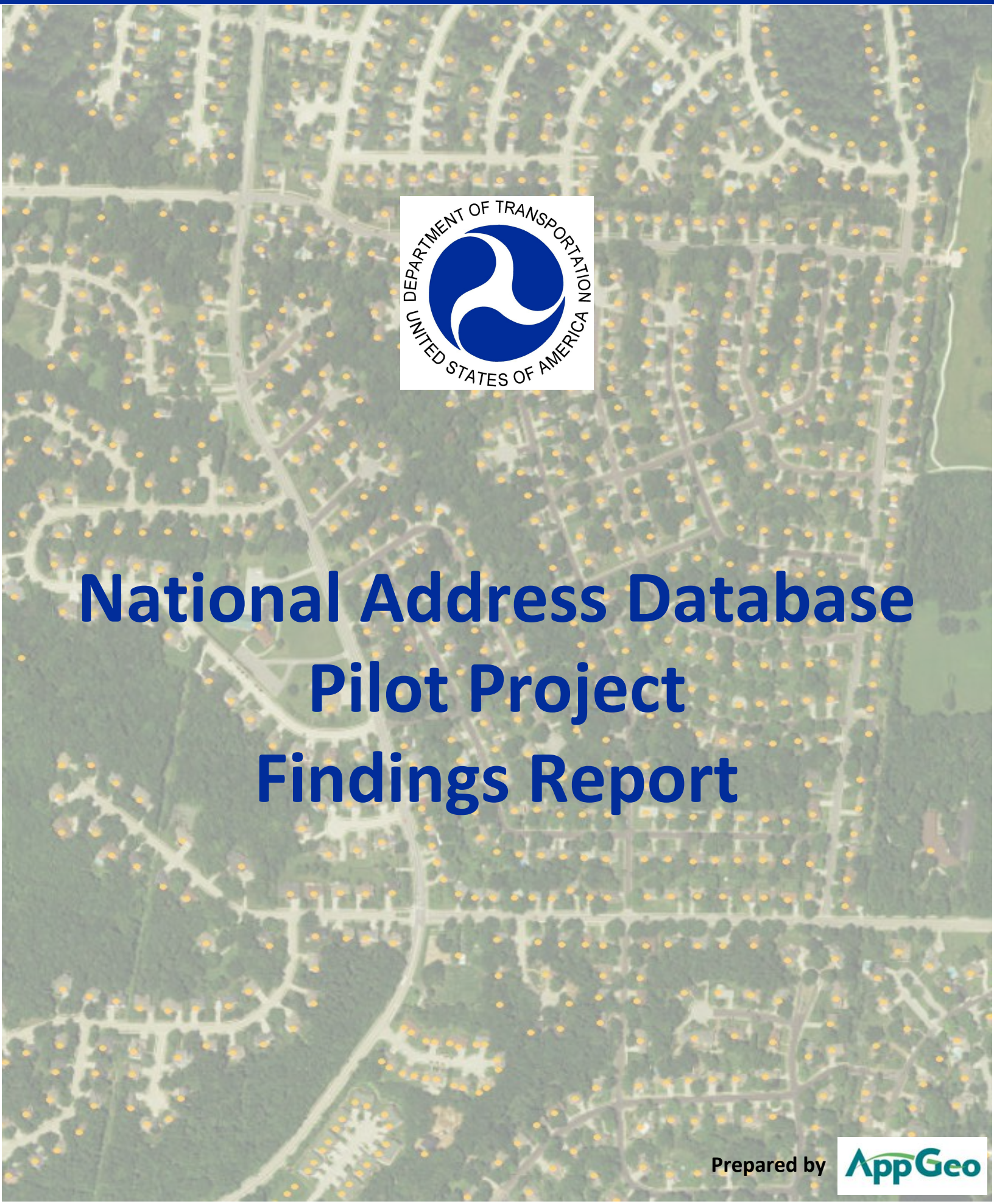


National Address Database Pilot Project Findings Report

Prepared by AppGeo

Table of Contents

Executive Summary

Background

The 2012 NGAC report titled *The Need for a National Address Database* set forth the vision of a National Address Database (NAD) as "*an authoritative and publicly available resource that provides accurate address location information".* This vision was further tested and discussed at the 2015 National Address Database (NAD) Summit. As a result of discussions at the NAD Summit, participants generally agreed that it was time to "stop talking and start doing" and that pilot projects were needed to test the feasibility of the NAD.

NAD Pilot Project Goals

- Determine minimum data content guidelines and schema
- Understand best practices for address roll-up
- Explore workflows for new address creation by jurisdictions that do not currently have address point data
- Assess the technical feasibility of the NAD

NAD Minimum Content Approach

The overarching ideals of the Minimum Content guidelines include

- The NAD is an **aggregation of authoritative address *point* data**
  - Authoritative data is considered to be the data emanating from the entity responsible for the address creation and maintenance (i.e., the Content Originator, most typically a city, town or county).
- It will maintain **low barriers to participation** (while maintaining quality standards)
- The NAD schema will be **intentionally simple** to the extent possible, while adhering to agreed-upon best practices (e.g., full parsing, domains for validation, etc)
- There is no intent for, or expectation that address authorities will manage their data in the NAD schema. Rather, agencies are encouraged to use whatever schema works best for them, and make their data available for aggregation into the NAD.

NAD Pilot Participants

- The following state and local government partners *with existing* comprehensive address point data provided their data for standardization and loading into the NAD schema
  - State of Arkansas
  - State of Arizona
  - Boone County, Missouri

- The following county that *did not* currently have address point data participated in pilot phase involving new address point data creation
  - Jackson County, Arkansas

- The following state and local governments volunteered data that was pre-loaded into the NAD Schema
  - State of VA
  - State of NJ
  - Washington DC
  - State of UT
  - A subset of counties in OH (via the State of Ohio)
  - A subset of counties in MO (via Boone County)

Executive Summary (cont.)

Findings

Participation and Data Sharing

- Tribal participation is likely to be a challenge due to data sharing concerns
- Data sharing agreements to keep NAD data publically available may also be a challenge

Feasibility of Aggregation

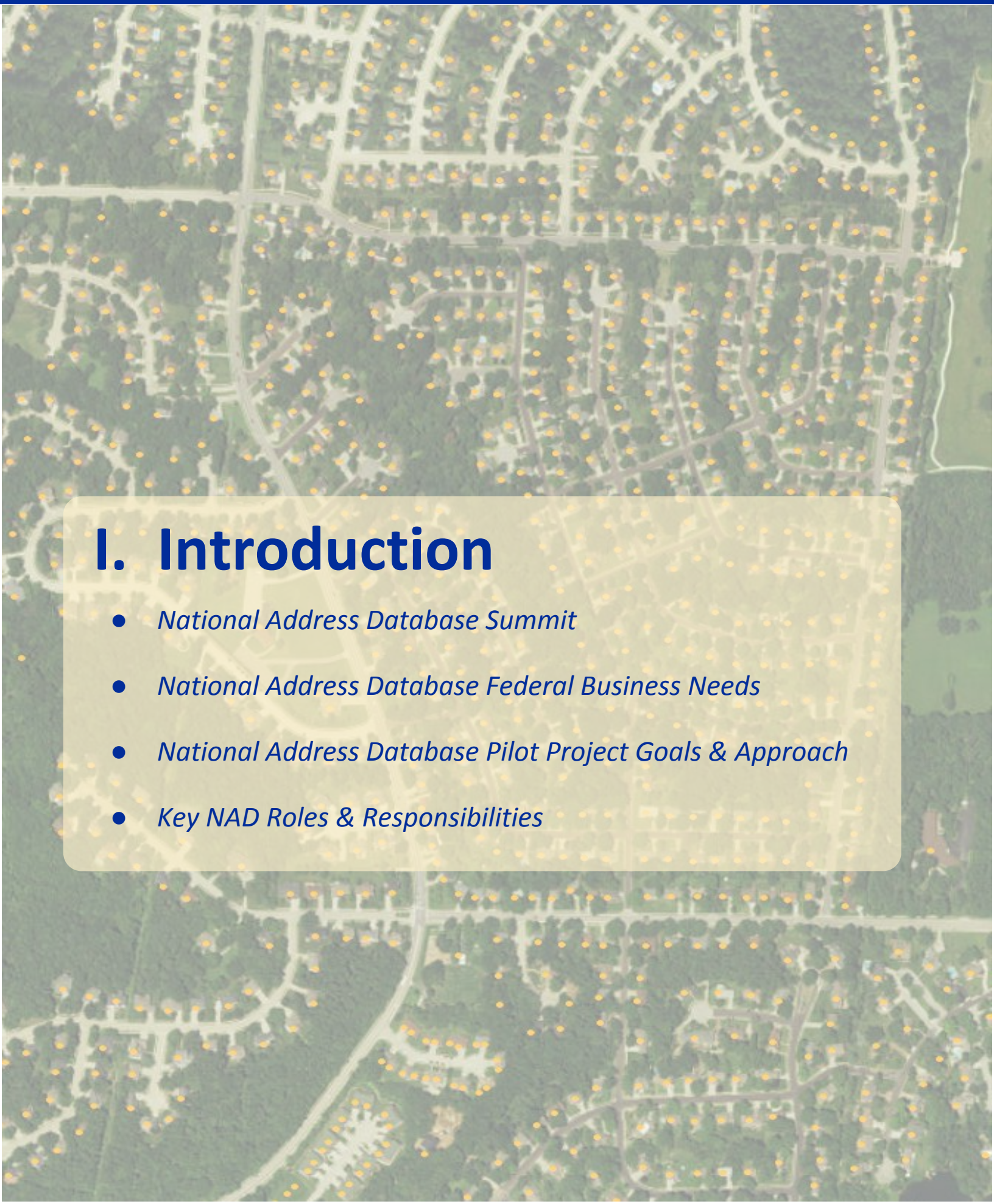
- Developing the NAD is technically feasible but will involve overcoming some obstacles
- The pilot schema will evolve, but needs to remain consistent with leading address standards and schemas to allow for streamlined ETL (Extract, Transform and Load) and aggregation
- Aggregating existing statewide collections was straightforward for pilot datasets, but will become more complex as more participants with varying levels of technical capability get involved

Data Creation and Maintenance

- The ease with which new address point data can be created depends largely on quality of existing source materials (e.g., address lists; parcel data; street centerline data; etc.)
- To build the most comprehensive NAD, communities without existing address data may need funding support to get started with data creation

Next Steps

- Continued and sustained education and outreach on the NAD effort should be pursued
- Standing up a preliminary NAD database within the USDOT and Establishing the ingest process and acceptance criteria
- Outreach to existing statewide aggregators to obtain voluntary contributions to the NAD should be pursued
- Tackling regular data updating within the NAD via recurring contributors needs to start
- Identifying funding and/or grant programs for address data creation that can be made available to jurisdictions that do not yet have address data is required



I. Introduction

- *National Address Database Summit*
- *National Address Database Federal Business Needs*
- *National Address Database Pilot Project Goals & Approach*
- *Key NAD Roles & Responsibilities*

National Address Database Summit

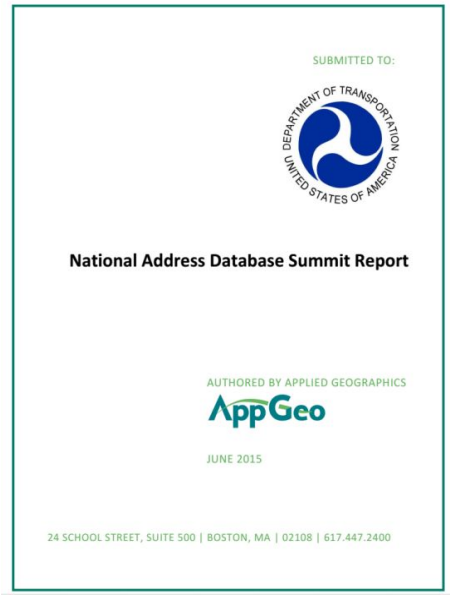
The 2012 NGAC report titled *The Need for a National Address Database* set forth the overarching vision that: "***The National Address Database is an authoritative and publicly available resource that provides accurate address location information to save lives, reduce costs, and improve service provision for public and private interests***".

To further that vision, **The National Address Database (NAD) Summit** was held in April 2015, sponsored by the United States Department of Transportation's (USDOT) Bureau of Transportation Statistics (BTS). The Summit provided a specialized forum for generating ideas and gathering input on the feasibility and format of a shared address database for the nation. The stated objective of the summit was to "identify and discuss possible options for developing a National Address Database" (NAD) and to discuss feasibility, possible approach, and next steps.

Summit participants came to broad agreement on four key points that can help guide the direction a NAD initiative may take:

1. **Local authorities** are the **authoritative source** for address assignment and are data set originators.
2. **State authorities** should be **statewide aggregators** of county and local data sets.
3. Given the vast and complex nature of the United States it is **critical to recognize the role of non-state governmental entities** such as **Tribal Nations, US Territories** and the **District of Columbia** play in an NAD.
4. **Federal leadership** and support is needed for there to be a sustainable national approach.

The key outcome of the summit was that action and activity are required to move the NAD forward. To that end, the participants agreed that a priority next step would be to **pursue pilot projects as quickly as possible** to both tackle unresolved issues and demonstrate the feasibility of a NAD database. The US Department of Transportation was able to fund this NAD Pilot project.



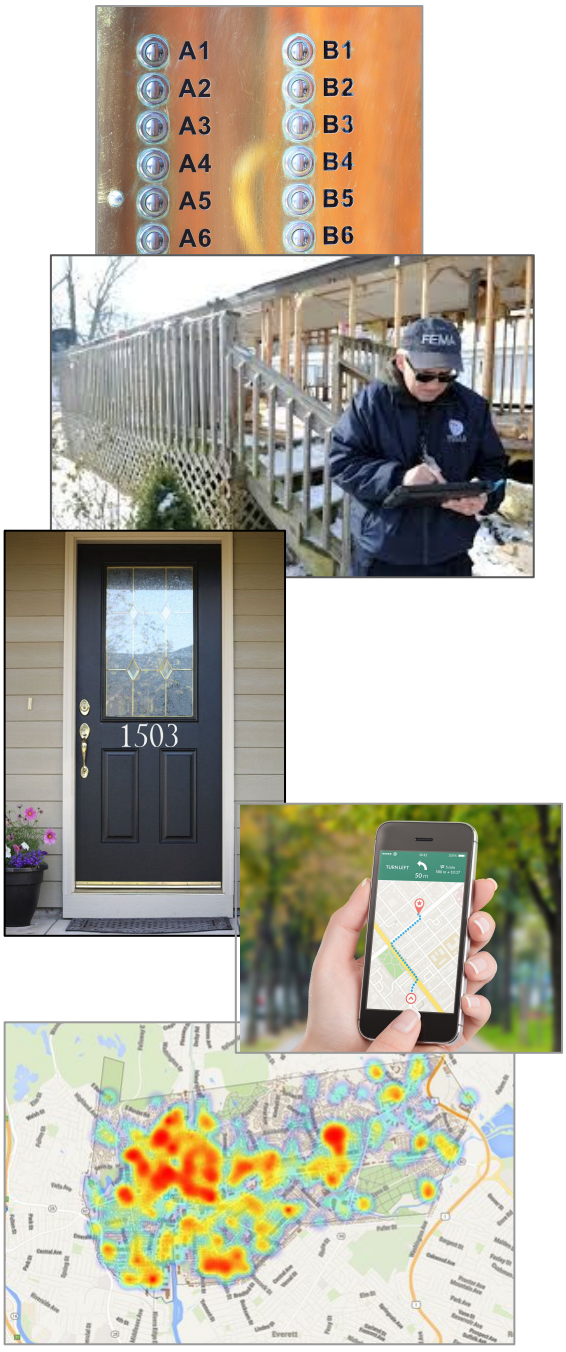
Click here to download the National Address Database Summit Report

National Address Database Business Needs

Address data is critical at all levels of government. NSGIC members helped to document address business needs in 2014, and compiled the results into the following list by the NSGIC Address Committee: https://www.nsgic.org/public_resources/Address_Business_Needs_101714_Final.pdf. This document lists many of the business functions for Local, State and Federal government that depend on quality address data.

