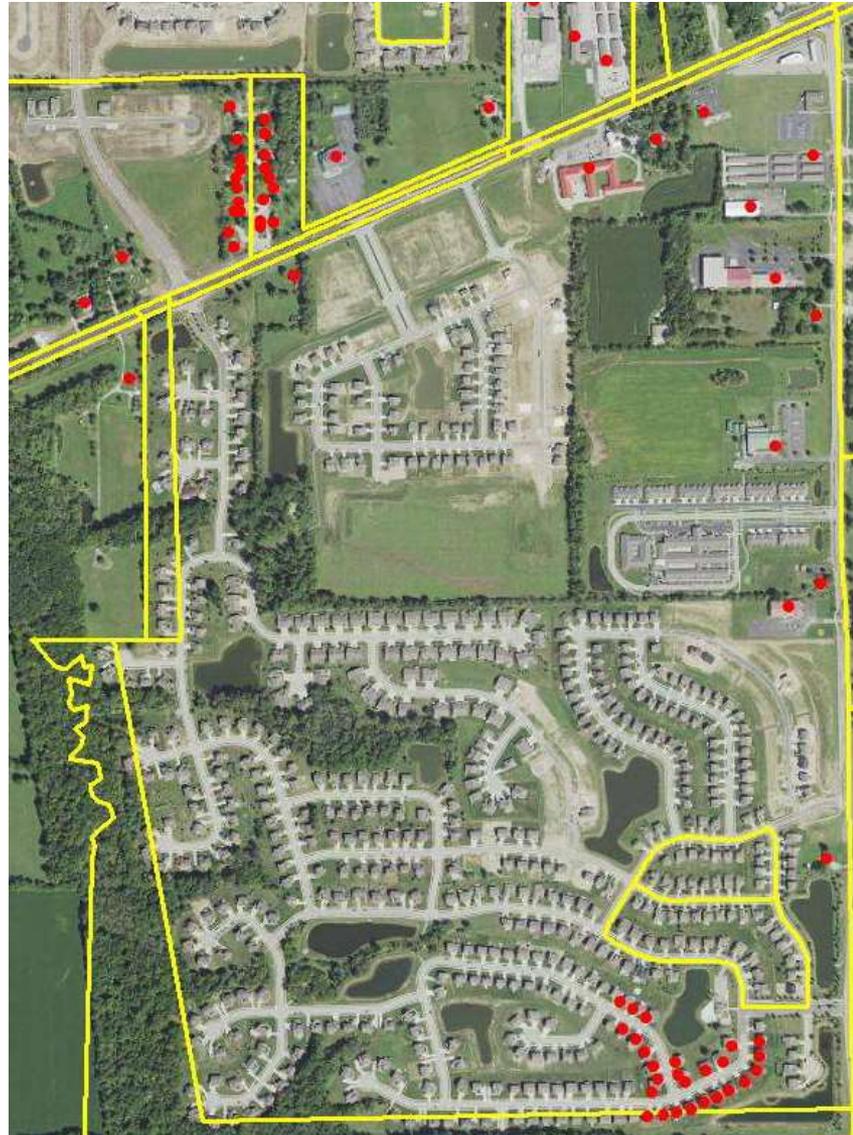
NAD Missing Dozens of BSAs from Apartment Complex / Parcel Centroid

