The Address Theme and National Address Database Update

GIS for Transportation Symposium
March 21, 2018

Presented by
Lynda Liptrap, U.S. Census Bureau
Steve Lewis, U.S Department of Transportation
Topics

• Address Theme activity update

• National Address Database update
Federal Geographic Data Committee (FGDC)
FGDC Address Theme

Addresses

Biodiversity & Ecosystems
Cadastre
Climate & Weather
Resources
Elevation

Geodetic Control
Geology
Governmental Units
Imagery
Land Use – Land Cover
Real Property

Soils
Transportation
Utilities
Water - Inland
Water - Oceans & Coasts
The Address Theme and the National Address Database (NAD)

The Address Theme: a collection of datasets pertaining to addresses and point locations

The NAD: an open and publicly available nationwide database of residential, commercial, and postal delivery addresses with location data (geospatial coordinates)
Address Theme and NAD Goals for FY18

- Make the NAD available on the FGDC GeoPlatform
- Conduct a NAD GeoPlatform Pilot
- Develop a Theme Strategic Plan
- Continue monthly Address Subcommittee meetings
- Establish a NAD Content Subgroup of the Address Subcommittee
- Continue Workflow Subgroup meetings
Address Subcommittee Subgroups

• Workflow Subgroup
  • Objective - recommend workflows for processing address updates from providers for the NAD and other Address Theme datasets
  • Chair, Matt Zimolzak (Address Dataset Manager, Census Bureau)
  • Co-chair, Dan Ross (Chief GIO, State of Minnesota)
  • Established summer of 2017

• Content Subgroup
  • Objective – re-evaluate and recommend minimum and optimal content for the NAD
  • Chair, Dave Cackowski, Census Bureau
  • Established February 2018
NAD Update Topics

• Background Info
• Contract for NAD ETL/Database Developer
• QA/QC/ETL Procedures
• Current NAD Status
• NAD Production Environment
First Step: The NAD Summit

- Held April 8-9, 2015 at the Maritime Institute in Linthicum, MD
- Funded by DOT/Bureau of Transportation Statistics
- Objective: To identify and discuss possible options for developing a National Address Database (NAD)
- Broad stakeholder representation
  - Government (Federal, State, Local, and Tribal)
  - Private Sector
  - Non-Profits and Trade Organizations
How the NAD Would Work

“Vetted” Crowdsourcing

Data quality and error feedback

Federal

State 1

Local 1

State 2

Local 2

Local 3

National geocoder

Data flow
Pilot Launch October 2015

• DOT Funded
• Included State and Local Partners
  • Arizona
  • Arkansas
  • Boone County, Missouri
  • District of Columbia, New Jersey, Ohio, Utah, Virginia, and 10 additional Missouri counties/cities joined after the pilot started
• Goals
  • Explore workflows and understand best practices for address roll-up
  • Compare Data Schemas
  • Develop minimum content guideline and associated data schema
## NAD Minimum Content Guideline

### The Address Itself
- Address Number
- Street Name
- Subaddress
- City/Town/Place
- County
- State
- Zip

### Geographic Location of the Address
- Lat/Long
- National Grid Coordinates

### Metadata About the Address
- Address authority
- Address source
- Address date
- Unique ID
- Type (residential, commercial, etc.)
- Placement (rooftop, driveway access, etc.)
NAD Schema
Contract for NAD ETL Developer

- Awarded to GIS Inc. in June 2017
- Onsite developer reported to DOT in July 2017
  - Initial focus on QA/QC of data submissions
  - Now assisting states with ETL development
- Additional GIS Inc. resources available for “crunch” times
DOT Quality Checks for Data 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
DOT Quality Check Report

- Report is an accounting of the QC process results
  - Summary information, including accepted and flagged (rejected) address records
  - Detailed results