The expected main use case of the NAD will be to serve as a **publically available nationwide geocoding data source based on authoritative data**. Additionally, the NAD can help support many of the Federal business needs identified in the NSGIC list mentioned above, for example:

- **US DOT Federal Highway Administration**
  - Office of Safety - Crash Reporting and Analysis
  - Office of Planning - Transportation Planning, Land Use
- **US Census**: Decennial enumeration of US population, maintaining Master Address File
- **FEMA**: Determine cost and need for individual assistance after natural and manmade disasters
- **Centers for Disease Control and Prevention**: Regulatory enforcement, visualization and analysis, surveillance activities, etc.
- **Consumer Financial Protection Bureau**: Fraud mitigation, risk assessment
- **Department of Homeland Security**: Preparedness and planning, situational awareness
- **Federal Housing Finance Agency**: Match transactions for the purpose of estimating FHFA's suite of house price indexes
- **Veterans Administration**, Office of Policy and Planning: Visualization and analysis, policy development and planning for delivery of services



NAD Pilot Project Goals and Approach

The major GOALS of the NAD Pilot Project are to:

- Determine minimum data content guidelines and schema
- Understand best practices for address roll-up
- Explore workflows for address creation
- Assess the technical feasibility of the NAD

The OUTCOMES of this pilot project include:

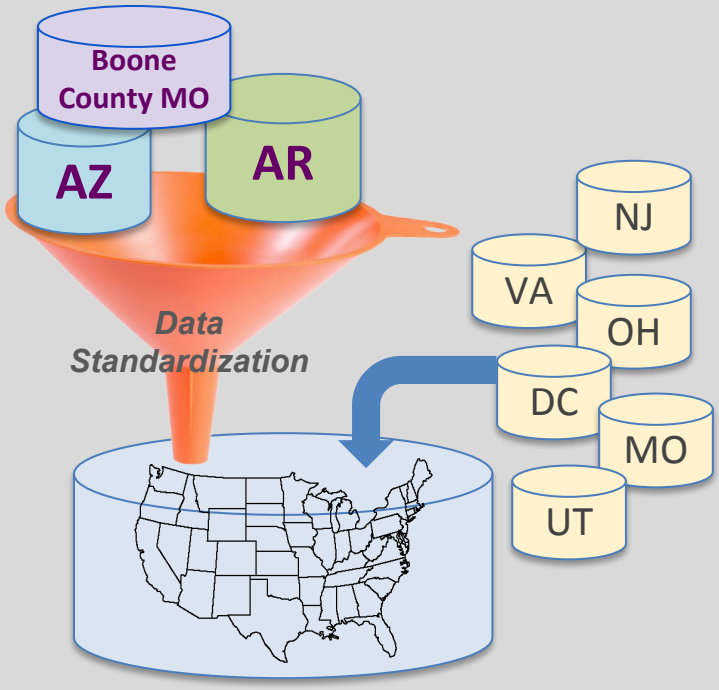
- A pilot address dataset which includes a small number of "test case" address datasets from participating jurisdictions
- Workflows and tools for address data ETL (extract-transfer-load) for aggregation
- Documentation (i.e., this report) that will inform next steps for performing these activities on a broader scale, and ultimately at the national level

The pilot will be considering two different address source cases

Jurisdictions that have address points - these would be used to test standardization and "rollup" procedures to a national dataset

Participants:
Arizona, Arkansas, Boone County Missouri

Additional NAD Volunteers:
Virginia, New Jersey, Utah, Washington DC, State of Ohio, and various counties in Missouri



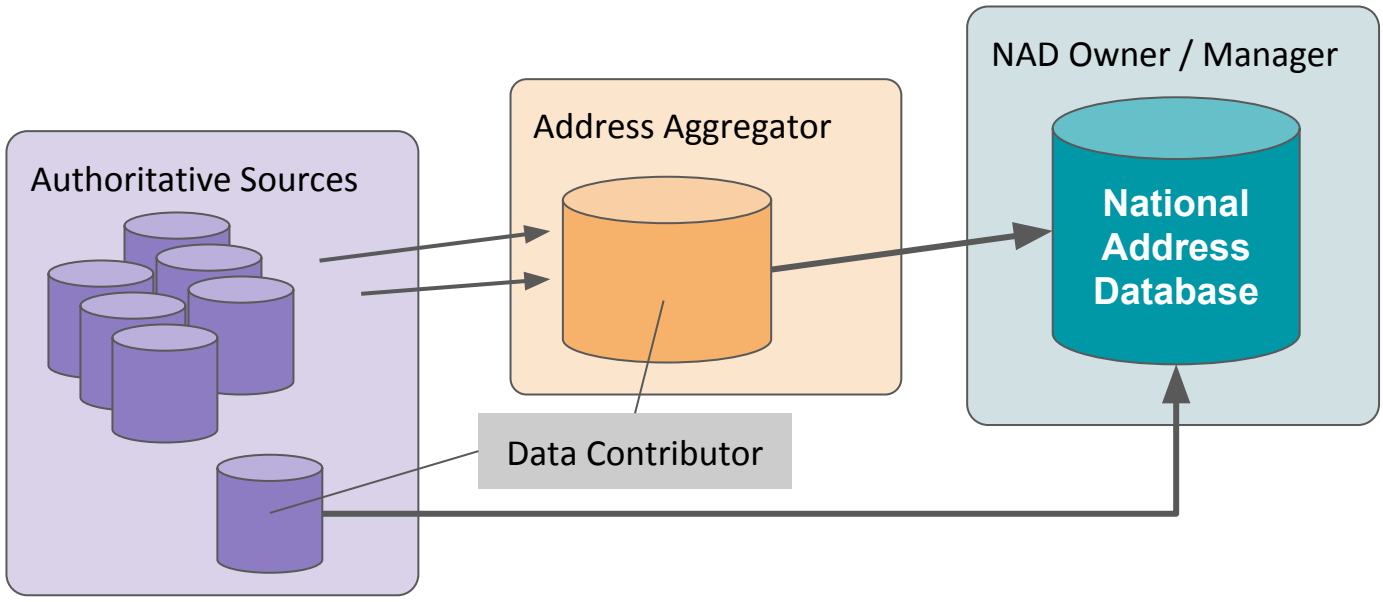
Jurisdictions that have NO address points these would be used to test streamlined ways of developing initial address data

Participant: *Jackson County, AR*

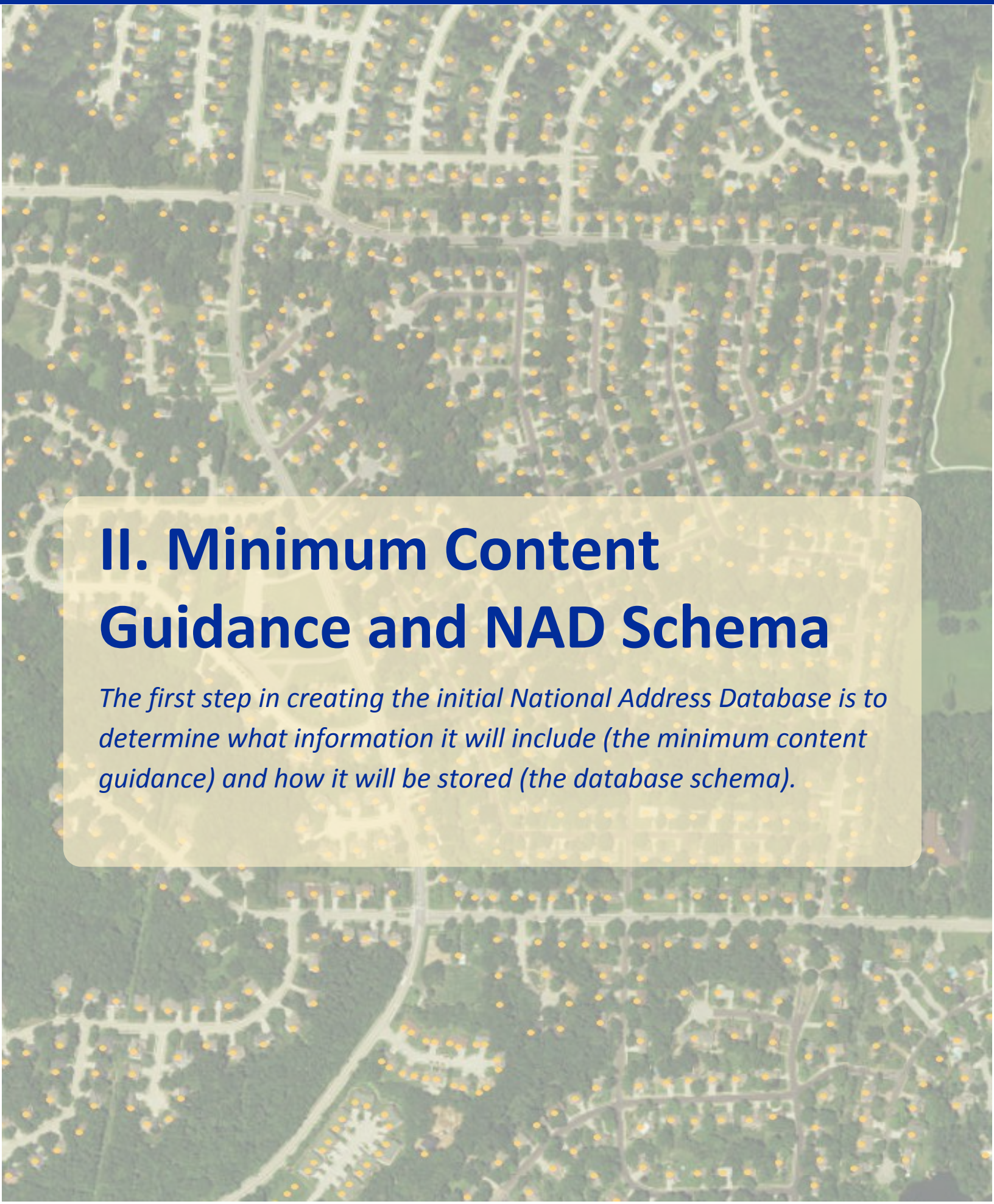
A	B	C	D	E	F
Name	**Address**	**Address2**	**City**	**State**	**Zip**
John Doe	123 Anystreet		Anycity	MN	12345
Betty Johnson	456 That Road	#108	Sometown	KY	67890
Tom Anderson	789 Broad Street		Dullsville	CA	23456
Sally Stevens	1011 Main Street	Unit 3	Pleasantville	NY	17890
Richard Townsend	1213 Mulholand Drive		Anywhere	WA	34567
Maria Sanchez	1415 Circle Road	Apt. 2	Sometown	NV	28901
Antoine Dodson	1617 Oregon Trail		Nowhere	FL	45678
Jane Doe	456 Anystreet	Unit 7	Anycity	MN	12345
Bart Johnson	123 That Road		Sometown	KY	67890

Key NAD Roles & Responsibilities

The diagram and table below provides a summary of the various roles involved in the creation, collection, aggregation and publishing of the NAD and how each role relates to the others.



Authoritative Source	Address Aggregator	NAD Owner / Manager
The Authoritative source is the entity responsible for the address point data creation and maintenance (i.e., the Content Originator, most typically a city, town, county or tribal area).	An address aggregator is often a state (or a regional entity), collects data from the local authoritative sources, and aggregates it to a standard format for submission to the NAD.	The NAD Owner / Manager is the entity that takes responsibility for accepting and reviewing submissions to the NAD, hosting and publishing the data, and managing updates.
Authoritative Sources *or* Aggregators can be a **Data Contributor**, depending on the submission path to the NAD.		



II. Minimum Content Guidance and NAD Schema

The first step in creating the initial National Address Database is to determine what information it will include (the minimum content guidance) and how it will be stored (the database schema).

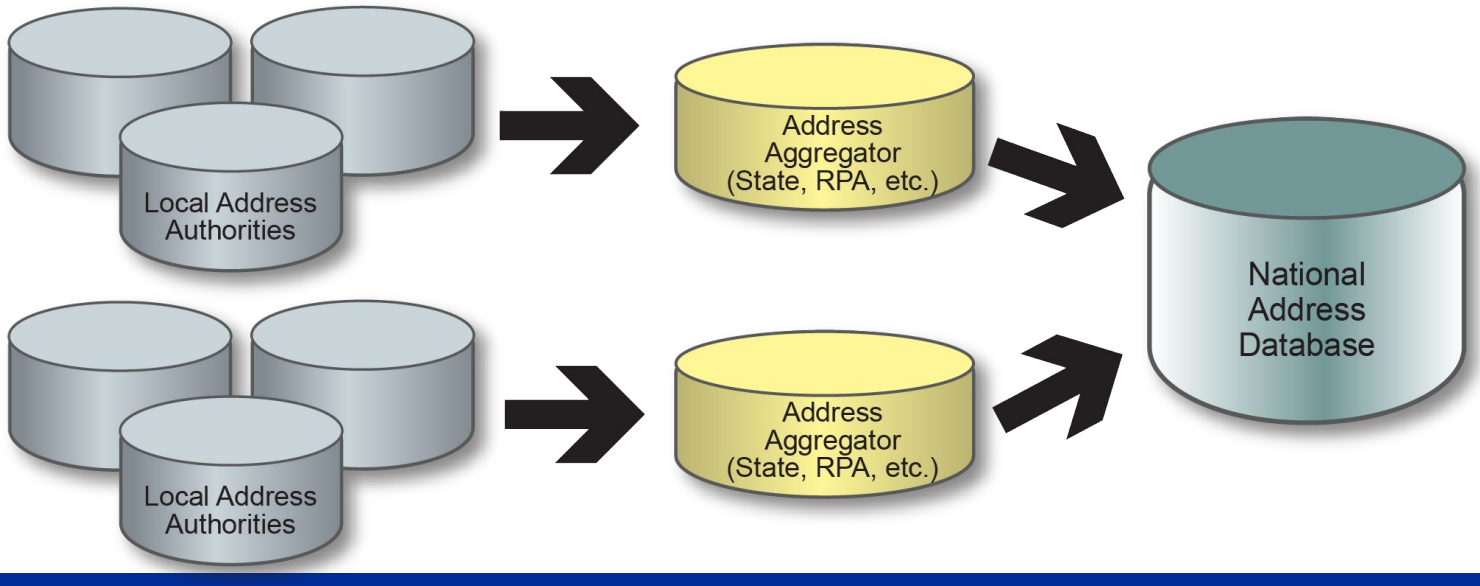
NAD Overview

The goal of the NAD is to identify the "Minimum Viable Product" (MVP) for aggregated address content that will be most generally useful

- MVP is a simple address point database, with one point per address
- The NAD Schema (and minimum content guidance) is one aspect of the MVP
- The MVP represents a common base that everyone shares
  - Many are still struggling to just get started
  - Some advanced contributors may have more advanced address datasets that go beyond the MVP, such as
    - "Non-street addresses" identifying parking lots or rest areas along highways, trail heads, etc.
    - Emergency Response zone information
    - Multiple points per address (e.g., road access point *and* building entry)
- The MVP and other products can be designed to work with one another to provide a more complete solution, over time
- There should be an attitude of continual process improvement (i.e., bringing more products, and more addresses on-line over time)

NAD Rollup

- NAD is a roll-up, **harvested from *authoritative*** data. Authoritative data is considered to be the data emanating from the entity responsible for the address creation and maintenance (i.e., the Content Originator, most typically a city, town or county).
- **The NAD will use the best practice of full street address parsing** similar to what is laid out in the Federal Geographic Data Committee (FGDC) *United States Thoroughfare, Landmark, and Postal Address Data Standard* (FGDC-STD-016-2011) and the National Emergency Number Association (NENA) *Civic Location Data Exchange Format* (CLDXF) standard (NENA-STA-004), including domains for street type and directionals for data validation.
- **For collection, the NAD will keep a low barrier to participation** (e.g., will not require a specific parsing schema for data contributors)



Minimum Content Guidance

The goal of the minimum content guidance is to include only the information needed to identify an address as well as some basic identifying/metadata information about the address. The objective is to avoid a complex schema as it is not intended to be a schema used for data *management*, but rather for national aggregation and rollup. In general, the NAD will contain three main components, seen at the right.

These data elements represent the ideal data content for the NAD. This ideal will be considered the goal to work towards and not an absolute requirement. As such, if a given jurisdiction (county, state, tribal area, etc.) has address data that does not contain all of these elements, the data will be accepted into the NAD to the extent possible. For example, certain components such as Address Type and Address Placement and subaddressing will not be required. However, datasets will be rejected if they don't contain the key basic address information as well as key contact information and metadata elements such as in address authority, address source and address date. Furthermore, if a jurisdiction has additional information in their dataset that is not included in the NAD at the time, they are encouraged to submit their data in its entirety as it may inform future additions to the NAD.

The NAD guidance will specify a target accuracy for both Long/Lat and United States National Grid (USNG) coordinates. For example, an ideal accuracy to 1 meter, but no greater than 10 meters. National Grid coordinates can be derived and don't necessarily need to be provided by the jurisdiction submitting the data.

NAD Minimum Content Guidelines Feedback Review:

- Round 1: NSGIC NAD Project Advisory committee (see appendix 1 for participants)
- Round 2: All NAD Summit attendees
- Guidelines were revised/refined in response to each round of comments

The NAD Minimum content guidelines contain three types of data elements:

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

- Unique ID (e.g., GUID)
- Address type (residential, commercial, etc.)
- Address placement (rooftop, driveway entrance, structure entrance, etc.)
- Address authority (i.e., data creator)
- Address source (i.e., data aggregator)
- Address date (i.e. date updated, valid date)

NAD Schema

This is the proposed (v1) NAD Schema for storing data in the NAD, including all attribute fields and domains. This schema is in line with the NAD Minimum Content approach.

The full schema is included in appendix 5.

A few key points regarding this schema:

- The data will be stored in **WGS 1984 Web Mercator**. All submissions must be projected properly into WGS 1984 Web Mercator
- This schema stores address points, with one point per address record.
- It is anticipated that this schema will evolve over time as the NAD becomes a reality.
- This is not a minimum schema for submission - it is anticipated that many of the fields may be null for many records, or perhaps nonexistent depending on the source data.
- The proposed schema is a single flat table (non-relational) based largely on the NENA CLDXF standard.
- Data may be harvested or contributed that has more content than the NAD schema. In these cases some data elements provided by the contributor will be stripped out of the NAD incarnation.