No Title 13 Data



Captured some SFHs in the SE Part of Block, but Most SFH Addresses Missing

No Title 13 Data



Multiple Blocks, No NAD Addresses Mixed SFH & Multi-Unit

No Title 13 Data



1 BSA Only, NAD Missing 3 BSAs from Apartment Complex / Parcel Centroid

No Title 13 Data



Same BSA Repeated 4 Times No Unit or Building Designations

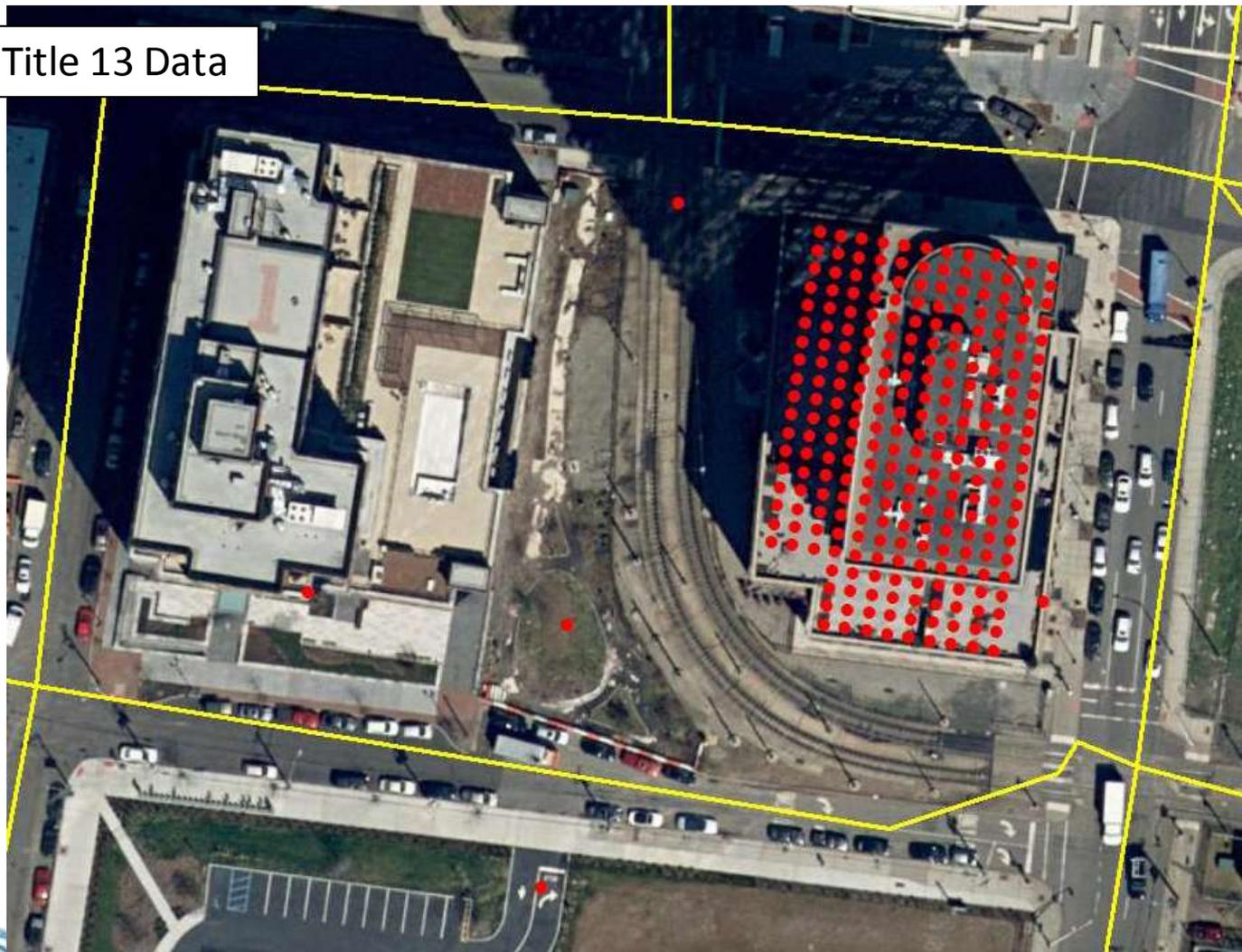
