

Address Subcommittee Meeting



April 10, 2019

11:00 am– 12:30 pm Eastern

US Census Bureau

Meeting Agenda

- 1. Welcome and Introductions/Roll Call**
- 2. NAD Updates – Steve Lewis**
- 3. Puerto Rico Address Data Working Group Update – Lynda Liptrap**
- 4. UUID Proposal from Address Content Subgroup/Workflow Subgroup – Matt Zimolzak**
- 5. Update from Workflow Subgroup – Matt Zimolzak**
- 6. Update from Address Content Subgroup – Dave Cackowski**
- 7. Action item review – Dave Cackowski**
- 8. Adjourn**

NAD Updates

Texas Parsing

- Using an open source address parsing script hosted on GitHub developed by DataMade and the University of Sussex
 - <https://github.com/datamade/usaddress>
- Parsing is being completed in 100,000 feature blocks to minimize rework due to parsing errors
- Base code was turned into a script to minimize the parsing translation from the native format to the NAD schema

Texas Parsing – Part 2

- The parsing is struggling with features with invalid names, <NULL> names, COUNTY ROADS, and private roads with multiple numbers or directionals
 - 100 COUNTY ROAD 1 creates an incorrect parse by dropping COUNTY from the data
 - 100 PRIVATE ROAD 1 S creates an error that terminates the parsing script
 - 100 SMITH creates an error that terminates the script
 - 100 <NULL> creates an error that terminates the script

Texas Parsing – Part 3

- The address parser can be trained with new sample data to improve results
- To date, 4.2 million of the 9.3 million records have been processed, with only 111 points flagged with <NULL> values
- Anticipating 15% – 20% incorrect parsing with current model which will improve as the model is retrained
- Have to decide when to move on

New Partner – Wyoming

- No statewide address program, but WYGEO is helping to gather county data
- To date, 3 of the 24 Wyoming counties have agreed to provide data (~44,000 address points)
- Another 3 counties have parcel layers that could be turned into addresses
- Outreach continues

New Partner – Wisconsin

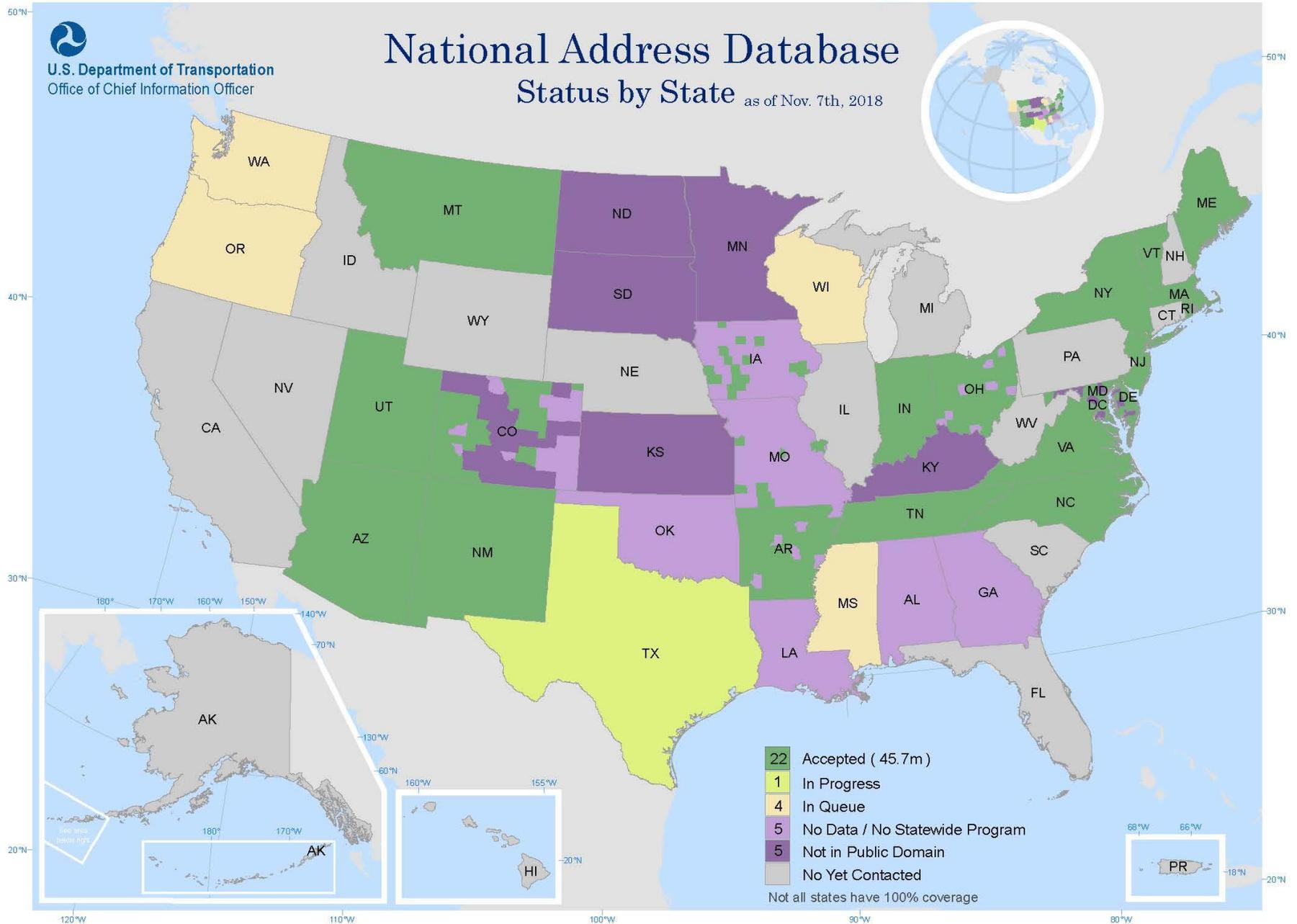
- Has been a committed partner for several months, but wasn't able to provide data until yesterday (4/9/2019)
- Will add 3 million address points to the NAD
- In the queue for processing