The "Expected Use" column is included simply to indicate whether a field is generally "always used", "commonly used", "occasionally used", or "rarely used" within a dataset to give submitters a sense of expectations. (Note: a "commonly used" attribute may be null for the majority of records, but is still likely to be utilized within an address database).

	Field Name	Field Alias	Type	Length	Domain	Expected Use
Address	State	**State**	Text	2	✔	always used
	County	**County**	Text	40	✔	always used
	Inc_Muni	**Incorporated Municipality**	Text	100		commonly used
	Uninc_Comm	**Unincorporated Community**	Text	100		commonly used
	Nbrhd_Comm	**Neighborhood Community**	Text	100		commonly used
	Post_Comm	**Postal Community Name**	Text	40		commonly used
	Zip_Code	**ZIP Code**	Text	7		always used
	Plus_4	**Zip Plus 4 Addition**	Text	7		occasionally used
	Bulk_Zip	**Bulk Delivery ZIP Code**	Text	7		rarely used
	Bulk_Plus4	**Bulk Delivery ZIP Plus 4 Addition**	Text	7		rarely used
	StN_PreMod	**Street Name Pre Modifier (PRM)**	Text	15		commonly used
	StN_PreDir	**Street Name Pre Directional (PRD)**	Text	50	✔	commonly used
	StN_PreTyp	**Street Name Pre Type (STP)**	Text	25	✔	commonly used
	StN_PreSep	**Street Name Pre Type Separator (STPS)**	Text	20	✔	commonly used
	StreetName	**Street Name (RD)**	Text	60		always used
	StN_PosTyp	**Street Name Post Type (STS)**	Text	15	✔	commonly used
	StN_PosDir	**Street Name Post Directional (POD)**	Text	50	✔	commonly used
	StN_PosMod	**Street Name Post Modifier (POM)**	Text	25		commonly used
	AddNum_Pre	**Address number prefix (HNP)**	Text	15		commonly used
	Add_Number	**Address number (HNO)**	Long	6		always used
	AddNum_Suf	**Address number suffix (HNS)**	Text	15		commonly used
	LandmkPart	**Landmark Name Part (LMKP)**	Text	150		occasionally used
	LandmkName	**Landmark (LMK)**	Text	150		occasionally used
	Building	**Building (BLD)**	Text	75		commonly used
	Floor	**Floor (FLR)**	Text	75		commonly used
	Unit	**Unit (UNIT)**	Text	75		commonly used
	Room	**Room (ROOM)**	Text	75		rarely used
	Addtl_Loc	**Additional Location Info (LOC)**	Text	225		rarely used
	Milepost	**Milepost**	Text	50		rarely used
Location	Longitude	**Address Longitude**	Float	12		always used
	Latitude	**Address Latitude**	Float	11		always used
	NatGrid_Coord	**National Grid Coordinates**	Text	50		always used
Metadata	GUID	**GUID**	GUID			always used
	Addr_Type	**Address Type**	Text	50	✔	commonly used
	Placement	**Address Placement**	Text	25	✔	commonly used
	Source	**Address Source**	Text	75		always used
	AddAuth	**Address Authority**	Text	75		commonly used
	UniqWithin	**Unique Within**	Text	75		occasionally used
	LastUpdate	**Date Last Updated**	Date	26		always used
	Effective	**Effective Date**	Date	26		commonly used
	Expired	**Expiration Date**	Date	26		commonly used

NAD Schema Considerations: FGDC and CLDXF

While developing the NAD Schema, the project team looked closely at two of the most commonly used national address schemas - FGDC and CLDXF. The following concepts were critical when considering the NAD schema.

1. Address **parsing** is a key best practice

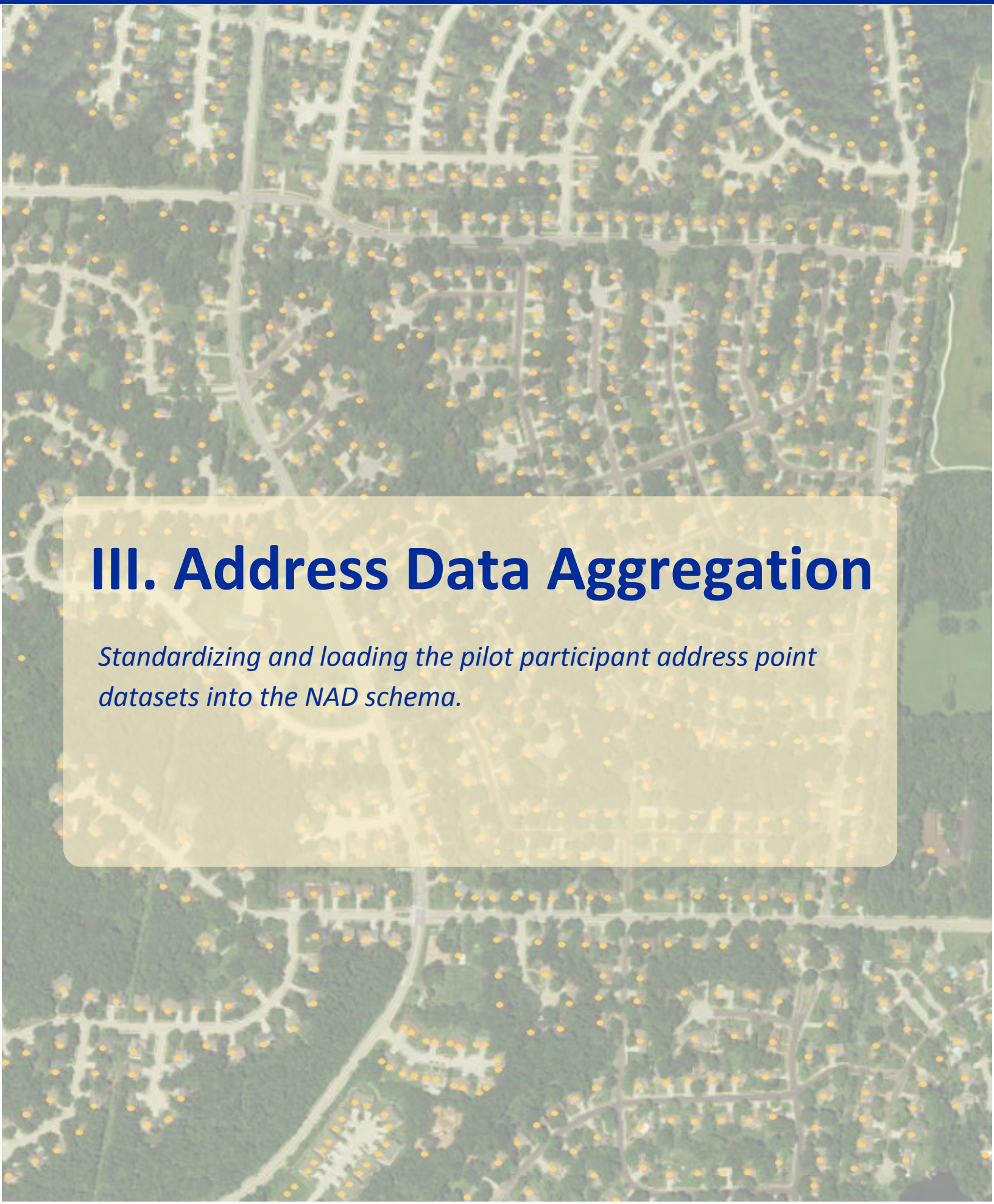
- FGDC and CLDXF standards identify a similar comprehensive parsing approach which was followed in the NAD Schema
- Non-parsed data that is submitted will be parsed during ingestion

2. There are however significant differences in the way that "**Place**" is handled in each schema, and in the way that **Subaddress** data is parsed.

- **Place**
  - In **FGDC**, "Place" is stored in related data element pairs. There is a "Place Name" (e.g., New York City) and a "Place Name Type" (e.g., Municipality). These pairs may be repeated to denote the hierarchy of "place" (e.g., the County, the Municipality, the Postal Community of an address). This method allows for more efficient storage in a relational tabular format and the greatest flexibility in terms of which "place" types are used or needed for a given address, as many types do not have to be used or could be used more than once.
  - The **CLDXF** standard separates each type of place (County, Municipality, etc) into hierarchical elements. This method allows easier storage in a flat-file tabular format but presumes the hierarchy is applicable to all addresses, allows one and only one value for each type of place name and assumes no other types of placenames are necessary or useful to describe the location. This method is simpler to implement and may be easier to maintain than the relational method described above.
- **Subaddress parsing**
  - In **FGDC**, subaddresses are stored in related data element pairs (e.g. value ="Eastman Cancer", type="Wing") in a similar manner to place names (described above). These pairs may be repeated to denote the hierarchy of subaddresses (building, floor, suite, desk, etc.). This method allows for more efficient storage in a relational tabular format and the greatest flexibility in terms of which subaddress types are used or needed for a given location, as many types do not have to be used or could be used more than once.
  - The **CLDXF** standard separates each type of subaddress (Building, Floor, Unit) into hierarchal elements. This method allows easier storage in a flat-file tabular format but presumes the hierarchy of subaddress type is applicable to all subaddresses, allows one and only one value for each type of subaddress and assumes no other types of subaddresses are necessary or useful to describe the location. This method is simpler to implement and may be easier to maintain than the relational method described above.

These differences are laid out in detail in a document titled "Profile Reconciling the FGDC *United States Thoroughfare, Landmark, and Postal Address Data Standard* and the NENA *Next Generation 9-1-1 (NG9-1-1) Civic Location Data Exchange Format (CLDXF) Standard"* (provisional draft available on the FGDC website).

For the NAD schema, in both cases, the project team and commenters preferred the intuitive nature of the CLDXF approach.



III. Address Data Aggregation

Standardizing and loading the pilot participant address point datasets into the NAD schema.

Address Parsing and Aggregation

As previously stated, the NAD is a roll-up, **harvested from authoritative data.**

A universally agreed-upon best practice is that data created and managed by authoritative sources should meet well documented standards. A key element of this is creating and maintaining a fully parsed database with field validation, etc.

However, the reality is that not all address aggregators use the same parsing schema. In many cases, a local schema is perfectly acceptable. Thus, one of the challenges of the NAD is to standardize and aggregate different parsing schemas into a single database.

The image below shows several state parsing schemas and how they compare.

Arizona	Arkansas	North Carolina	Utah	Virgina
AddNum_Pre				PREADDRNUM
Add_Number	adr_num	ADDR_HN	AddNum	ADDRNUM
AddNum_Suf	adr_num_suf		AddNumSuffix	ADDRNUMSUF
StN_PreMod				STREET_PREMOD
StN_PreDir	pre_dir	ADDR_PD	PrefixDir	STREET_PREFIX
StN_PreTyp				STREET_PRE_TYPE
StN_PreSep				
StreetName	pstr_name	ADDR_SN	StreetName	STREET_NAME
StN_PosTyp	pstr_type	ADDR_ST	StreetType	STREET_TYPE
StN_PosDir	psuf_dir	ADDR_SD	SuffixDir	STREET_SUFFIX
StN_PosMod	pstr_mod			

The ideal scenario is that a contributor's data is already using the NAD parsing schema for street address components and can go directly into the NAD, as is. However, if a data contributor is *not* using the NAD address parsing, they can either:

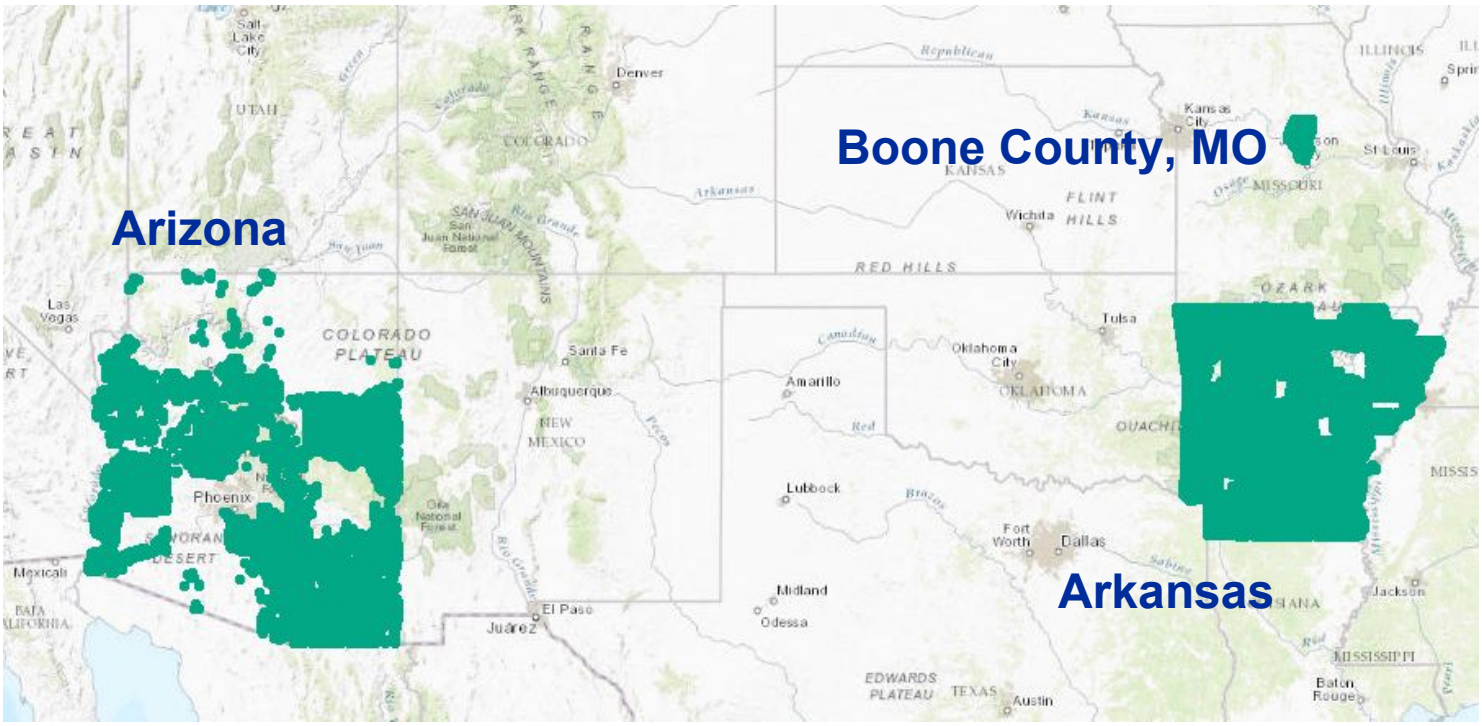
- Convert their data to use the NAD parsing schema through field mapping and ETL and submit to the NAD, or
- Submit non-parsed (concatenated) data which will then be run through an FGDC/CLDXF parser to format it for inclusion in the NAD. Submitted data that have issues being parsed or converted to the NAD format would be returned to the submitter for review. The reason for allowing this option is to create a low barrier to participation for an agency that does not already have NAD-parsed address data.

In terms of data responsibility, for the pilot, the project team has developed ETL tools for aggregating the pilot datasets (AZ, AR and Boone County). The additional data volunteers (NJ, DC, UT, OK, VA, OK, MO) developed their own ETL processes to migrate data into the NAD Schema. For the national implementation, it would be the responsibility of the NAD owner/manager to perform parsing and processing to match the NAD Schema if the source data providers are unable to provide it already in the NAD Schema. Resources will be necessary to build and maintain the right tools (e.g., for parsing, ETL, etc).

ETL = Extract, Transform and Load

ETL refers to a database process or set of processes that **extracts** data from one or more data sources, **transforms** the data to a standardized format, and then **loads** the data into the final database.