- Report is shared with data provider
## DATA COMPLETENESS

*required* Flagged items are highlighted in red. Flagged items not required to be populated at the state level are highlighted in yellow.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Total Count</th>
<th>Populated</th>
<th>Empty/NULL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Feature</td>
<td>2,190,243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAUID</td>
<td>99.92%</td>
<td>2,188,597</td>
<td>1,646</td>
</tr>
<tr>
<td>PlaceID</td>
<td>68.80%</td>
<td>1,506,956</td>
<td>683,287</td>
</tr>
<tr>
<td>Longitude*</td>
<td>98.21%</td>
<td>2,151,068</td>
<td>39,175</td>
</tr>
<tr>
<td>Latitude*</td>
<td>98.21%</td>
<td>2,151,068</td>
<td>39,175</td>
</tr>
<tr>
<td>AddrNum*</td>
<td>99.02%</td>
<td>2,168,697</td>
<td>21,546</td>
</tr>
<tr>
<td>NumSuf</td>
<td>0.43%</td>
<td>9,985</td>
<td>2,180,858</td>
</tr>
<tr>
<td>PreDir</td>
<td>36.75%</td>
<td>804,813</td>
<td>1,385,430</td>
</tr>
<tr>
<td>PreType</td>
<td>2.21%</td>
<td>48,353</td>
<td>2,141,890</td>
</tr>
<tr>
<td>StreetName*</td>
<td>98.65%</td>
<td>2,160,783</td>
<td>29,460</td>
</tr>
<tr>
<td>PostType</td>
<td>87.69%</td>
<td>1,920,655</td>
<td>269,588</td>
</tr>
<tr>
<td>PostDir</td>
<td>0.64%</td>
<td>14,006</td>
<td>2,176,237</td>
</tr>
<tr>
<td>UnitType</td>
<td>9.21%</td>
<td>201,718</td>
<td>1,988,525</td>
</tr>
<tr>
<td>UnitNumber</td>
<td>13.56%</td>
<td>297,094</td>
<td>1,883,149</td>
</tr>
<tr>
<td>AddrFull</td>
<td>99.09%</td>
<td>2,170,295</td>
<td>19,947</td>
</tr>
<tr>
<td>PlaceName</td>
<td>67.59%</td>
<td>1,480,330</td>
<td>709,513</td>
</tr>
<tr>
<td>Zipcode*</td>
<td>93.52%</td>
<td>2,048,246</td>
<td>141,997</td>
</tr>
<tr>
<td>IsCAI</td>
<td>0.00%</td>
<td>0</td>
<td>2,190,243</td>
</tr>
<tr>
<td>ParcelID</td>
<td>6.63%</td>
<td>145,118</td>
<td>2,045,125</td>
</tr>
<tr>
<td>MOD_DATE*</td>
<td>80.08%</td>
<td>1,754,049</td>
<td>436,194</td>
</tr>
<tr>
<td>ACT_STAT</td>
<td>80.08%</td>
<td>1,754,049</td>
<td>436,194</td>
</tr>
<tr>
<td>PROC_STAT</td>
<td>80.08%</td>
<td>1,754,049</td>
<td>436,194</td>
</tr>
<tr>
<td>County*</td>
<td>100.00%</td>
<td>2,190,243</td>
<td>0</td>
</tr>
</tbody>
</table>
DOT Quality Review Summary Statistics

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

# Excludes New York City Addresses

- Selected results.
- All summary statistics are for Beta NAD versions.
NAD Production Environment

• DOT and Census will partner with DOI to leverage the Geospatial Platform

• Bulk download from an AWS S3 bucket is available now
  https://s3.amazonaws.com/nationaladdressdata/NAD_20180215.gdb.zip

• Map based viewer and collaboration site will follow (GeoPlatform Pilot)
NAD Timeline

From Sprint to Snooze to Sprint.....

NAD Summit – April 2015

NAD Pilot – September 2015

Pilot Ends, as do resources – July 2016

Contract Awarded for ETL Developer – July 2017

.....we need additional partners to keep sprinting!
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