National Address Database Pilot Project

NACo GIS Subcommittee

February 2016

Agenda

Two simple things

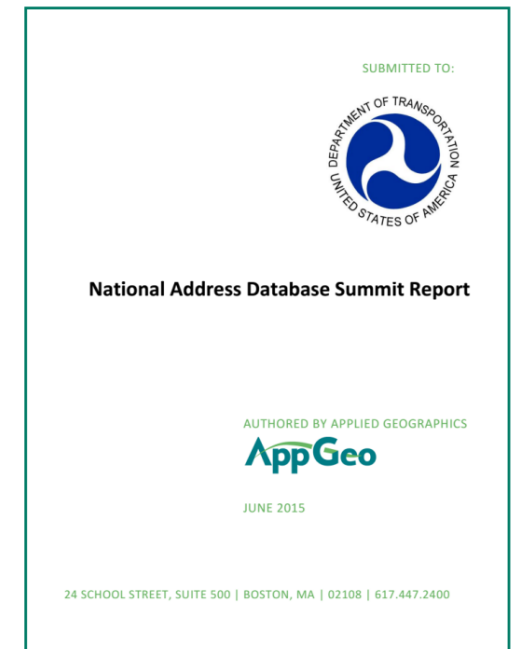1. Share status and progress

2. Answer questions

Background
National Address Database Summit

- Held in April 2015

- Key outcomes:
  - Enough talking; let's start *doing*
  - Continue to build momentum

- Recommended next step:
  - ***Pilot Projects***

- USDOT was able to fund the NAD pilot

Figure 1. Participants at the Summit shown in the large group discussion.

SUBMITTED TO:

National Address Database Summit Report

AUTHORED BY APPLIED GEOGRAPHICS

AppGeo

JUNE 2015

24 SCHOOL STREET, SUITE 500 | BOSTON, MA | 02108 | 617.447.2400

Purpose of the NAD Pilot Project

Four major work goals:

1. Determine **minimum data content** standards

2. Explore **workflows for address creation**

3. Understand best practices for **address roll-up**

4. Assess the **technical feasibility** of the NAD

5. Philosophical goal: keep NAD in public domain!

NSGIC Advisory Group Collaboration

- Group made up of key stakeholders/experts, including

  - Andy Rowan
  - Bert Granberg
  - Curtis Pulford
  - Cy Smith

  - Gene Trobia
  - Shelby Johnson
  - Tim Trainor

- Met a handful of times between Oct – Dec 2015

- Discussed project status, provided input and guidance on minimum content standard, etc.

Two distinct components, for two audiences

1. **The "haves"**

   – Agencies who have existing address data

   – Key questions are *how to standardize and roll-up*
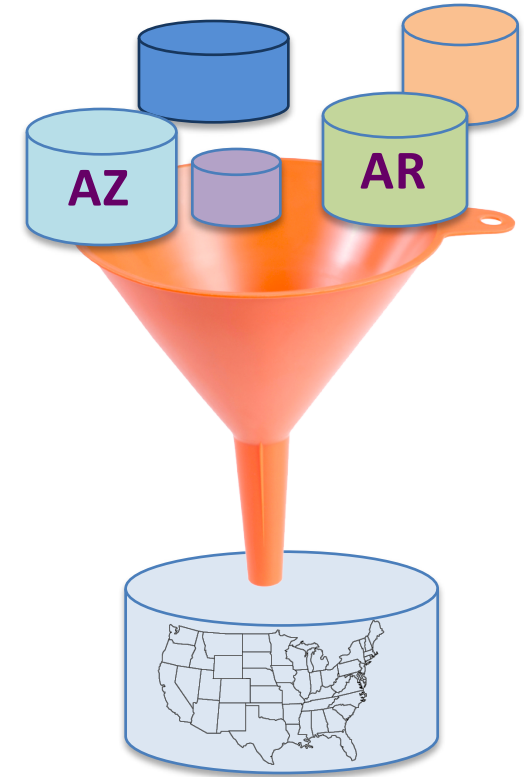
2. **The "have nots"**

   – Agencies interested in addressing but have not started

   – Key question is *how to get started*

   – Streamlined way of developing "initial addresses"