Aggregation of Pilot Address Data

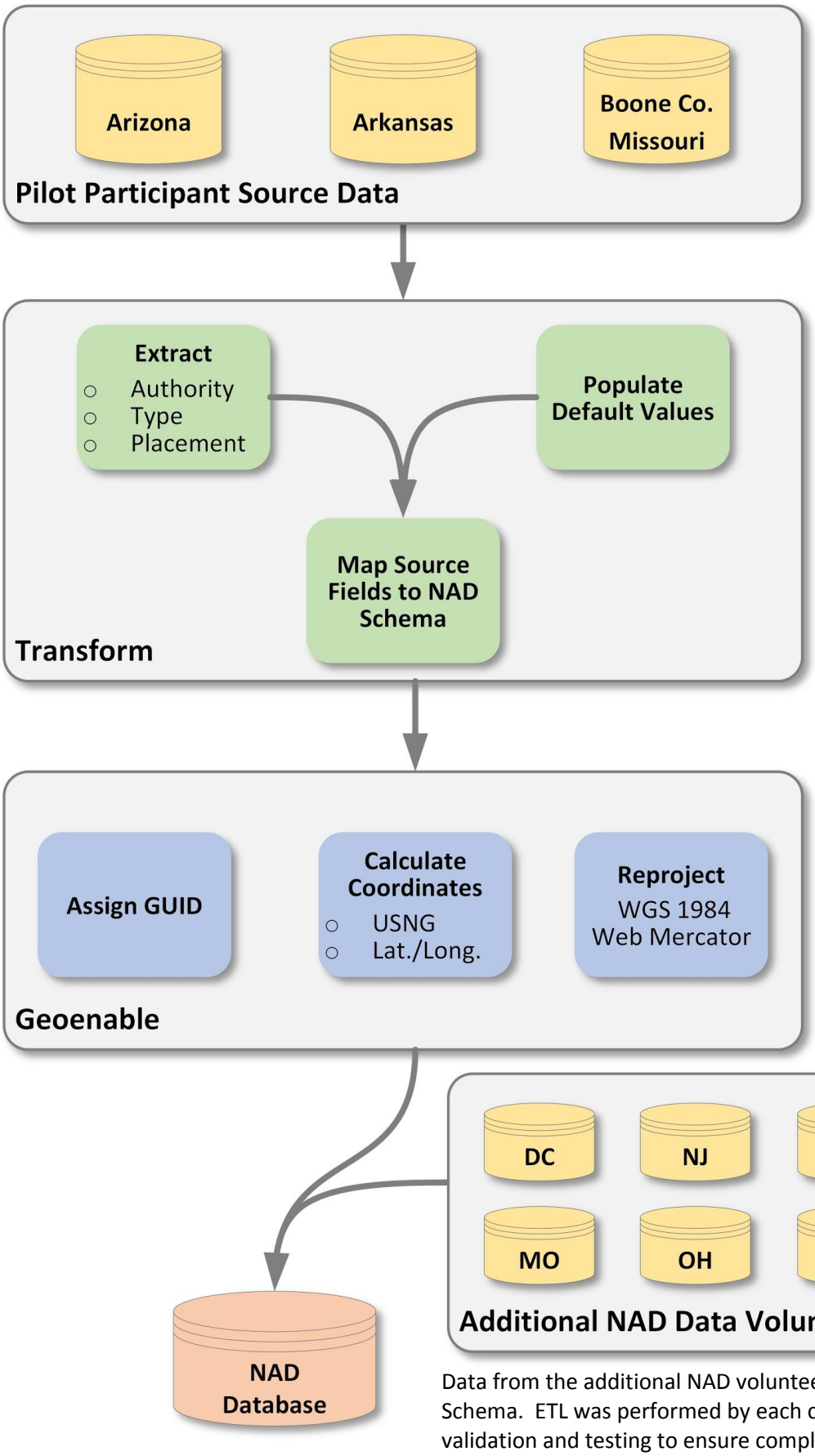


The three jurisdictions depicted above - the states of Arizona and Arkansas, and Boone County, MO - had existing address databases and were formal participants in the pilot project. Each contributed both their data and ideas to the project. All have agreed to have their data be publicly available following the pilot project. See Section VI, Participation and Data Sharing for further details on Arizona's challenges with public data release.

For the three participating pilot agencies, the following is a summary of the aggregation tasks completed during this pilot project:

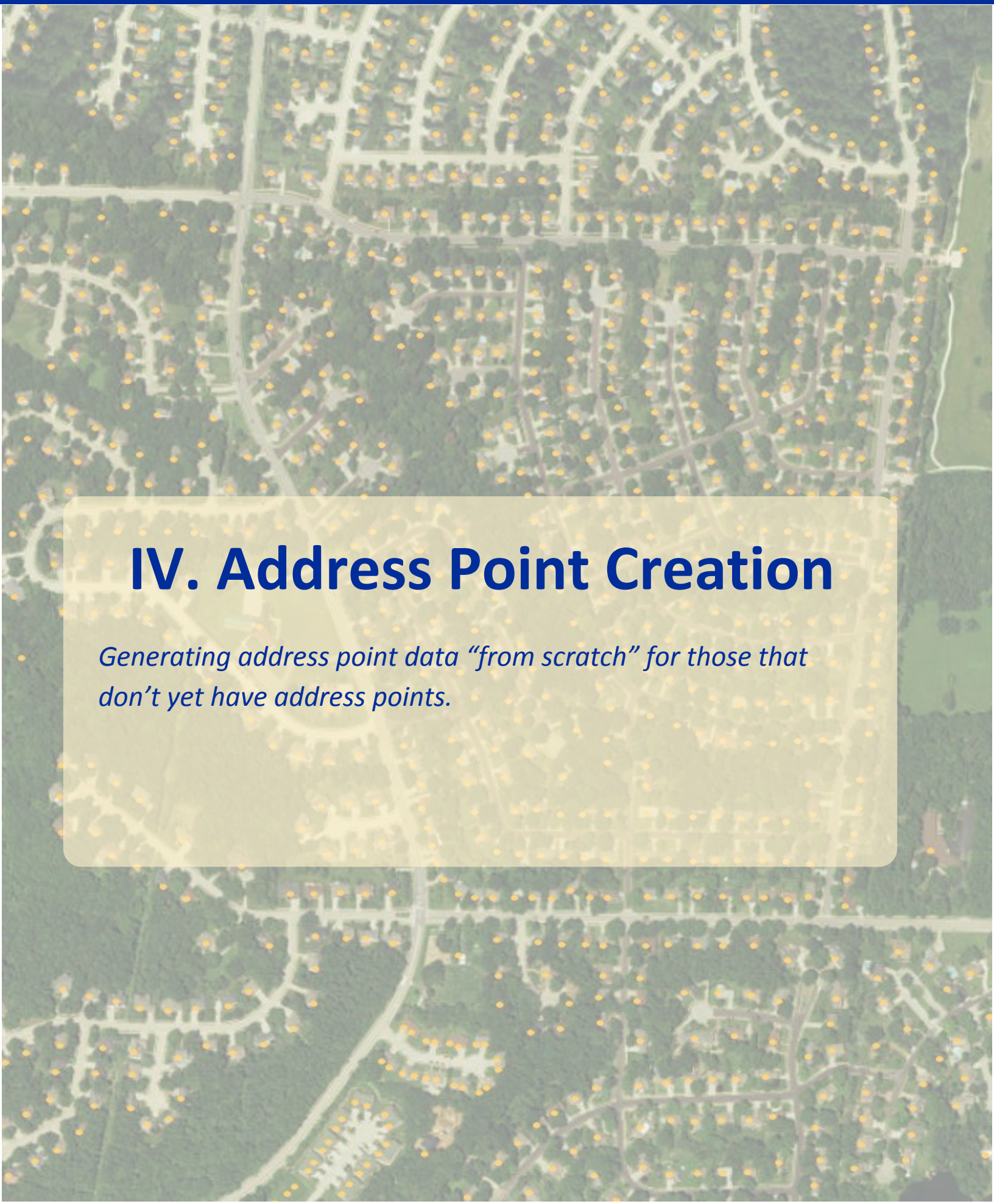
1. Address point data was collected from each source (Arizona, Arkansas and Boone County)
2. Data and metadata were reviewed, and the project team discussed internal workflows with data source contacts to understand the full picture of data creation and maintenance (see appendix 3 for more details)
    a. Note: address maintenance workflows and best practices were also discussed with other agencies, and these findings are documented in appendix 3.
3. The project team began field mapping work from source datasets to the NAD schema
4. Address source and authority details were identified, where possible
5. Reviewed any remaining data questions (e.g., translating source values for Address Type into NAD domain values)
6. Built ETL processes using FME (an integration and data manipulation software by *Safe Software*, https://www.safe.com/) to convert the source Esri Geodatabases into the NAD PostGIS database. See details of ETL workflow on next page. See appendix 7 for further details on some of the data standardization tasks that were involved in loading data into the NAD.

ETL Workflow



Note: see appendix 7 for additional details on data standardization and preparation for loading into the NAD.

Data from the additional NAD volunteers was submitted in the NAD Schema. ETL was performed by each contributor. The pilot team did some validation and testing to ensure compliance with the schema, reported any issues, and loaded each final source dataset into the NAD.

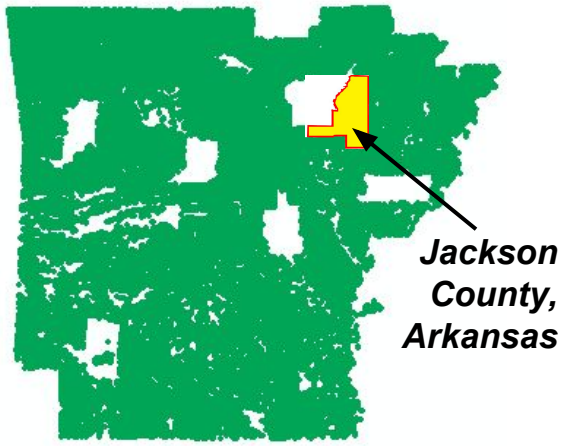


IV. Address Point Creation

Generating address point data "from scratch" for those that don't yet have address points.

Address Creation for Jackson County, AR

Finding suitable participants that didn't already have address data was an early challenge of this task. The goal was to find agencies that haven't yet created their address points, but were actively *interested, motivated*, and *willing to work* to improve and maintain the data after it was created. The NAD Project team did extensive outreach (at conferences, using existing networks, etc.) to find potential candidates.



Ultimately, Jackson County Arkansas proved to be the correct candidate. They needed address points, had a full database (E911) of all existing addresses, a centerline file with address ranges, and a county-wide parcel file with some addresses. They also had the interest, motivation, and willingness to work on data improvement after the data were delivered. In addition, as one of only 7 counties in Arkansas without address point data, the state was interested in their success. With these key ingredients, as well as support from the State GIS office, the NAD Pilot team was able to use the county as a test case for address *point* creation.

Jackson County Existing Source Datasets

1. Countywide E911 address list

- 18,469 total records (including some known duplicates)
  - Roughly 2,500 didn't have a City assigned
  - Roughly 2,800 did not have zip code assigned (thus only appx. 15,000 have a good chance of successful geocoding)
  - Issues with Address Number vs Address field inconsistencies
- Data Cleaning:
  - Address number & Address fields - standardized into concatenated field to remove inconsistencies in data capture
  - Stripped leading/trailing/extraneous spaces or '-' from addresses and zip codes

2. Countywide Parcels

- Parcel Addresses have varying levels of completeness - only 6,664 parcels had an address number
  - City Names are not official names, but indicate if parcel is inside or outside city bounds
  - Zip Codes - only 11 records had zip code information
- Data Cleaning:
  - Added fully populated State Abbreviation field
  - Concatenated all address elements (number, street name, street type) into a single field to remove inconsistencies in data capture
  - Stripped/replaced any leading/trailing spaces
  - Cleaned up City names where possible (which was critical since no Zip Codes were available)

3. Countywide Centerlines

- High quality reference data source - No data cleaning performed
- 861 segments missing primary street name as well as address data. Many of these appear to be cemetery, prison, or park roads.
- City Left, City Right and Zip Left, Zip Right codes fully populated

Jackson County, AR Geocoding Workflow

Once the project team assessed and cleaned the county's source data, the resultant address list was then geocoded using a variety of geocoding sources.

Due to the varying accuracy within each of the different geocoding data sources, the project team took the best results from each dataset and incorporated them into a "best point" dataset. Thus, the final dataset was based on the following hierarchy:

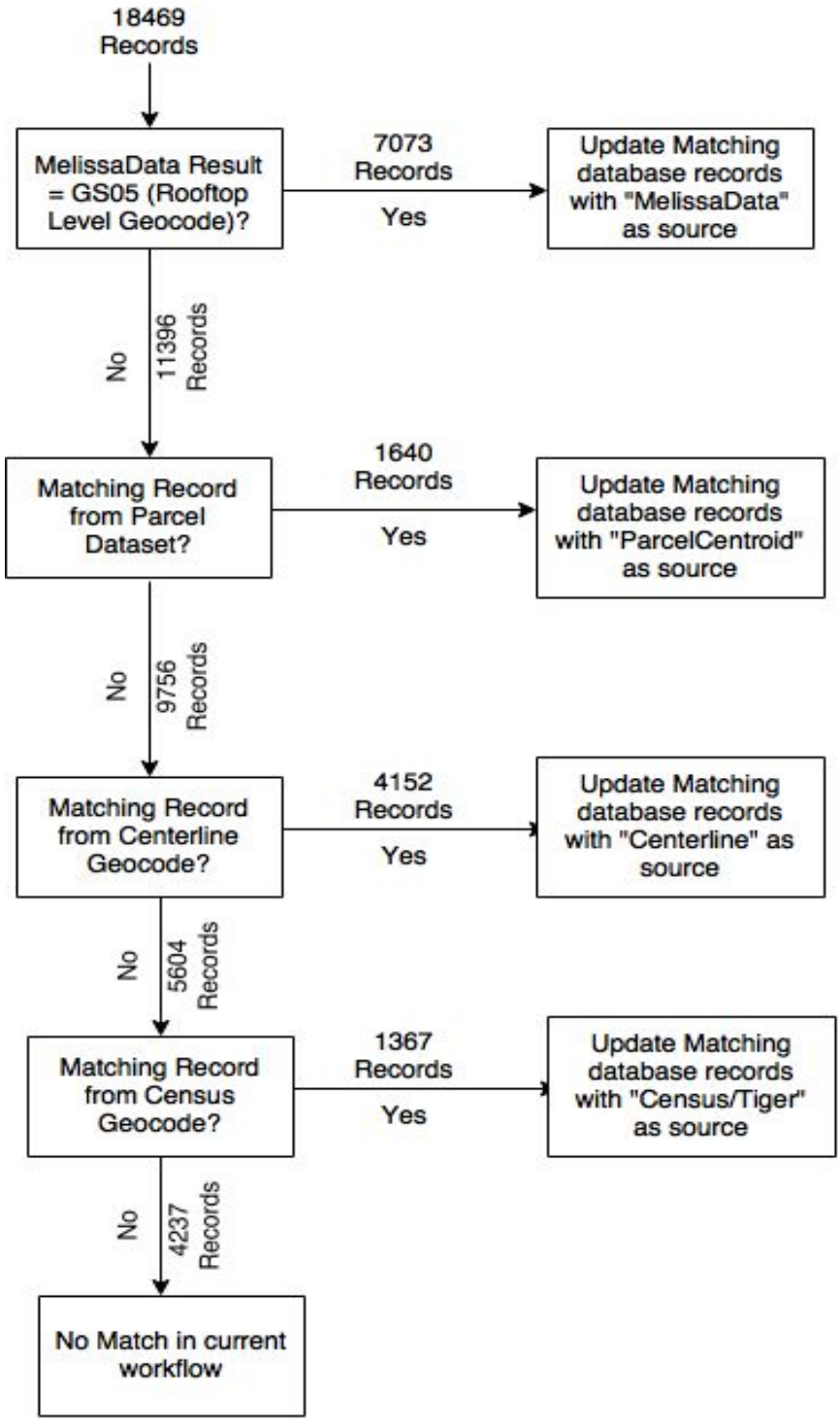
1. MelissaData (i.e., commercial source)
2. Parcels
3. County Centerlines
4. Census/TIGER Centerlines

The diagram at the right details this process, and the results of each step.

The entire address dataset was geocoded against all sources, and each address was tagged as to whether it got a match from each source, as well as the lat/long that was created when a match did exist.

Finally, for each E911 address, **the lat/long location that was derived from the "best possible" source was used in the final "best point" dataset.**

