NATIONAL ADDRESS DATABASE (NAD)
MINIMUM CONTENT STANDARD

The NAD Pilot project is currently underway and is aimed at investigating the requirements, workflows, challenges and best practices related to building the NAD. The first critical task in the pilot project is to determine the NAD Minimum Content Standard. This document outlines the content standard and approach in order to gather input and feedback from stakeholders. Once this first critical step is completed, the other key pilot objectives (workflows for address creation, best practices for address roll-up, etc.) will be explored further.

SUMMARY

The goal of the minimum content standard 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 rollup. In general, the NAD will contain three main components:

- The address itself
  - Address Number, Street Name, Subaddress
  - City/Town/Place, County, State, Zip
- The geographic location of the address
  - Lat/Long, National Grid Coordinates[1]
- Further identification and character attributes about the address location
  - Unique ID (e.g., GUID)
  - Address type (based on a domain, including residential, commercial, etc)
  - Address placement (based on domain, including rooftop, driveway entrance, structure entrance, etc)
  - Address authority (i.e., data creator), address source (i.e., data aggregator), address date

These data elements represent the ideal data content for the NAD, which will be held as the goal to work towards. However, 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 likely 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 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 it as it may inform future additions to the NAD.

NAD ROLLUP APPROACH

- 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).
- **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*

[1] The NAD guidance will specify a target accuracy for both Long/Lat and 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.

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. The exact schema to be used for the NAD is still being discussed (see below).

- **For collection, the NAD will keep a low barrier to participation** (e.g., don't require a specific parsing schema for data contributors)
- 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 and submit, 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 has 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 will develop and use parsing tools for non-parsed submissions. For the national implementation, it would be the responsibility of the NAD owner/manager. Resources will be necessary to build and maintain the right tools (e.g., for parsing, ETL, etc).
- 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.

DATABASE SCHEMA CONSIDERATIONS– FGDC AND CLDXF

The project team is still considering which schema may better fit the needs of the NAD – FGDC or CLDXF, or perhaps some combination thereof. The project team, project advisors and the NAD Pilot Project Advisory Group (which is made up of several NSGIC members and other stakeholders) are weighing in on which data schema might be best for the NAD. In some ways, the two are very similar. For example, the address number and street name address parsing both use roughly the following fields:

- Address Number Parsing
    - Address Number Prefix
    - Address Number
    - Address Number Suffix
- Street Name Parsing
    - Street Name Pre Modifier
    - Street Name Pre Directional
    - Street Name Pre Type
    - Street Name Pre Type Separator
    - Street Name
    - Street Name Post Type
    - Street Name Post Directional
    - Street Name Post Modifier

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 (which will be parsed in the NAD)
  - 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)”.

Based on recent meetings with the NAD Pilot Project Advisory Group, the preference of the group is leaning towards the intuitive nature of the CLDXF model. However, further evaluation and input on this matter is needed.