U.S. Department of Transportation
Office of Chief Information Officer

National Address Database Status by State

as of Nov. 7th, 2018



Puerto Rico Address Data Working Group Update

Puerto Rico Address Data Working Group Update

- Working Group Identified as a need at a workshop in October 2018
- Objective
 - Document and socialize proven practices for collecting, managing and sharing Puerto Rico address data within federal agencies
 - Encourage the implementation and update of existing national address standards for Puerto Rico addresses managed by federal agencies
 - Publish recommendations for address content that corresponds to the Proven Practices gathered across federal agencies.

Puerto Rico Address Data Working Group Update

- Kick-Off Meeting, April 2, 2019
 - Opening Remarks – Dr. Nancy Potok
 - Building a Proven Practices Inventory – Census Bureau Example
 - Next Steps – Build an inventory based on each agency's Proven Practices

UUID Proposal from Address Content Subgroup/Workflow Subgroup

UUID Requirements

- UUID shall be unique for each record in the address database, and shall not be duplicated for a different address record in said database.
- UUID shall persist for its address record in each iteration of the address database, immutable.
- If an address record in the address database is retired, its UUID is retired with it and cannot be used again for another address record.
- If the provider does not have the capability to maintain a persistent UUID for individual address records across iterations of the address database, the responsibility for assigning a persistent UUID to address records in the database evolves to the state level aggregator and as necessary, the federal level aggregator.

UUID Considerations

- Prefer decentralized UUID assignment and management, facilitates wider adoption.
- Most database software packages used by governments offer the capability to create and assign UUID with extreme low probability of duplication by another decentralized entity.
- Additional information on UUID can be found in the FGDC Address Standard, Section 2.3.1.1.

FGDC Address Standard – Address ID

2.3 Address Attributes

2.3.1 Address ID

2.3.1.1 Address ID

Element Name	AddressID
Other common names for this element	
Definition	The unique identifier assigned to an address.
Definition Source	New
Data Type	characterString
Existing Standards for this	None

Element	
Domain of Values for this Element	No
Source of Values	Primary key, issued locally
How Defined (eg, locally, from standard, other)	Locally
Example:	Integer ID: 1243286 UUID: 550e8400-e29b-11d4-a716-446655440000
Notes/Comments:	<ol style="list-style-type: none"> 1. The Address ID is a required element of an address data record. The ID must be unique for each address assigned by an Address Authority. In cases where an Address Authority does not assign an Address ID, it may be assigned by an address aggregator, such as a regional government, state government, federal agency or a commercial address aggregator. The Address ID may be either a locally generated unique ID, or it may be a Universally Unique ID (UUID) which is machine-generated within the database environment. 2. IDs are almost always integers, and integer ID's are much easier to manage. However, some ID schemes use hyphens, leading zeros, or other non-integer characters, so the standard also accommodates alphanumeric IDs. <p>Notes and Reference Information on UUID</p> <ol style="list-style-type: none"> 1. A UUID is presented as a 16-byte (128-bit) number written in hexadecimal form computed according to a UUID algorithm. At least five algorithms have been developed. 2. UUIDs are documented in two standards, ITU-T X.667 and IETF RFC 4122 (see Appendix A for complete references). The two standards are technically consistent. 3. This standard provides for a UUID as a means to identify an address while it is passed from the originating source through a chain of intermediaries to the end-user. The need arises because there exists within the United States no central coordinating body to identify and register addresses. There is not even a registry of the authorities empowered to create addresses, nor is one likely to be created. 4. "The intent of UUIDs is to enable distributed systems to uniquely identify information without significant central coordination. Thus, anyone can create a UUID and use it to identify something with reasonable confidence that the identifier will never be unintentionally used by anyone for anything else. Information labeled with UUIDs can therefore be later combined into a single database without need to resolve name conflicts." (quoted from Wikipedia, "Universally Unique Identifier", as posted 4 September 2010 at:

Workflow Subgroup Update



Address Content Subgroup Update

Action Items

Thank You



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

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