From presentation at NSGIC Annual conference

Current status & what comes next

- ✓ Connecting with participants in AZ and AR
    - – Including here at NSGIC

- ✓ Trying to find interested communities
    - – Both "haves" and "have nots"

- ✓ Reaching out to Tribal participants via NTGISC

- ✓ Commencing detailed research on existing systems/programs
    - – OpenAddresses.io
    - – Community TIGER

- ✓ Identifying best geocoding and address list data sources

- ✓ Beginning work on "minimum data content standard"

Connecting with participants in AZ and AR

Agency Name	Type	Data Shared?	Status Comments
State of Arizona	State	Yes	Have extensive documentation on their internal processes and have regular contact with them. Data have been made available *for NAD pilot project use only.* Working internally on making addresses available in public domain.
State of Arkansas	State	Yes	Have obtained their data and have gathered initial information about their data collection processes.

Trying to find additional, interested have/have-not communities

Agency Name	Type	Data Shared?	Status Comments
Boone County, MO	County	Yes	We have their data and have gathered initial information about their data collection processes.

- **Tallahatchie & Union Counties, Mississippi**
  - Are sharing their "address data creation story" and lessons learned

- **City of Meridian, Idaho**
  - Have had discussions and gathered valuable feedback on their addressing best practices, lessons learned, etc.

"Have-not" status

- Goal to find agencies (likely counties or tribes) that haven't yet created their addresses, but are **interested, motivated**, and **willing to work with us**. We don't want to create addresses that will then sit on a shelf.

✓ **Jackson County, AR**

- – Working on geocoding E911 address **list** to their centerlines/parcels
  - • Initial geocoding done with ArcGIS geocoding tools

- – Match rates are low, working on data scrubbing to improve match rates

- – Looking at other geocoders (e.g., MapQuest) in attempt to improve and compare results

- – Lessons learned for data creation will be included in pilot findings

- **Marion County, AL** as potential second have-not county

Tribal participation

- Conducted joint webinar with National Tribal Geographic Information Support Center (NTGISC) on 12/1/15 with a handful of participants

- Gene Trobia attendance at the National Tribal GIS Conference, held November 16th - 20th 2015 in Albuquerque, NM

Agency Name	Type	Data Shared?	Status Comments
Gila River Indian Community	Tribe	Pending	We have received agreement that they will participate in the pilot, as well as some initial information about their addressing challenges. They are working on getting permission to share their data.
Navajo Addressing Authority	Tribe	Pending	We have been in touch with them and are working on getting their participation. Dialog is open, however, no data or information has been exchanged yet.

Detailed research on existing systems/programs

Do not reinvent the wheel / learn existing lessons

- **OpenAddresses.io**
  - Open dialog with Ian Dees
  - Reviewed documentation
  - Conferred with CFPB on their observations

- **Community TIGER**
  - Received deep briefing/demo from Census
    - Final report will contain an overview of the program
  - Have remained in close contact with Census team

D R A F T

Initial findings/observations on OpenAddresses

- Important data source with 200M+ addresses in DB

- **Global** in scope; not just USA
  - Their schema needs to reflect that

- Solid meta data on data sources and dates
  - But, date is not part of the address record itself
  - Data can be downloaded in original format and/or OpenAddress format

- May house **overlapping data**
  - E.g., data collected from NY State and New York City

- Data may be harvested from a **non-authoritative source**
  - E.g., in MA, the OpenAddress data are derived from statewide **parcels**; not from the MA Master Address Database
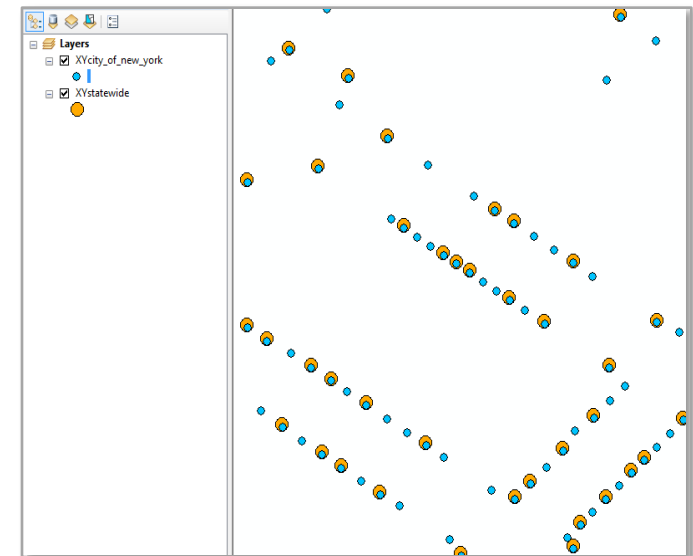
Work completed to get the "minimum data content standard" out for review

