

Meeting Notes  
FGDC Address Subcommittee  
August 12, 2020  
Webinar

Attendance (29 Total):

Carl Anderson, URISA  
Andrew Bailey, Department of the Interior  
Florinda Balfour, Department of Veterans Affairs  
Dierdre Bevington-Attardi, U.S. Census Bureau  
Keri Brennan, Michael Baker International  
Dave Cackowski, Census Bureau  
Jonathan Duran, State of Arkansas  
Monique Eleby, Census Bureau  
Ross Epstein, SafeGraph  
Ashley Hitt, Connected Nation  
Laura Hogberg, Census Bureau  
Phil Markert, Department of Homeland Security  
Jeremy McMullen, State of Vermont  
James Meyer, State of Arizona  
Alesha Perdomo, U.S. Postal Service  
Chris Portell, FEMA  
Raúl Ríos-Díaz, iCasaPR  
Karla Riso, Census Bureau  
Richard Robinson, Housing and Urban Development  
Karen Rogers, State of Wyoming  
Dan Ross, State of Minnesota  
Jill Saligoe-Simmel, Esri  
Diane Snediker, Census Bureau  
John Sperling, iCasaPR  
Thomas Springsteen, HIFLD/Booz Allen Hamilton  
Marilia Valdes, FEMA  
Ed Wells, URISA  
Martha Wells, URISA  
Matt Zimolzak, Census Bureau

## Meeting Summary

### **Address Workflow Subgroup Update, Matt Zimolzak (Census Bureau):**

- The group will present the second tab of the Address Workflow Validations Matrix to the Address Subcommittee later in this meeting.

### **Puerto Rico Civic Address Vulnerability Evaluation (PRCAVE) Update, Raúl Ríos-Díaz (iCasaPR):**

- Given everything that is going on, including Hurricane Isias, Covid, and earthquakes, which have led to Census response and primary election issues, there is great need to improve addresses. Approximately 50 percent of structures in Puerto Rico may lack addresses.
- PRCAVE is planning a data release in September. There will be more information on this at the September Address Subcommittee meeting.
- PRCAVE is also starting to parse addresses according to the FGDC standard and USPS Publication 28.

### **Puerto Rico Address Data Working Group (PRADWG) Phase 3 Update, Dave Cackowski (Census Bureau)**

- PRADWG Phase 3 will result in a report on federal agency Puerto Rico address data needs and goals.
- The first step will be to send agencies a questionnaire to ask about their needs and goals for Puerto Rico address data.

### **Address Workflow Validations Matrix, FGDC Validations Tab, Martha Wells (URISA)**

- Data Quality Process
  - Data originates with the Local Address Authority (LAA), which is the only group that has the local knowledge.
  - State tests, reviews, and aggregates the data and may accept or return to LAA to fix.
  - NAD receives and reviews test results, conducts limited further testing and may accept or return to State.
- Identifying Elements and Aspect to be tested.
  - A Matrix was created to identify each element or attribute to be tested, and using the Validation ID (Val ID) to describe the needed test and link the test to the appropriate Data Quality test or tests.
  - In addition, the Workflow Group identified the level (Local, State or NAD/Federal) that would test each item, and whether tests are required or recommended.
- Address Workflow Validations Matrix – Main Tab – we reviewed for the subcommittee about a year ago. Only the state workflow is complete so the other boxes are not filled in on the matrix.

- Local Address Authority (City/County/Region)
  - Address Creator and maintenance authority
  - Local knowledge
  - Address Reference System (business rules) and street centerline network
  - Can document known anomalies
  - Some aggregation from towns/cities to County or regional agency may occur.
  - Tests:
    - Completeness
    - Temporal accuracy
    - Positional accuracy
    - Logical consistency
    - Attribute (Thematic) accuracy
    - Lineage
- State Geographic Information Office
  - 2<sup>nd</sup> level aggregation
  - Standardization of data into NAD Schema
  - Tests:
    - Logical Consistency
    - Attribute (Thematic) Accuracy
    - Positional accuracy (within submitting Address Authority jurisdiction)
    - Temporal accuracy (based on date stamps)
    - Completeness (are critical fields filled?)
  - Data failing tests returns to Local Address Authority
  - State level requirements for LAA submissions may vary.
- State Level
  - Some states require centerline with spatial and tabular address records.
    - Allows testing of parity, sequence, and ranges
  - Other states are limited to spatial and tabular data provided for individual addresses.
    - Allows testing of administrative boundaries (county, city/town, region, postal ZIP Code)
  - Some states do not have a current program to aggregate Local Address Authority data
    - Some testing may be done at LAA level, some may be done at NAD
- NAD
  - Data received from State GIO
  - Tests for:
    - Logical consistency
    - Attribute (Thematic) accuracy
    - Positional accuracy (within State)
    - Completeness (all not-null fields are populated)
  - Data that fails test(s) is returned to State GIO, or Local Address Agency for review and potential resubmittal.