No Title 13 Data



Different Representations of Multi-Unit Structure Addressing

Left has 1 Point, Right Has All Points but No WUIDs

No Title 13 Data



Missouri - MAF Geocoding Error, Accounts for MAF / NAD Count Difference



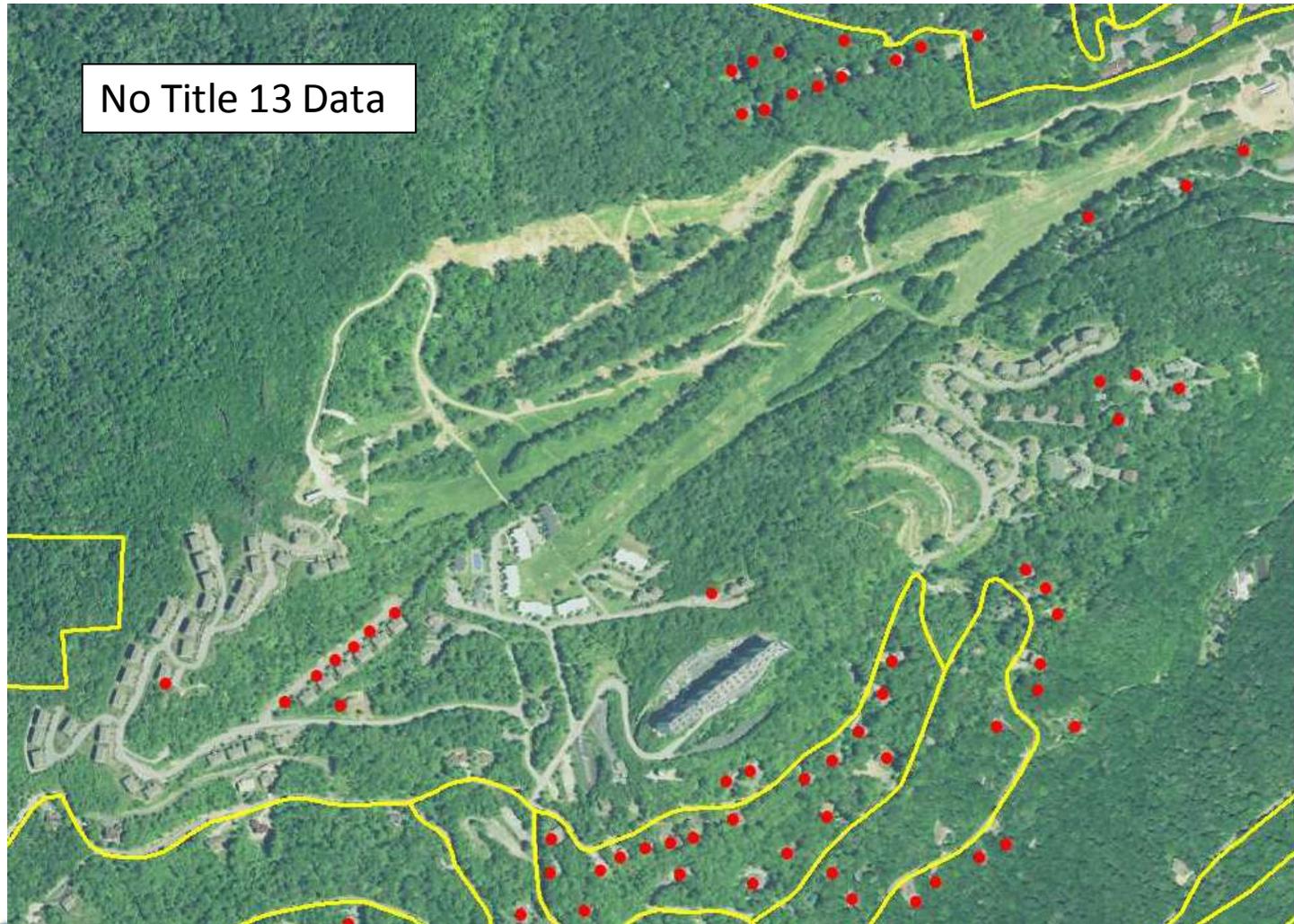
No Title 13 Data

District of Columbia - MAF Addresses Geocoded to Wrong Block, Accounts for MAF / NAD Count Difference