- First round of comments from NSGIC and Census have been incorporated

- Broader distribution to all Summit attendees was first week of Feb

  – Have received many comments

  – Most are supportive

D R A F T

Minimum content standard

In general, the NAD will contain three main components:

The Address itself

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

Geographic Location of the address

- Lat/Long
- National Grid Coordinates

Metadata about the address

- Address authority
- Address source
- Address date
- Unique ID
- Type (residential, commercial, etc.)
- Placement (rooftop, driveway access, etc.)

How did we arrive here?
Info gathering on existing address database

1. **Philosophy:** *Keep it simple!*

   – Broadest participation possible

   – Lowest barrier to entry

2. **Evaluation of existing address schemas and initiatives**

   – Understand lessons already learned

   – Schema comparison

     • FGDC and CLDXF

     • Review of state schemas: AR, AZ, MA, NC, NY, RI, UT, VA, VT

     • Other schemas (e.g., DC, various counties)

   – Initiatives

     • OpenAddresses

     • CommunityTIGER

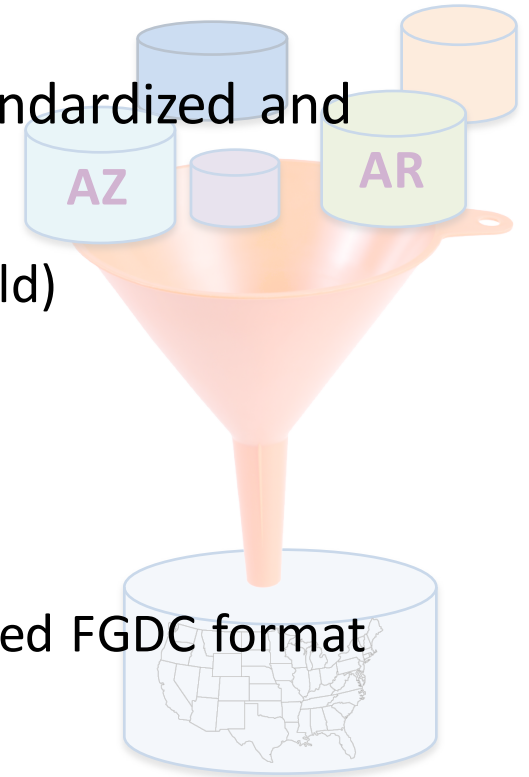# Key issues and planning factors
Address parsing, or not

- ## Address **parsing** is a key best practice
  - Breaking the address into its "parts": Street number + Street name + Type
  - FGDC standard identifies a comprehensive parsing approach

- ## Not everyone uses FGDC parsing
  - Not all FGDC components are found everywhere
  - Thus, local parsing schema may be more appropriate
  - The NAD must be able to accommodate a variety of parsing approaches

FGDC	Arizona	Arkansas	North Carolina	Utah	Virgina
Address Number Prefix	AddNum_Pre				PREADDRNUM
Address Number	Add_Number	adr_num	ADDR_HN	AddNum	ADDRNUM
Address Number Suffix	AddNum_Suf	adr_num_suf		AddNumSuffix	ADDRNUMSUF
Street Name Pre Modifier	StN_PreMod				STREET_PREMOD
Street Name Pre Directional	StN_PreDir	pre_dir	ADDR_PD	PrefixDir	STREET_PREFIX
Street Name Pre Type	StN_PreTyp				STREET_PRE_TYPE
Seperator Element	StN_PreSep				
Street Name	StreetName	pstr_name	ADDR_SN	StreetName	STREET_NAME
Street Name Post Type	StN_PosTyp	pstr_type	ADDR_ST	StreetType	STREET_TYPE
Street Name Post Directional	StN_PosDir	psuf_dir	ADDR_SD	SuffixDir	STREET_SUFFIX
Street Name Post Modifier	StN_PosMod	pstr_mod			