- The second tab is the FGDC Validations
- Data Quality Tests
  - Based on tests defined in FGDC United States Thoroughfare, Landmark and Postal Address Data Standard
  - Tests are compliant with **standard ISO metadata** definitions and quality parameters
  - Different tests are conducted at different levels of aggregation based on the ability of testing agency to have sufficient information to conduct the test.
- Address data Quality Tests
  - Tests are identified In the Data Quality Part of the United States Thoroughfare, Landmark and Postal Address Data Standard
  - Tests that apply to the NAD Content and Data Schema will be finalized after the Content Working Group's recommendations are reviewed and the final Schema is established.
- Organization
  - Tests are listed in 4 areas as identified in the State and Federal Workflows
  - Seventeen different categories of data were identified to be validated
- Testing Protocols
  - Each of the validations to be performed is listed, and referenced to one or more tests that are found in the FGDC Standard.
    - Some components of an address have only one relevant test, others have as many as 5 or more appropriate tests.
    - The validation and sub-validation columns indicate the parameter being tested, and the expected results of the test that validate the individual data item in the record.
    - The tests are listed by name, and the section number of the FGDC Standard
- Test References: Columns
  - Tests are listed in alphabetical order, not order of execution.
  - "Domain" tests examine values found in the record against the allowable values in the domain of values for that field.
  - Other tests examine business rules that are included in the data and examine records to be certain that the data values match the rules.
- Spatial Domain Tests
  - Identify address values that do not match their location information
  - Values are tested as point-in-polygon: Does the address record's point, expressed as a pair of x/y coordinates, fall within the polygon that the record contains for that spatial area.
- Tabular Checks
  - Cover domains of values and definitions of individual fields in data.
  - Conducted at all three levels (Local Address Authority, State, NAD)
  - Conformance with local business rules occurs at Local Address Authority (LAA)-logical consistency and thematic accuracy

- Conformance with defined values (street name directionals, types, place names, state names, etc.) can be done at LAA, State or NAD—thematic accuracy
- Missing values (LAA, State, NAD)—completeness check
- Misplaced values (LAA, State, NAD) (wrong type of data in the field— such as a town name in the street name field.)
- Data Tests
  - These include specific tests of individual data records to determine their validity.
  - Relationships can also be tested: If the address lifecycle status is “active”, is the “start date” for the address earlier than today, and is the end date null?
- Goals of the Validation Process
  - Ensure data quality and trustworthiness
  - Provide feedback to Local Address Authorities on anomalies or other data irregularities found
  - Assist in standardizing address information across multiple jurisdictions and states
- Discussion
  - Raul Rios – Do you have any guidelines for checking data quality? Martha – All tests are fully described in the FGDC Standard. There are about 38 tests. Ed – There is a distinction between logical tests and empirical tests. Martha has been describing logical tests. An example of an empirical test: ‘is the address on the correct side of the street?’
  - Karen Rogers – NSGIC has a geospatial maturity assessment story map. This is a good place to see what the states are doing.

## Content Recommendations for the National Address Database (NAD) (Continued), Ed Wells (URISA)

- Subaddresses
  - Simple Elements
    - Subaddress Type (e.g., *Room* in “Room 536”)
    - Subaddress ID (required) (e.g., 536 in “Room 536”)
  - Complex Elements
    - Subaddress Element (e.g., “Fifth floor”; or “Room 536”) = { Subaddress Type + Subaddress ID\* }
  - Complete Subaddress (“Fifth floor, Room 536”) = A series of one or more Subaddress Elements
  - Optional in:
    - Numbered Thoroughfare, Unnumbered Thoroughfare, Landmark, Community addresses.
  - Prohibited in:
    - Intersection, Two-number Address Range addresses
- Landmark Names
  - Simple Elements
    - Landmark Name