No Title 13 Data



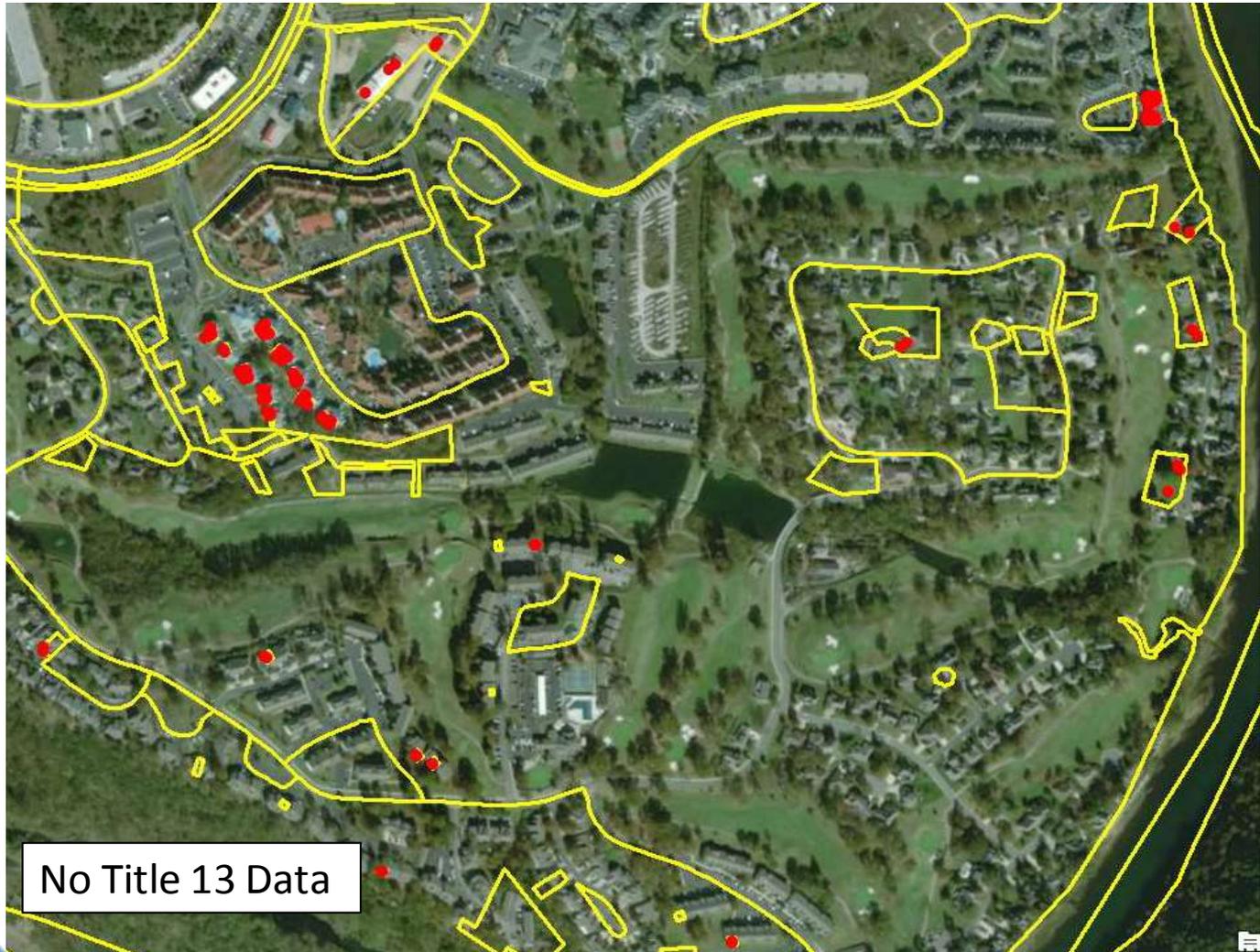
NAD Represents Some, but Missing Many Addresses at Resort Area