Overall, a **77% match rate was achieved** from 18,469 records

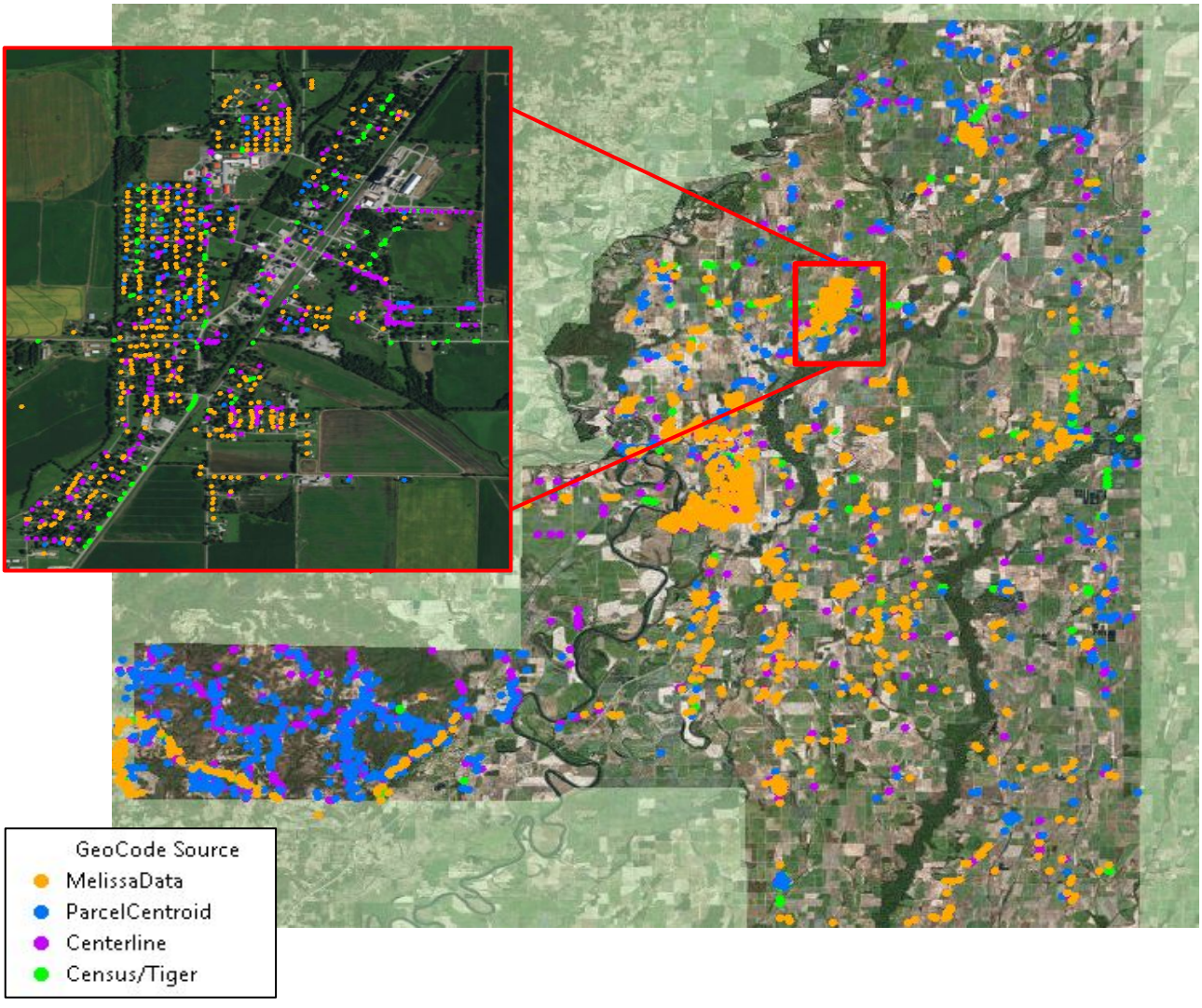


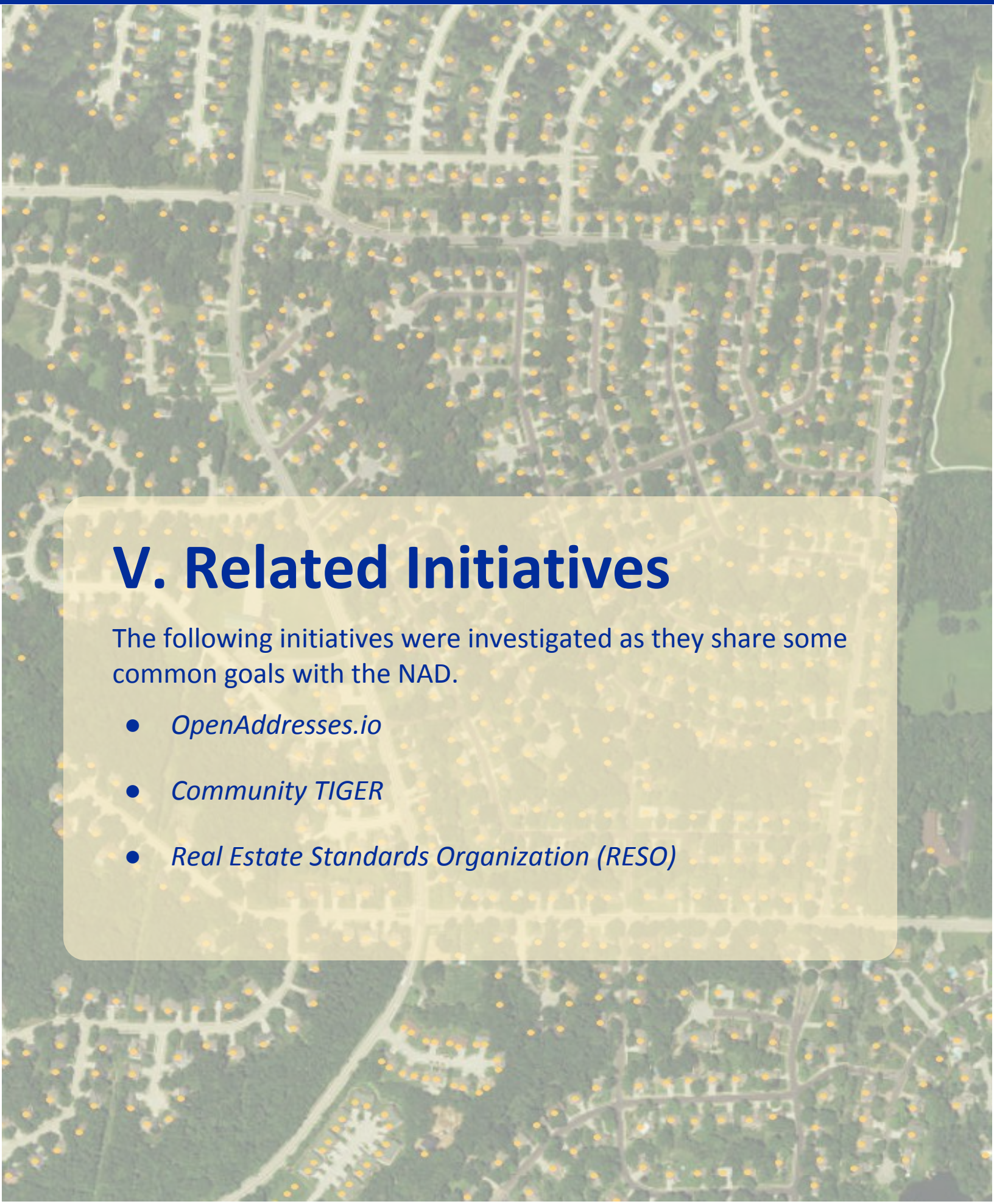
Geocoding Options
AppGeo assessed several geocoding options to enhance not only the match rate, but also the accuracy of the matches provided. These included *Google* (licensing terms of service put too many restrictions on data use) and *Nominatum* (the accuracy, particularly in rural areas was no better than the existing source county centerlines). AppGeo then found MelissaData, which provided "rooftop level" results and had very good turn-around time in processing (24 hours or less). MelissaData is a commercial data supplier that offers several address data-related services, including bulk geocoding.

Jackson County, AR Results

Source	Total Records Matched	% Matched*
MelissaData	7,073	38%
Parcel Centroids	1,700	9%
County Centerline	4,112	23%
Census/TIGER Centerlines	1,347	7%
Totals:	**14,232**	**77%**

*% matched of 18,469 total E911 records provided





V. Related Initiatives

The following initiatives were investigated as they share some common goals with the NAD.

- *OpenAddresses.io*
- *Community TIGER*
- *Real Estate Standards Organization (RESO)*

Open Addresses.io

Key Commonalities with the NAD Effort:

- Aimed at collecting and publishing open and freely available address point data
- Proves that it ***is*** very technically feasible to aggregate and publish nationwide addresses
- Adheres to minimum content concept

Key Differences from the NAD Effort:

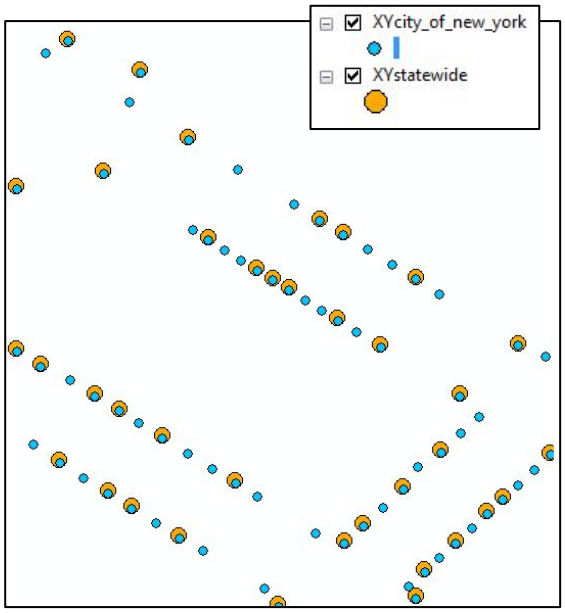
- NAD is collecting ONLY from authoritative sources
- NAD employs full address parsing and feature-level metadata tracking
- OpenAddresses.io is *global* in scope

OpenAddresses provides a global collection of **free and open** addresses that are harvested from a variety of data sources and represents a wealth of address data. Over **200 million addresses** are available worldwide, and are very easy to download and use.

Key Items for Users to be Aware of when using data from OpenAddresses.io (OA)

1. Overlapping Source Data

OpenAddresses pulls from a wide variety of sources, and in certain cases (e.g., data from overlapping jurisdictions) this could present some issues that users must understand when using the data. In the example at the right, data from both New York City and the State of New York overlap, and the user must determine the accuracy of each.



Example of overlapping data in OpenAddresses.io for New York City

2. Lacking feature-level metadata

The data schema itself does not include an attribute showing the data source or the date updated or cached, so users should understand where and how to access this information (i.e., elsewhere in the OpenAddress.io web-site, see appendix 8 for details) so that they are aware of where the data came from and how recently the data have been updated.

3. Non-Authoritative Data

Address data may have been harvested from somewhere other than the authoritative source. For example, the State of Massachusetts maintains a Master Address Database, but the data available through OpenAddresses was derived from the statewide *parcels*. While these parcels may contain good address information, they are not considered the authoritative address data source.

4. Data Lacking Essential Address information

- Many records are missing street address but instead contain PO boxes
- Many Records are missing Zip Code information, which would complicate using OA data for geocoding.

Community TIGER

Key Commonalities with the NAD Effort:

- Aimed at aggregating data from authoritative sources
- Based on nationally recognized address standards

Key Differences from the NAD Effort:

- NAD is in the public domain, not limited by Title 13 use restrictions
- NAD is focused on low barrier to participation; Census efforts are aimed at supporting the Decennial Census

Community TIGER is a set of cloud-based tools and processes (both desktop and web-based) that allow agencies to standardize and submit their address point and centerline data to the US Census Bureau. The goal is to have local governments managing and sharing their own data. The Community TIGER tools allow them to manage their data locally with their own, preferred tools and then submit to the Census when they wish where they will be translated to a variation of the Esri Local Government Data Model. The Census "cloud-based tools" provide access to customized Esri desktop software running on virtual machines within the cloud.

The data collection and aggregation aspect of Community TIGER is *outside* of the Title 13 environment. However, once the data has been ingested into the Census dataset, where the data may be improved through the Census enumeration process, the data are then be limited by Title 13 public access restrictions.

There are 3 main aspects of the Community TIGER Tool

- Desktop tools to create new addresses, as well as ETL existing addresses
- Desktop tools to standardize, de-duplicate, run quality checks, and perform batch data reviews on data. For address points, one of the "data cleaning" activities is to delete duplicate points. There are other geometry cleanup functions that are performed on the centerline data as well.
- Web-based dashboard for interrogating status, reviewing error messages, and securely transferring data

The NAD pilot project team was given a full demo of the Community TIGER tools by Census. Based on this demo, some initial observations of Community TIGER include:

- It's a somewhat complicated process that less technical participants might find challenging (i.e., familiarity with Esri desktop software is needed for some capabilities)
- It makes some assumptions that users may not be comfortable with. For example, the tools automatically perform "data cleansing" (i.e. topology cleanup and edits) without allowing the user to see what changes are actually being made
- The main benefit that we see is it provides an alternative to LUCA (Local Update of Census Addresses) for communities.

The NAD Pilot project team had attempted to get feedback on Community TIGER from actual participants in order to understand the user experience better, but has been unable to get any direct feedback to date. During the June, 2016 National Geospatial Advisory Committee meeting in Washington, DC, Tim Trainor of Census reported that there were "about a dozen, or so" active Community TIGER users.

Real Estate Standards Organization

Key Commonalities with the NAD Effort:

- Focus on national standards and best practices for data sharing
- Long-term vision for ongoing maintenance and upkeep

Key Differences from the NAD Effort:

- NAD is aimed at focusing on capturing addresses, which can vary from real estate property records
- NAD is only collecting data from authoritative sources (not commercial suppliers)

According to their website (http://www.reso.org/), "RESO actively develops, adopts and implements open and accepted data standards and processes across all real estate transactions." Their primary mission is to "To create and promote the adoption of standards that drive efficiency throughout the real estate industry."

RESO representatives and the NAD Pilot project team communicated throughout the pilot project to discuss shared goals and priorities in order to find any areas of potential overlap or mutual benefit. **The main focus of these conversations was around the shared need for unique identifiers for address and property records.** The NAD will require some type of unique ID for each address (i.e., a "globally unique ID", or GUID), and likewise, RESO is aiming to assign a unique identifier to all real estate properties.

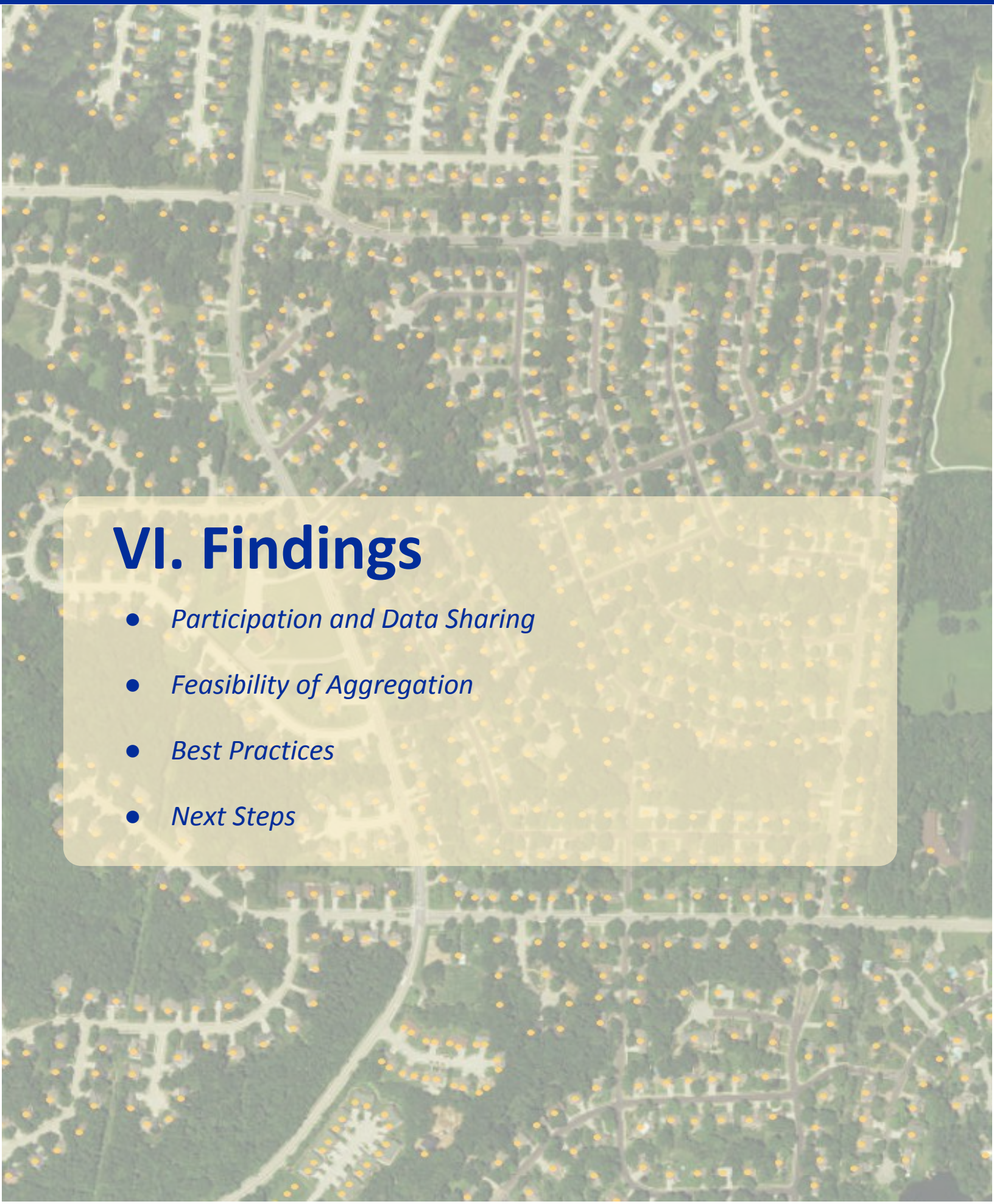
The following is a summary of the findings, based on discussions and communication with RESO:

- RESO's main motivation for a nationwide unique property ID is to answer two business cases:
  - De-duplicate real estate listings on sites such as Zillow and Realtor.com
  - Be able to create accurate property history
- RESO is working with commercial entities (CoreLogic http://www.corelogic.com/ and Black Knight Financial services (http://www.bkfs.com/data-and-analytics/Pages/default.aspx) to acquire nationwide property information
- RESO is currently considering how such a unique identifier might be generated (e.g., randomly, like a GUID, or somewhat "human consumable", such as a combo of lat/long or other variables).
- The ultimate goal is to develop and implement an open web service or API that would allow the public (and sites such as Zillow) to get access to the Unique ID. Methods for ID lookup might be
  - by municipality + parcel ID
  - by address
  - by lat/long

NAD Addresses vs RESO Properties

In some cases, the NAD address corresponds directly with a real estate property (e.g., single family homes, condos, etc.). However, there is not always a one-to-one match between address and property. For example, an apartment building may be a single record from a real estate property perspective (it would be owned, bought, sold as a single entity), but the same apartment building would likely have multiple address records in the NAD - i.e., one "subaddress" for each unit. Further, a vacant tract of land may have no address assigned to it, but it would represent a real estate property.

For these reasons, the RESO unique ID *won't* replace the need for a GUID in the NAD. But, it could be something that NAD consumes (i.e., pulls into the data via the planned RESO web service) if it ends up being a valuable link to other datasets.