Key issues and planning factors
Data aggregation and parsing

- Parsing tools exist/can be developed to create standardized and fully-parsed aggregations

  - To parse, un-parsed data (i.e., full address is one field)

  - To map one parsing scheme into the FGDC scheme

- Three options

  1. State/authoritative data is submitted in fully parsed FGDC format

     a. State **maintains** in FGDC format (e.g., AZ)

     b. State **converts** to FGDC before submitting to NAD

  2. Un-parsed/non-FGDC parsed data are submitted and **NAD tools** parse the data before aggregation
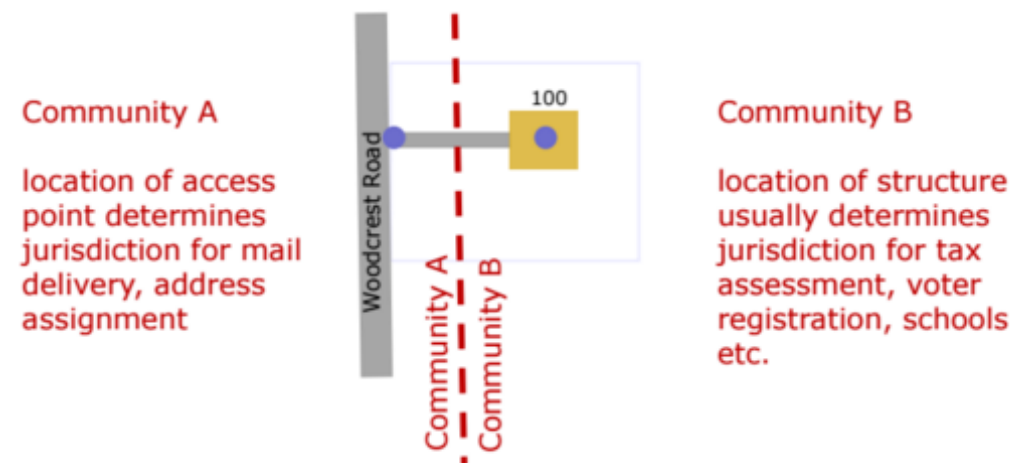
Important, unresolved issues

- Places where **CDLXF** and **FGDC** have significant differences
  - Neither was designed as a "schema", both are "transfer standards"

- **Subaddress**
  - **FGDC**: type/value pairs (e.g. type="Wing", value ="Eastman Cancer Wing")
    - Very flexible and extensible
  - **CLDXF**: Building / Floor / Unit hierarchy

- **Place**
  - **FGDC**: place name type / value pair (e.g., type="Municipality", value="New York City")
    - Pairs can be repeated to denote the hierarchy of "place" (e.g., County, Muni, etc.)
  - **CLDXF**: Discrete fields for County, Municipality, etc.
  - Ability to geographically calculate place via overlay

Key issues and planning factors
Important, unresolved issues

- Polygons for **"address authority" boundary**
  - Issue identified by multiple states with mature address programs
  - Definitive extent of authority
  - Helps explain edge cases (e.g., address in one town; street access is in another)
  - **Key for QA/QC**:

    What happens if someone submits an address that's in another's jurisdiction?

Community A

location of access point determines jurisdiction for mail delivery, address assignment

Woodcrest Road

Community A
Community B

100

Community B

location of structure usually determines jurisdiction for tax assessment, voter registration, schools etc.

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

What comes next

- **Finalize** the minimum content standard

- Data work with "have", and "have not" partners
  - **Haves**: explore collection, standardization and aggregation into minimum content standard
  - **Have nots**: explore "shortcuts" and first steps for initiating address data creation
  - **Prototype NAD**: Incorporate have/have-not results into prototype NAD

- Begin **final report** writing
  - Summarize research and findings
  - Present lessons learned
  - Outline next steps for activating the NAD

Questions?

Steve Lewis

USDOT Geospatial Information Officer

steve.lewis@dot.gov