No Title 13 Data



Resort Area with Condominiums, Not in NAD



No Title 13 Data

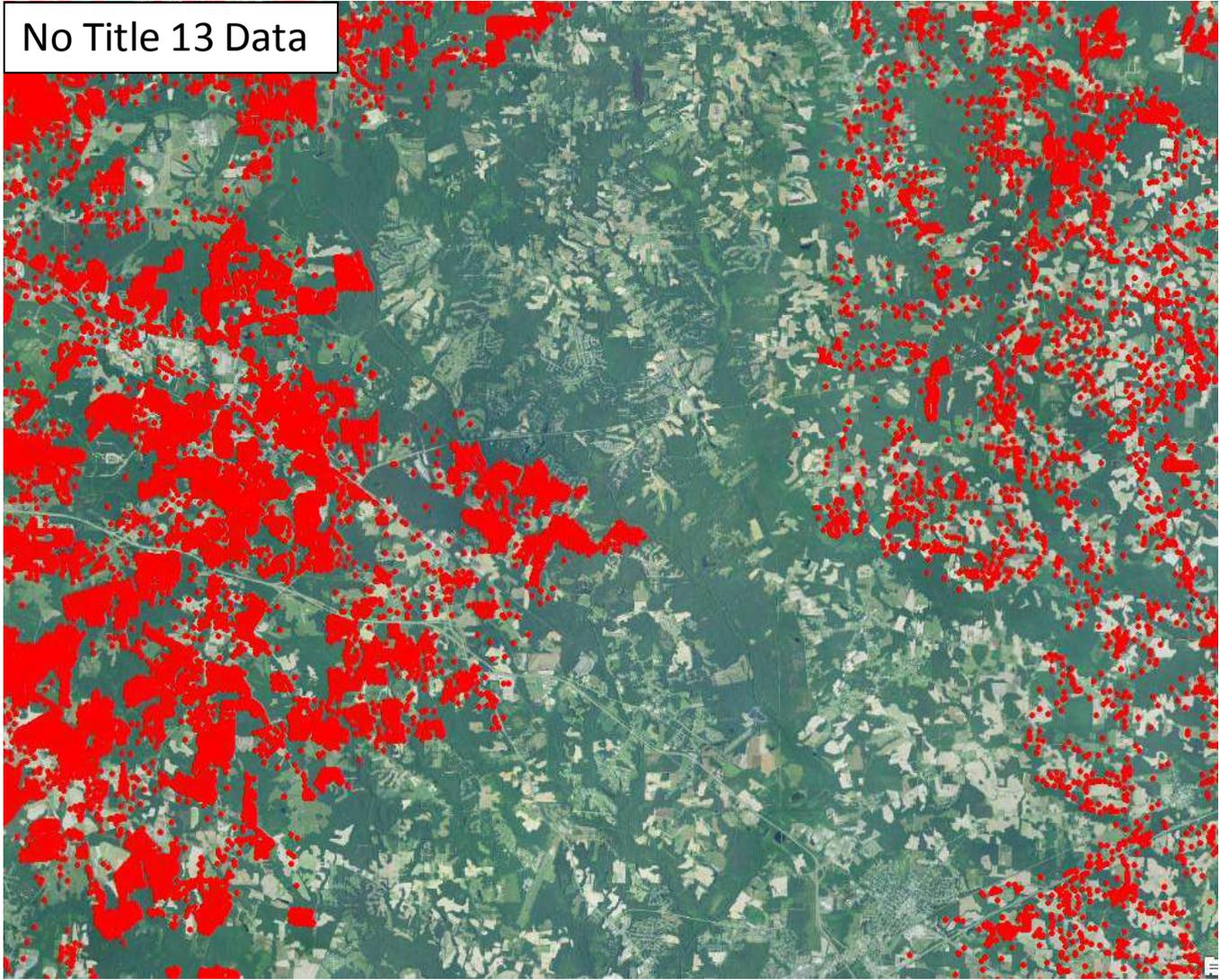
Apartments or Condos, Some Address Points Have Building Designations, Some Represent All Buildings with One Point

No Title 13 Data



Large Areas Devoid of NAD Address Points

No Title 13 Data



Partial NAD Coverage, SFH and Multi-Unit Significant Number of SFH Missing

No Title 13 Data



All Addresses (325+) in View Same BSA Only ~25 with WUIDs, Almost All in One Block

No Title 13 Data



NAD Missing Large Retirement Community

No Title 13 Data



Address Theme and NAD Team Contacts

Steve Lewis
steve.lewis@dot.gov
202-366-9223

Lynda Liptrap
lynda.a.liptrap@census.gov
301-763-1058

Matt Zimolzak
matthew.a.zimolzak@census.gov
301-763-9419

David Cackowski
g.david.cackowski@census.gov
301-763-5423