VI. Findings

- *Participation and Data Sharing*
- *Feasibility of Aggregation*
- *Best Practices*
- *Next Steps*

Findings Summary

Further details are found on the following pages

Participation and Data Sharing

- Tribal participation is likely to be a challenge
- Data sharing agreements to keep NAD data publically available may also be a challenge

Feasibility of Aggregation

- Developing the NAD is technically feasible but will involve some challenges
- The schema will evolve, but needs to remain consistent with leading address standards and schemas to allow for streamlined ETL
- Aggregating existing statewide/“have” collections was straightforward for pilot datasets, but will become more complex as more participants with varying levels of technical capability get involved

Data Creation and Maintenance

- Data Creation depends largely on quality of source data
- Communities without existing address data may need funding support to get started

Next Steps

- Continued education and outreach on the NAD effort
- Standing up a preliminary NAD database within the USDOT and Establishing the ingest process and acceptance criteria
- Outreach to existing statewide aggregators to obtain voluntary contributions to the NAD
- Tackling regular data updating within the NAD via recurring contributors
- Identifying funding/a grant program for address data creation that can be made available to jurisdictions that do not yet have address data

Participation and Data Sharing

The biggest challenge faced by participants that already have address data is that of data sharing. While some states, like Arkansas, have a culture and history of freely shared data, including address data, some of the agencies with existing address data that were involved in this project were not able to fully share their data to the public. For example, in Arizona, the statewide data were assembled for specific uses under agreements with the various contributing jurisdictions and address authorities (e.g., counties). And, in some cases these entities did not allow their contributed data to be further shared under those agreements, and thus it has been left out of the NAD.

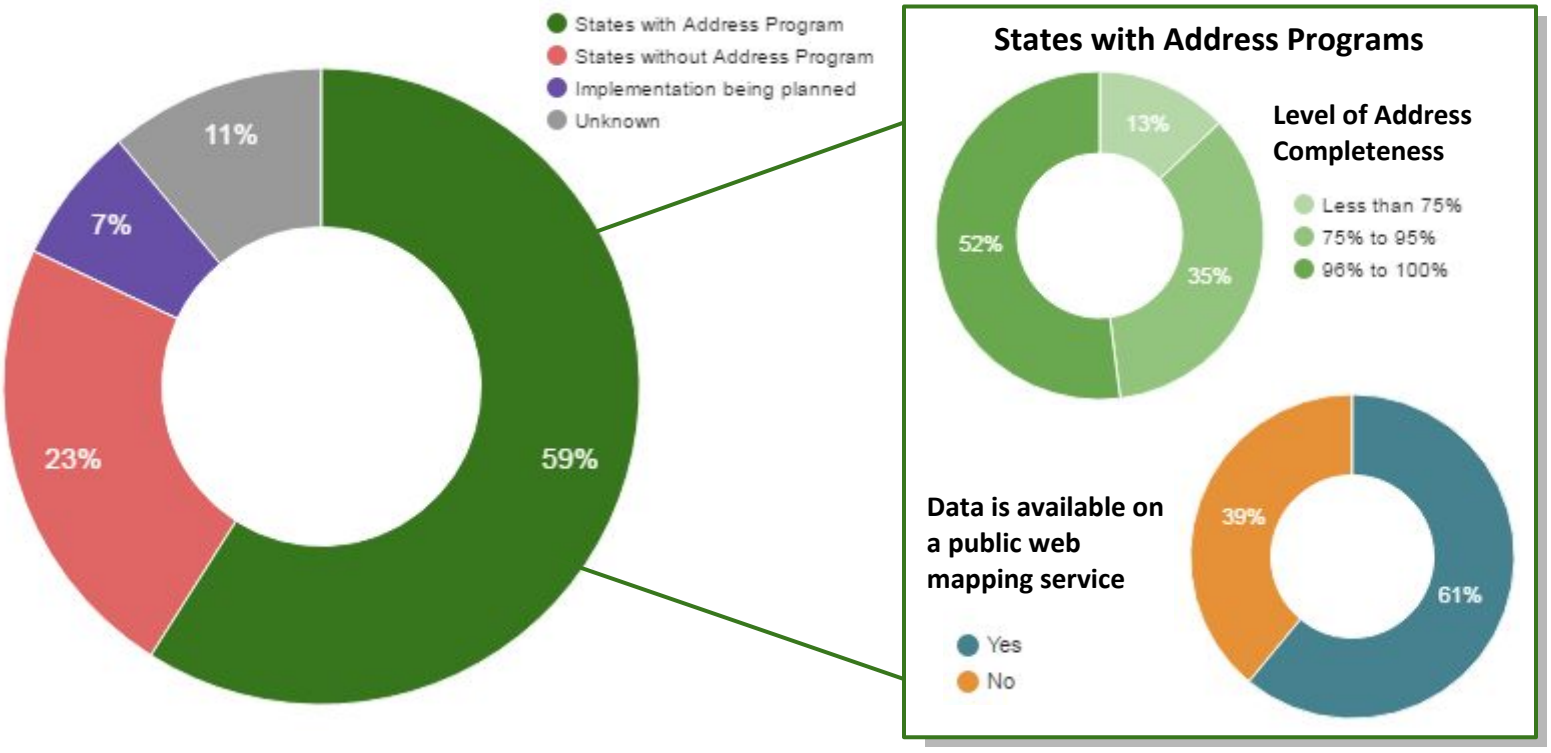
This will likely be a larger issue as the NAD moves from a pilot phase into a fuller roll-out. It should be expected that some states/counties/tribal agencies will have data sharing policies that may not allow public sharing. Ideally these more closed policies will begin to open up in the coming years as the NAD gains momentum and as large numbers of agencies show a willingness to share openly.

Similarly, the project team conducted extensive outreach to several tribal entities and authorities in an attempt to obtain some tribal address data for the NAD pilot. Outreach efforts included:

- Coordination with National Tribal Geographic Information Support Center (NTGISC) and their members
- Participation in the NTGISC conference
- Direct outreach to the Gila River Nation
- Direct outreach to the Navajo Nation

In the case of Arizona, tribal data *was* included in their statewide address data and will be included in the NAD pilot dataset. However, direct outreach and communication with the tribal entities was difficult and concerns about tribal data sharing will likely be a recurring theme as the NAD matures.

According to the 2015 NSGIC Geospatial Maturity Assessment, 31 respondents (out of 53) have a state-level address program. It will be very critical to work with these states to get their participation in the NAD. For the states without an address program, the NAD may provide an important forum for encouraging state-level organization and best practices.



Feasibility of Aggregation

- **Developing the NAD is technically feasible but will involve some challenges**
- **The schema will evolve**, but needs to remain consistent with leading address schemas to allow for streamlined ETL
- **Aggregating existing statewide collections was straightforward for pilot datasets**, but will become more complex as more participants with varying levels of technical capability get involved
  - In addition to the three pilot participants (AR, AZ and Boone County), the following additional contributors have volunteered to ETL their own data for inclusion in the pilot NAD database. These include
    - Virginia Geographic Information Network (VGIN)
    - New Jersey Office of GIS
    - Utah AGRC (Automated Geographic Reference Center)
    - Washington DC - Office of the Chief Technology Officer (OCTO)
    - Missouri (9 counties, 1 City): aggregated by Jason Warzinik, GIS Manager of Boone County, MO
    - Ohio (80 of 88 Counties) compiled by Jeff Smith, OSDI Manager of the Ohio Geographically Referenced Information Program.

Dan Widner, VGIN Coordinator: "The ETL was easy and it is repeatable in the future. We publish on a calendar quarterly basis so let us know when you want updates."

The final pilot NAD has a total of 16.8 million records across 8 states

- A key challenge will be in maintaining longer term **State-Local coordination and partnerships so that the statewide aggregations can remain updated**. In addition, these partnerships may need to go beyond simply exchanging data and involved providing technical assistance to locals, especially the smaller less technically enabled communities. This will be essential for long term maintenance and update of both statewide data and the NAD.
- Over time, additional **resources will be needed to build and maintain an expanding array of validation and ingestion tools as well as parsers and data update tool**s. As more kinds of data are contributed the sophistication of the tooling will expand and that tooling will also need to support updating from recurring contributors.

Shelby Johnson, GIO of Arkansas:
"Personnel are required to coordinate, hold hands, and help with the regional / state integration. Building a system based on a technological, web-based, online, pie in the sky pipe dream is a fool's errand. No two jurisdictions are the same, no two addressing intervals are the same and seldom do local naming conventions follow any convention. For those reasons and the fact that addressing authority personnel turnover is always a constant... a regional or state integrator MUST have a minimum capability to provide technical support, guidance and assistance."

Best Practices for Data Creation and Maintenance

Data Creation

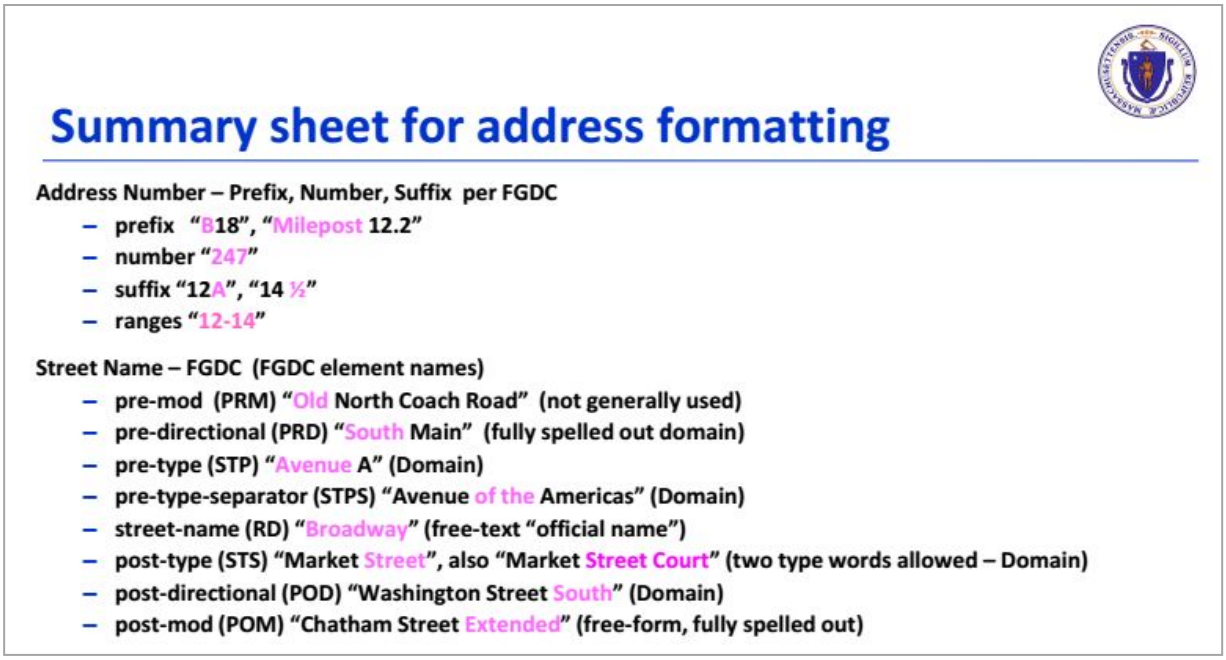
- New address data creation depends largely on quality of source data
- More scrubbing and standardization of the source E911 address list before geocoding may have yielded better results
- Local knowledge and field work is an absolute necessity to review and verify the existing matched addresses, and to locate the unmatched addresses.
- Address creation that begins with geocoding of 911 lists is *only a start.* Entities that pursue this approach should expect a need for further field validation to verify the geocoding locations and to identify addresses that were not geocoded.

Data Maintenance

As described above, the NAD is a roll-up of authoritative data. In order for the NAD to be a quality dataset, the information coming from the data contributors needs to follow certain guidelines. Below are two key best practices. (See appendix 9 for some advanced address database characteristics)

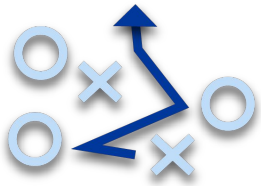
In general, addresses should:

- Not use abbreviations. For example, spell out "Street" (not "St.") and spell out "Road" (not "Rd."). This helps to ensure there is no ambiguity in addresses.
- Utilize full parsing of address information to allow for validation and use of domains (e.g., for street type and direction). See image below for parsing of street number and street name, using the FGDC parsing schema.



CLDXF Parsing (image credit: Christian Jacqz, MassGIS)

Next Steps



- **Continued education and outreach on the NAD effort**

One of the key findings from the NAD summit was that there had been “enough talking” and it was time to start “doing.” This pilot represents the beginning of the “doing” and it will be important to make that message clear and to encourage more activity.

In addition, during the project timeframe the FGDC created a new “addressing theme” as part of the National Spatial Data Infrastructure (NSDI). This is the first expansion of NSDI themes and USDOT and Census were identified as co-theme leads. As such, both USDOT and Census have cause to work with Federal agencies to advance the NAD.

- **Standing up a preliminary NAD database within the USDOT and Establishing the ingest process and acceptance criteria**

The USDOT has made a commitment to stand-up a cloud-based environment to house the pilot NAD database created through this project as well as make continued efforts to find additional data contributors. As this effort takes root, USDOT will work to make the data available and to expand the ETL and ingestion tooling necessary to process contributions (e.g., acceptance/rejection criteria, compliance testing and reporting).

- **Outreach to existing statewide aggregators and individual address authorities to obtain voluntary contributions to the NAD**

Explicit data outreach efforts should be initiated to continue to build the list of address data contributors to the NAD. Given the increasing number of states that are assembling statewide data sets outreach should begin there to obtain the most data via a single contribution. Also, further outreach should be made to individual counties that have mature address programs in states that don’t yet have statewide programs in place.

- **Tackling regular data updating within the NAD via recurring contributors**

The initial efforts of the NAD pilot are focused on *assembling* a national address data aggregation. But, due to regular changes in local address data, there needs to be an ongoing refresh program to ensure the contents of the NAD are accurate and current. As ingestion and ETL routines for data assembly begin to mature, it will be important to begin addressing the complexities of data refresh from contributors, and also mechanisms for users of the NAD to communicate issues/errors with NAD data back to the data originators.

- **Identifying funding/a grant program for address data creation that can be made available to jurisdictions that do not yet have address data**

The work done with Jackson County, AR was a key part of the pilot project and demonstrated that there are cost effective means of getting “have not” communities started with address point data creation. Encouraging and supporting the development of more address point data is an important component of a full NAD and identifying funding support to get smaller, less technologically advanced places started would support the long term success of the NAD.

VII. Appendices

List of Appendices

1. NAD Pilot Advisory Group Participants
2. Important Related Materials
3. Address Entity Interview Findings
    a. State of Arkansas
    b. State of Arizona
    c. Boone County, Missouri
    d. City of Meridian, Idaho
    e. East Baton Rouge Parish, Louisiana
4. NAD Minimum Content Feedback
5. Proposed NAD Schema
6. Proposed NAD Schema Domains
7. Data Preparation and Standardization for Loading into the NAD
8. OpenAddresses.io Data Source Information
9. Advanced Address Database Considerations

Appendix 1

NAD Pilot Advisory Group Participants

The following NSGIC members participated in regular NAD Pilot Project status meetings and provided early feedback on the Minimum Content Approach, NAD Schema, etc.

Many thanks for these folks for participating and lending their expertise!

Bert Granberg, State of Utah

Shelby Johnson, State of Arkansas

Curtis Pulford, State of Arizona

Andy Rowan, State of NJ

Cy Smith, State of Oregon

Timothy Trainor, US Census

Gene Trobia , NSGIC

Appendix 2

Important Related Materials

National Geospatial Advisory Committee (NGAC): The Need for a National Address Database (December 2012)
https://www.fgdc.gov/ngac/meetings/december-2012/NGAC%20National%20Address%20Database%20Paper.pdf

National Geospatial Geospatial Advisory Committee (NGAC): The Need for a National Address Database - Use Cases (December 2014)
https://www.fgdc.gov/ngac/meetings/december-2014/ngac-national-address-database-use-case-paper-december-2014.pdf

National States Geographic Information Council (NSGIC): A National Address Point Database Will Improve Government Services (May 2014)
https://www.nsgic.org/public_resources/Address_Point_Database_Values_V1_051914.pdf

US Government Accountability Office Report to Congressional Requesters: GEOSPATIAL DATA - Progress Needed on Identifying Expenditures, Building and Utilizing a Data Infrastructure, and Reducing Duplicative Efforts (February 2015)
http://www.gao.gov/assets/670/668493.pdf

National Address Database Summit Report (June 2015)
https://sites.google.com/a/appgeo.com/nationaladdressdatasummit/home/presentationsandreports/NAD_Summit_Report.pdf?attredirects=0&d=1

FGDC United States Thoroughfare, Landmark, and Postal Address Data Standard (February 2011)
http://www.fgdc.gov/standards/projects/address-data/index_html

NENA NENA Next Generation 9-1-1 (NG9-1-1) United States Civic Location Data Exchange Format (CLDXF) Standard (March 2014)
https://www.nena.org/?NG911CLDXF

NENA Development of Site/Structure Address Point GIS Data for 9-1-1 (September 2015)
https://www.nena.org/?SSAP

Position Paper for FGDC Theme on Addresses (April 2016)
https://www.fgdc.gov/organization/steering-committee/meeting-minutes/april-2016/position-paper-fgdc-theme-on-addresses-dot-sc.pdf

Appendix 3

Address Entity Interview Findings

The sections below contain summaries of the findings based on discussions and communication with pilot project participants. The first three (Arkansas, Arizona and Boone County) provided data for the NAD pilot data project. The final two (City of Meridian, ID and East Baton Rouge Parish, LA) participated in phone interviews to share their unique experiences and lessons learned related to address data collaboration, creation and maintenance.

State of Arkansas

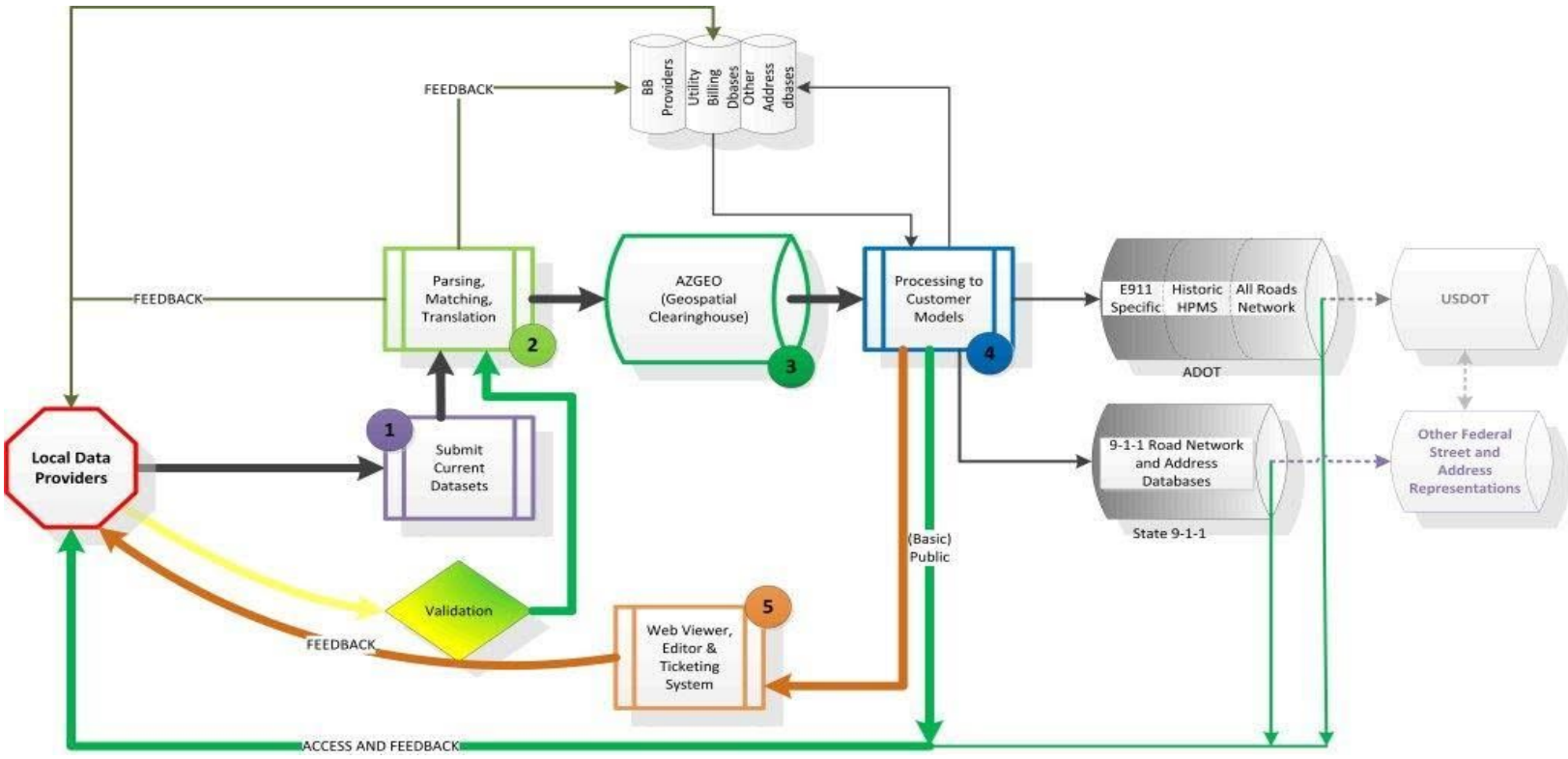
- Contact: Shelby Johnson, Geographic Information Officer, Arkansas GIS Office
- Most counties are using the State's standard
- Data comes mostly from counties; but some cities (e.g. Jonesboro) handle the data themselves.
- Refresh frequency varies. Some counties refresh their data monthly, some annually.
- Counties have different placement of points; buildings vs. access to road.
- There is current legislation effort to build a statewide address DB.
- The state uses FME for ETL of the county data into the statewide dataset.
- Regarding AR counties that do not have address data yet, they have found that the key is for that county to be interested and motivated, otherwise they are just creating data that no one will take ownership of.

State of Arizona

- Contact: Curtis Pulford, State Cartographer, Arizona State Land Department
- Arizona is made up of 15 counties + tribal lands
- Typically, data is updated on a quarterly basis from custodians
- The AZ schema is based on the NENA standard plus unique, homegrown sub-addressing (bldg+floor+unit).
- They have an emphasis on metadata, including a minimum metadata standard
- The original funding for address data came from 9-1-1. From there, specific accommodations for DOT-related needs were added. The system was designed to be extensible in consideration of other point/line/polygon features. The goal of their approach is for a state clearinghouse to efficiently partner with custodians and disseminate data for multiple purposes.
- Arizona's address data is currently not publicly available, so while they provided a copy of their data for the purposes of the pilot project, it can not be shared publicaly in the NAD. However, the state is currently editing their data request document to notify custodians that the address data will be used additionally for NAD purposes. At that point, the data will hopefully be openly available.
- Workflow:
  - The state takes data as-is from the counties. The main reason for this is so that the custodians will not have to change their business workflows
  - The state then applies their data parsing workflow to put county data into the state schema (using FME for automated ETL)
    - Note: FME-Pro is given to local 9-1-1 authorities for aggregating county data from local sources
  - After the data are migrated to the state schema, the state then gives the data back to the county in the standardized format and encourages schema adoption.
  - See detailed workflow diagram below (provided by AZ):

Address Entity Interview Findings (cont.)

Arizona Address Database Workflow



1. To obtain local agency participation, requires definition of the program (standards and target schema), refinement of program components, development of agreements among participating parties, on-going review and refinement of program.

2. Local data is run through a series of programs and procedures to quality check, using other databases (e.g. ALI, Broadband, utility and others) as available.

3. AZGEO s a secure Clearinghouse with naming conventions, written procedures, intelligent web services with built in criteria checks and meta data tracking.

4. (in development) Processes through Spatial ETL, pushed to public and customer specific models

5. (in development) Web viewer, tied to custodian ticketing system, for adding/deleting, moving or editing individual record attributes

Images courtesy Curtis Pulford, State of Arizona

Address Entity Interview Findings (cont.)

Boone County, Missouri

- Contact: Jason Warzinik, GIS Manager of Boone County, MO
- The dataset is shared between Boone County and The City of Columbia, each of which has a specific process for assigning addresses within their jurisdiction
- Issue reporting is handled through online viewer
- Address point placement is reviewed periodically by both the County and City as new imagery becomes available
- Addresses are added on an ongoing basis.

City of Meridian, Idaho

- Contact: Matt Tenold, GIS Analyst, City of Meridian, ID
- Address data cleanup has been a major focused effort
- Began with getting buy-in from all City departments. Use started slowly but began to increase as trust in the dataset grew.
- The initial address data cleanup initiative started by combining addresses from various sources (Utility billing, department spreadsheets, GIS layers, public safety data, etc.)
- Address verification and cleanup was largely a manual process, performed initially via SQL queries to remove duplicates. The dataset began with nearly 50k records and was scrubbed down to 34k.
- This work took 2 people nearly 4 months of dedicated effort to clean up discrepancies and remove duplicate addresses.
- A desktop application was created for address maintenance as well as a mobile application for collecting and verifying address data in the field
- Key lessons learned:
    - The cleanup effort needed to be a collaborative process city departments (highway, DPW street signs, Public Safety, etc.)
    - Legacy Addresses that don't follow "the rules" need to be accounted for and kept
    - City of Meridian, City of Boise and Ada County (which contains both cities) all use the same schema to ease data sharing

Address Entity Interview Findings (cont.)

East Baton Rouge Parish, Louisiana

- Contacts:
  - Joe Thompson, Chief Applications Administrator, 911 Data Management Lead, East Baton Rouge Parish, Louisiana
  - Jim Mitchell, Geospatial Services Manager, State of Louisiana Office of Technology Services
- Louisiana has no statewide 911. Every parish manages addressing and 911 independently, which is good and bad. It allows for some local flexibility, but hinders cross-parish data sharing and response. The NAD will be a great source for emergency response activities when dealing with mutual aid to neighboring parishes.
- Address data cleanup in East Baton Rouge parish for NG911 has been a major focus. There are 30+ agencies within the parish that need to be coordinated with (local fire departments, EMS, public safety, universities, etc.). All of the municipalities were handling addresses differently and have varying address ordinances. Data was collected from the local municipalities and standardized and aggregated up to the parish-wide dataset.
- They are using an add-on to Esri desktop called *Address Quality Extension* (from InfoGeographics, Inc.) to perform data review and cleanup. It allows them to compare MSAG data to GIS data to 911 Dispatch data to identify and remove inconsistencies.
- In some cases, re-addressing is needed to ensure emergency response can find a house. Some legacy addresses simply don't make sense (numbering skips around as you drive along the road, making it difficult to find an address). To make this happen, the E911 office would communicate directly with the homeowners to explain why re-addressing their house was critical for their safety in an emergency response situation.
- A Civic Address Committee maintains a centralized database to track addresses across the parish to ensure addresses are unique. The ultimate goal is not duplicate street names within the parish. Existing duplicate names are unlikely to be changed except when absolutely necessary. Going forward, new street names cannot be a duplicate of a street that already exists within the parish.

Appendix 4

NAD Minimum Content Feedback

When the Minimum Content Guidelines were shared with various groups (NAD Advisory Group, NAD Summit Participants), the project team received a lot of feedback. Most of the feedback was positive with some negative input as well. Below is an except of some of this feedback, and notes about whether, or not, changes in the approach were made.

Examples of feedback that ***did*** lead to improvements/changes in the minimum content include:

- The project team was urged to not characterize the leading standards as a competition of CLDXF vs. FGDC. Rather, the feedback urged that it be acknowledged that the standards differ for purposeful reasons while sharing many of the same qualities.
- Similarly, reviewers urged that calling CLDXF "more flexible" was not accurate. Indeed, the report now acknowledges that CLDXF is actually more rigid, and instead describes that it is perhaps more intuitive to many address users.
- Reviewers observed that that the introductory text clearly identify that the NAD pilot schema is for storing the aggregated addresses in the pilot NAD, and to make clear that it is not minimum requirement for submitting data to the NAD.
- It was observed that the "Seat" data element in CLDXF goes largely unused and thus many reviewers suggested that it should be removed from the NAD Schema.
- In addition to including an "address type" (i.e., residential, commercial, etc.) data element, it was also suggested that the minimum content guideline also include an optional field to capture "address placement method" (i.e., building centroid, parcel centroid, property access, etc. see Address Placement Examples).

NAD Minimum Content Feedback (cont.)

Critiques that were ***not*** incorporated into the document for various reasons include:

- *Comment:* The schema should support common local government requirements including property tax management, utilities, elections/voters, schools, planning and development, inspections, etc.
- *Response:* In the NAD aggregation model, local authorities can choose to use FGDC (or whatever model they would like) to support their specific business needs. The NAD aggregation is not designed to support all of these specific use cases. Rather, it is optimized to support the construction of the largest, lowest-common-denominator national aggregation. We expect that the authoritative source data, collected from local authorities will be used to support those specific local government activities.

- *Comment:* In order for the NAD to support the normalization of data from many sources, it will need a complex schema.
- *Response:* Extract-transform-load (ETL) work to standardize the data can be done prior to loading data into the NAD. We are intentionally avoiding a complex schema to maintain the lowest possible barrier to participation. It is not intended that data will be *managed* and *maintained* inside the NAD. Rather, the NAD schema will house data that are *aggregated* and then *shared* out for public use. The NAD will have database fields to provide a reference back to the original authoritative sources (i.e. feature-level metadata to allow users to know exactly where the data came from and when).

- *Comment:* Why wouldn't the NAD simply use the FGDC standard?
- *Response:* We acknowledge that the FGDC is a valuable standard and one that we hope will be used by many addressing authorities. But the current reality is that the FGDC standard is not universally used by this community right now, and in fact many address authorities and states are choosing aspects of the CLDXF standard. It is also worth noting that a large majority of the NAD Summit participants and NAD document reviewers voiced a preference for CLDXF handling of places and sub-addresses. Ultimately, the NAD approach is strongly supportive of the use of the FGDC standard by address authorities for address creation and maintenance. This will help to insure good addresses are created and managed by locals and once harvested and aggregated, those data will be part of the NAD.

Appendix 5

Proposed NAD Schema

The section contains the proposed (v1) NAD Schema for storing data in the NAD, including all attribute fields and domains. This schema is in line with the NAD Minimum Content Guidance laid out in section II. It is anticipated that this schema will evolve over time as the NAD becomes a reality.

Please note that this is not a minimum schema for submission - it is anticipated that many of the fields may be null for many records, or perhaps nonexistent depending on the source data. The "Expected Use" column indicates whether a field is generally "always used", "commonly used", "occasionally used", or "rarely used" within a dataset to give submitters a sense of expectations. (Note: a "commonly used" attribute may be null for the majority of records, but is still likely to be utilized within an address database). The proposed schema is a single flat table (non-relational) based largely on the NENA CLDXF standard (http://c.ymcdn.com/sites/www.nena.org/resource/resmgr/Standards/NENA-INF-014.1-2015_SSAP_INF.pdf).

Field Name	Field Alias	Type	Length	Domain	Description	Expected Use
State	State	text	2	Domain	Name of the state or state equivalent	always used
County	County	text	40	Domain	Name of county or county-equivalent where the address is located	always used
Inc_Muni	Incorporated Municipality	text	100	None	Name of the incorporated municipality or other general -purpose local governmental unit where the address is located	commonly used
Uninc_Comm	Unincorporated Community	text	100	None	Name of an unincorporated community, either within an incorporated municipality or in an unincorporated portion of a county, or both where the address is located.	commonly used
Nbrhd_Comm	Neighborhood Community	text	100	None	Name of an unincorporated neighborhood, subdivision or area, either within an incorporated municipality or in an unincorporated portion of a county or both, where the address is located.	commonly used
Post_Comm	Postal Community Name	Text	40	none	A city name for the ZIP code of an address, as given in the USPS City State file	commonly used
Zip_Code	ZIP Code	Text	7	none	For standard street mail delivery (with a corresponding geographic delivery area), the system of 5-digit codes that identifies the individual USPS Post Office associated with an address.	always used
Plus_4	Zip Plus 4 Addition	text	7	none	The ZIP plus 4 code (without the dash) Example: 1234	occasionally used
Bulk_Zip	Bulk Delivery ZIP Code	Text	7	none	For Bulk Delivery (e.g., government mailroom) zip codes with no corresponding geographical area, the system of 5-digit codes that identifies the individual delivery location associated with an address.	rarely used

Proposed NAD Schema (cont.)

Field Name	Field Alias	Type	Length	Domain	Description	Expected Use
Bulk_Plus4	Bulk Delivery ZIP Plus 4 Addition	text	7	none	For Bulk Delivery (e.g., government mailroom) zip codes with no corresponding geographical area, The ZIP plus 4 code (without the dash) Example: 1234	rarely used
StN_PreMod	Street Name Pre Modifier (PRM)	Text	15	None	Word or phrase that precedes and modifies the Street Name element or is placed outside the Street Name element so that the Street Name element can be used in creating a sorted list of complete street names	commonly used
StN_PreDir	Street Name Pre Directional (PRD)	Text	50	Domain	Word preceding the Street Name element that indicates the direction taken by the street from an arbitrary starting point or line, or the sector where it is located	commonly used
StN_PreTyp	Street Name Pre Type (STP)	Text	25	domain	Word or phrase that precedes the Street Name element and identifies a type of thoroughfare in a complete street name	commonly used
StN_PreSep	Street Name Pre Type Separator (STPS)	Text	20	domain	Preposition or prepositional phrase between the Street Name Pre Type and the Street Name	commonly used
StreetName	Street Name (RD)	text	60	None	The element of the complete street name that identifies the particular street (as opposed to any street types, directionals, and modifiers).	always used
StN_PosTyp	Street Name Post Type (STS)	Text	15	domain	Word or phrase that follows the Street Name element and identifies a type of thoroughfare in a complete street name	commonly used
StN_PosDir	Street Name Post Directional (POD)	Text	50	domain	A word following the Street Name element that indicates the direction taken by the street from an arbitrary starting point or line, or the sector where it is located.	commonly used
StN_PosMod	Street Name Post Modifier (POM)	Text	25	None	A word or phrase that follows and modifies the Street Name element, but is separated from it by a Street Name Post Type or a Street Name Post Directional or both	commonly used
AddNum_Pre	Address number prefix (HNP)	Text	15	None	An extension of the Address Number that precedes it and further identifies a location along a thoroughfare or within a defined area	commonly used
Add_Number	Address number (HNO)	Long	6	None	The whole number identifier of a location along a thoroughfare or within a defined community.	always used

Proposed NAD Schema (cont.)

Field Name	Field Alias	Type	Length	Domain	Description	Expected Use
AddNum_Suf	Address number suffix (HNS)	Text	15	None	An extension of the Address Number that follows it and further identifies a location along a thoroughfare or within a defined area.	commonly used
LandmkPart	Landmark Name Part (LMKP)	Text	150	None	The name or collection of names by which a prominent feature is publicly known.	occasionally used
LandmkName	Landmark (LMK)	Text	150	None	The name by which a prominent feature is publicly known.	occasionally used
Building	Building (BLD)	Text	75	None	One among a group of buildings that have the same address number and complete street name.	commonly used
Floor	Floor (FLR)	Text	75	None	A floor, story, or level within a building.	commonly used
Unit	Unit (UNIT)	Text	75	None	A group or suite of rooms within a building that are under common ownership or tenancy, typically having a common primary entrance.	commonly used
Room	Room (ROOM)	Text	75	None	A single room within a building.	rarely used
Addtl_Loc	Additional Location Information (LOC)	Text	225	None	A part of a subaddress that is not a building, floor, unit, or room.	rarely used
Milepost	Milepost	text	50	none	A posted numeric measurement from a given beginning point, they may or may not be actual milepost. Milepost numbers are useful for specifying locations along interstate highways, recreational trails, and other unaddressed routes, as well as stretches of county, state, federal, and other routes where distance measurements is posted. Milepost numbers may be given in place of or in addition to Address Numbers Example: Mile Marker 231.5	rarely used
Longitude	Address Longitude	Float	12	None	Address Longitude, derived based on point placement	always used
Latitude	Address Latitude	Float	11	None	Address Latitude, derived based on point placement	always used
NatGrid_Coord	National Grid Coordinates	text	50	None	National Grid Coordinate, derived based on point placement	always used
GUID	GUID	GUID		None	Globally Unique Identifier (also known as Universally Unique Identifier - *UUID*), automatically generated	always used

Proposed NAD Schema (cont.)

Field Name	Field Alias	Type	Length	Domain	Description	Expected Use
Addr_Type	Address Type	Text	50	domain	Indicates the general use of the address (residential, commercial, etc) The type of feature identified by the address	commonly used
Placement	Address Placement	Text	25	domain	Method used to place the address point	commonly used
Source	Address Source	text	75	None	Entity that provided the data to the NAD (this could be different than the authority. For example, the state may aggregate from the counties and then submit to the NAD. In this case, the state would be the source and the authority would be the county.)	always used
AddAuth	Address Authority	text	75	None	Entity responsible for the address assignment and maintenance	commonly used
UniqWithin	Unique Within	text	75	None	The name of the area within which the address should be assumed to be unique.	occasionally used
LastUpdate	Date Last Updated	date	26	None	Date that the address was last updated in the database	always used
Effective	Effective Date	date	26	None	Date that the address becomes effective (may be past or future)	commonly used
Expired	Expiration Date	date	26	None	Date that the address expires	commonly used

Appendix 6

Proposed NAD Schema Domains

Street Name Pre Directional
North
Northeast
East
Southeast
South
Southwest
West
Northwest

Street Name Pre Type
As referenced in the CLDXF standard, these values are found in the USPS Publication 28 Appendix C1 (http://pe.usps.gov/text/pub28/28apc_002.htm)

Street Name Pre Type Separator
of the
at
de las
des
in the
to the

Street Name Post Type
As referenced in the CLDXF standard, these values are found in the USPS Publication 28 Appendix C1 (http://pe.usps.gov/text/pub28/28apc_002.htm)

Street Name Post Directional
North
Northeast
Northwest
South
Southeast
Southwest
East
West

State
http://pe.usps.gov/cpim/ftp/pubs/Pub28/Pub28.pdf

County
www.census.gov/geo/reference/codes/cou.html

Address type
Residential (Housing)
Commercial (Office, Retail, Restaurant, Banking)
Multi-Use (mixed commercial/residential)
Open Space (Forest, vacant, cemeteries)
Industrial
Gov't/Public Services (Fire/Police, Library, Gov't offices)
Religious
Recreation (Ball fields, parks, golf courses, ski area)
Educational (Schools, Universities)
Institutional (Hospitals, group homes, Prisons, etc)
Other
Unknown

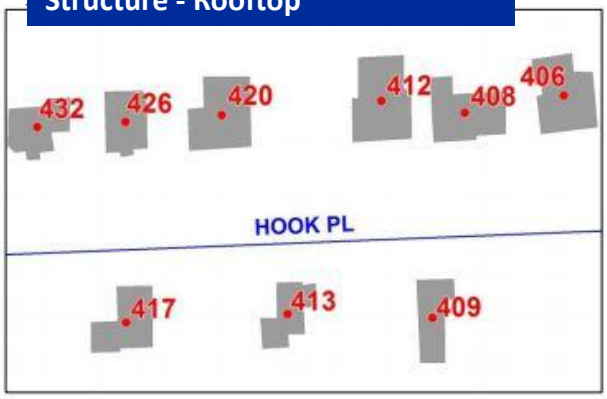
Address Placement
Structure - Rooftop
Structure - Entrance
Structure - Interior Unit Location
Parcel - Centroid
Parcel - Other/Manual Placement
Linear Geocode
Property Access Point
Site Placement
Other (some other method not listed)
Unknown (unknown address placement method)

(see next page for examples of address placement)

Address Placement Examples

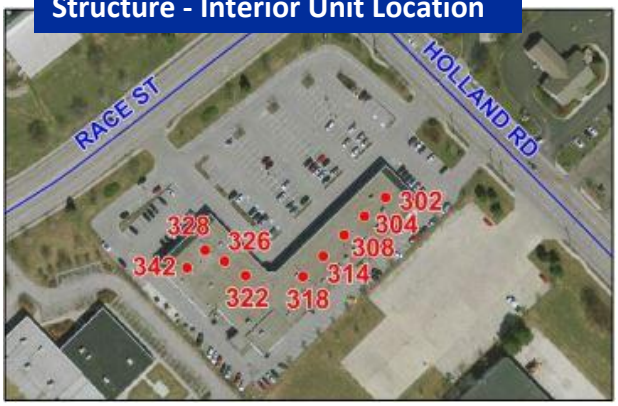
Images are from the *NENA Information Document for Development of Site/Structure Address Point GIS Data for 9-1-1*

Structure - Rooftop

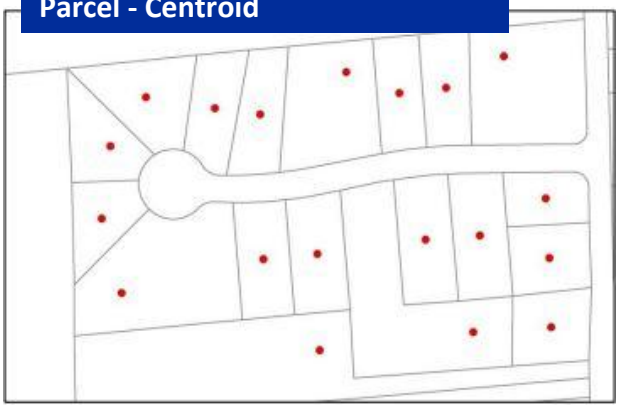


Structure - Entrance

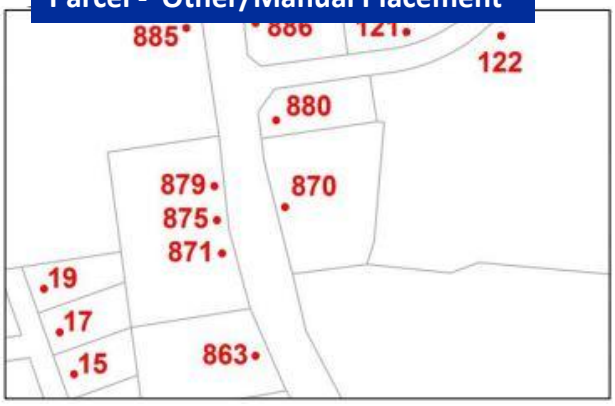
Structure - Interior Unit Location



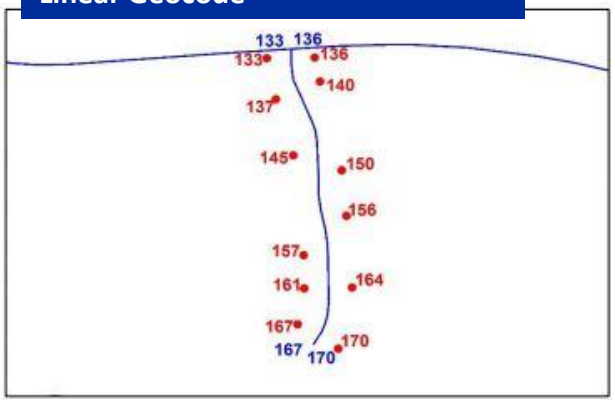
Parcel - Centroid



Parcel - Other/Manual Placement



Linear Geocode



Property Access

Site Placement

Appendix 7

Data Preparation and Standardization for Loading into the NAD

Based on the work done to prepare and standardize the three pilot datasets for loading into the NAD (AR, AZ, Boone County), the following are some **examples of typical data preparation and standardization activities** that address entities will likely need to undertake in order to load their data into the NAD Schema. The extent to which an entity's data schema matches the NAD will determine the amount of data work required.

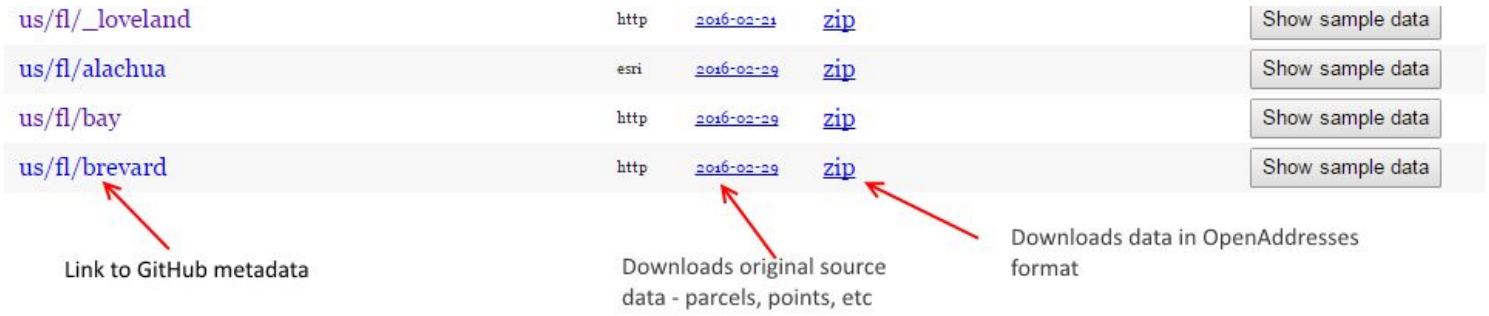
- Field Mapping to NAD Schema
- Set default values for fields such as State and Source
- Fully spell out abbreviations for pre/post direction fields
- Fully spell out Post Name Types (Road, Street, etc.)
- Assign Address Authority where available
- Convert source Address Placement and Address Type values (where available) to match NAD Domains
- Extract Lat/Long coordinate values from point location
- Convert Lat/Long values to National Grid coordinates
- Reformat date fields
- Generate GUID
- Trim spaces from values
- Change all characters to uppercase for consistency throughout the dataset
- Reproject data to EPSG:3857

Appendix 8

OpenAddresses.io Data Source Information

OpenAddresses provides a global collection of open and free addresses data sources. Over 200 million addresses are available worldwide. OpenAddresses pulls from a wide variety of sources, and this presents some issues that users must work around when using the data.

The website is not entirely intuitive. Data Sources and vintage is not readily available on the site, but most of this information can be found on GitHub.



The link to us/fl/brevard opens to the GitHub metadata page for the data. Attribution for source entity and data source location (URL) are highlighted below


```
"coverage": {
  "US Census": {
    "geoid": "12009",
    "name": "Brevard County",
    "state": "Florida"
  },
  "country": "us",
  "state": "fl",
  "county": "Brevard"
},
"attribution": "Brevard County",
"data": "https://www.bcpao.us/gisdata/county_wide/address.zip",
"type": "http",
"compression": "zip",
"website": "https://www.bcpao.us/gisdata/datasearch.asp",
"conform": {
```

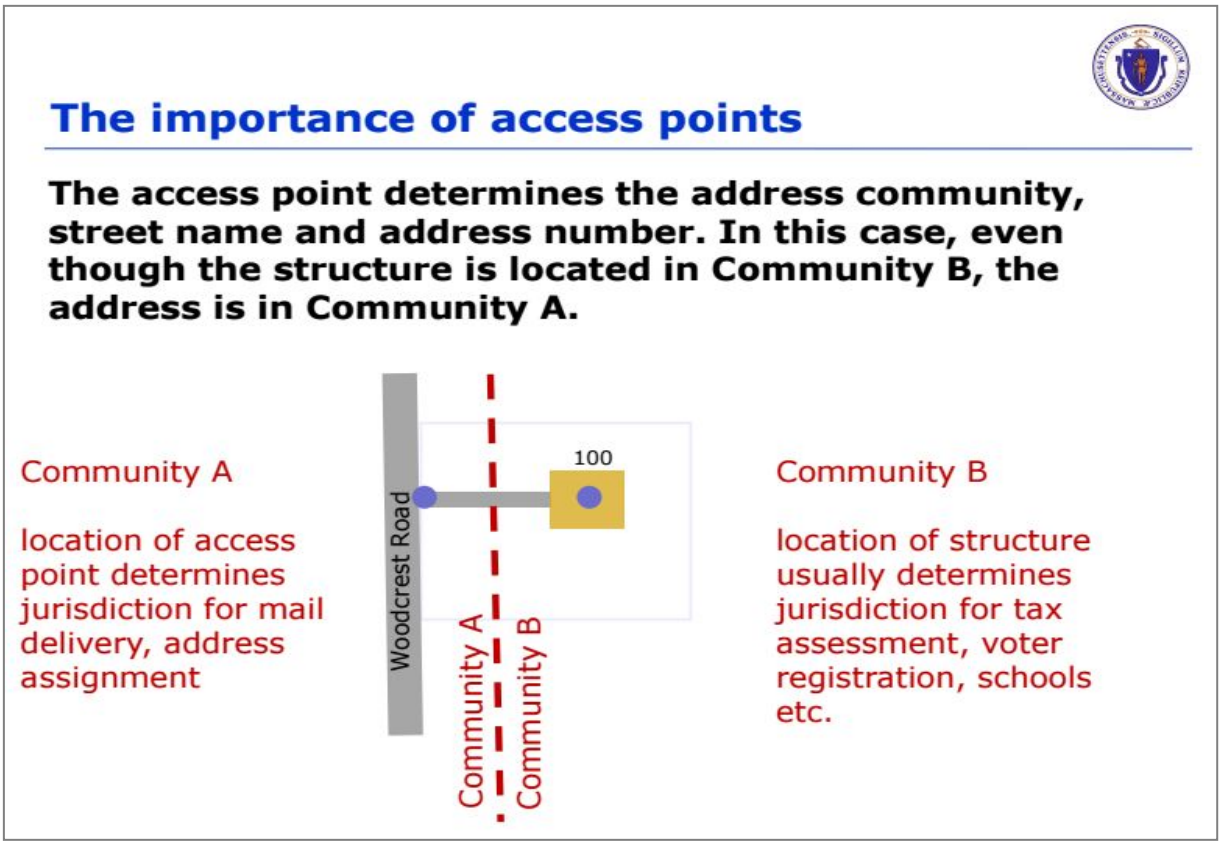

……..

Appendix 9

Advanced Address Database Considerations

In discussions with address experts and pilot project participant entities, several advanced address database characteristics were brought up. These might be included in sophisticated address databases that could go beyond the minimum. At this time, these elements are not needed in the NAD, but they will greatly increase the efficacy of address data at the local level. Therefor they should be considered by address authority entities when developing and expanding their local address data.

- Maintain *Address Authority* and Address "*Unique Within*" polygons
  - Address Authority polygons represent the area that a given address authority is reponsible for creating and maintaining address data within. These polygons can be a quality control tool to ensure all areas are covered by some entity, and no area is covered by more than one entity.
  - Polygons representing an area within which all addresses are assumed to be unique would be an additional useful dataset for validation and ensuring address uniqueness.
- Store Landmark- based addresses (e.g., highway rest areas, parking lots, trailheads)
- Store street name alias information via a lookup table (i.e. alternate street names, former street names)
- Store Legacy Addresses (i.e., addresses that no longer exist due to demolished buildings, re-addressing, road name changes, etc.)
- Maintain multiple points for a given address. In particular, capture the access point in addition to the building location
  - Access points from the road are critical for emergency responders
  - Ensure that the "place" (City, town, community) represents the access point for that address, which may be different than the address point itself. See image below.



The Importance of Access Points (image credit: Christian Jacqz, MassGIS)