- **Complex Element** [NAD recommendation deferred]
- *Complete Landmark Name* = A series of one or more Landmark Names
- **Mandatory in:**
  - Landmark and Community addresses
- **Optional in:**
  - Numbered Thoroughfare, Intersection, Two-number Address Range, and Unnumbered Thoroughfare addresses
- **Address Attributes**
  - Address ID, Source, and Reference Area
  - Address Coordinate Position(s)
  - Address Relationships
  - Additional Address Attributes (Mandatory)
  - Additional Address Attributes (Optional)
  - Attributes of Address Elements
- **Address ID, Source, and Reference Area**
  - **\*Address UUID** – Unique, persistent Address ID
  - **\*Address Authority** – Agency with authority to create/ alter/retire the address
  - **Address Reference System Name** – Name of the ARS within whose extent the address is located (and can be presumed unique)
  - *\* = Mandatory attribute*
- **Address Coordinate Position(s)**
  - **\*Address Longitude**
  - **\*Address Address Latitude**
  - **\*US National Grid Coordinate**
    - Address Elevation
  - **\*Address Placement** – Method used to place the address point within the addressed location (rooftop, main entrance, etc.)
  - **\*Address Point** – Mappable point or Point Z, constructed as a WKT Point or Point Z from the XY or XYZ coordinates, plus the Address Placement attribute
- **Address Relationships**
  - *Address-to-address Relationships*
    - **Related Address ID**
    - **Address Relationship Type** (official to alias; predecessor to successor, etc.)
  - *Address-to-parcel Relationships*
    - **Address Parcel Identifier Source**
    - **Address Parcel Identifier**
- **Additional Address Attributes (Mandatory)**
  - **\*Address Classification** – The class of address, as described in previous slides
  - **\*Address Last Revision Date** – The date the address record was last updated within the NAD

- **\*NAD Data Provider** – Organization or person providing data directly to the NAD
- **\*Data Set ID** – An identifier that associates each record of a transmitted dataset with the file-level metadata for the dataset
- **Additional Address Attributes (Optional)**
  - **Address Lifecycle Status** – e.g., potential, proposed, active, retired
  - **Address Start Date** – Earliest date on which the address is known to exist
  - **Address End Date** – The date on which the address is known to be no longer valid
  - **Address Anomaly Status** – Flag or note for a known exception to local addressing rules
  - **Location Description** – Text description of how to identify or find the addressed feature
  - **Address Feature Type** – The type of land use or feature at the address.
- **Attributes of Subaddress Elements (all Optional)**
  - **Subaddress Component Order** – the order in which the Subaddress Type and Identifier appear within a Subaddress Element
  - **CLDXFv1 Subaddress Type** – CLDXFv1 subaddress category (if any) into which a NAD Subaddress Element should be classified
  - **CLDXFv2 Named Location Type** – CLDXFv2 named location category (if any) into which a NAD Subaddress Element should be classified
- **Attributes of Address Elements (all Optional)**
  - **Element Sequence Number** – the sequence in which a series of Subaddress Elements, or Landmark Names, or Place Names should be ordered when they are presented together.
  - **Place Name Type** – The type of Place Name used in an address (County, Municipal, Postal City, etc.)
  - **Delivery Address Type** – Whether the Delivery Address includes or excludes the Complete Subaddress
- **Discussion**
  - (Q) Jon Sperling (iCasaPR) – What about temporal domains like timestamps?
  - (A) Ed – We have a start on covering it. Key information is captured, including last revision date, start date, and end date.
  - (Q) Jon – What if an address to parcel relationship changes?
  - (A) Ed – Need to trust the provider of parcel database to update us if relationships change. Matt – These are optional items. Ultimately it becomes buyer beware

**Content Recommendations for the NAD – Place Geography, Matt Zimolzak (U.S. Census Bureau)**

- **Why Include Geography in NAD?**
  - Technically not required
  - Can derive any geography for address points by point/polygon relationships with geographic area layers.

So why?

- Industry standard / Best practice
  - States and counties/equivalents minimum requirement in a spatial database of national scope.
- Convenience
  - Enables easy filtering/subsetting by included geographies
  - Enables multiple selections in a geography class, without selecting the whole class.
  - Enables cross-spatial geography relationships
- Useful in conveying the location of an address
- Guiding Principles for Inclusion
  - Include commonly used/referenced governmental units.
  - Primarily focused on legally defined and “functioning” governmental units.
  - Include some geographic areas defined for statistical, planning or commercial purposes (not legally defined), when it serves a useful or unique place geography purpose.
  - Must have documented and maintained spatial boundaries.
    - ZIP Code is the exception
- Guiding Principles in Execution
  - Single value for each Place Type (or null)
  - Consistency in variable element population
    - “Massachusetts” or “Commonwealth of Massachusetts,” but not both.
  - Positive indicator of null values (not global)
  - Positive indicator of unknown value (not global)
- States and Equivalents (Name)
  - Mandatory
  - No null or unknown values
  - “State” and “Commonwealth” treated the same/equivalent
  - District of Columbia is a state equivalent
  - US Territories are state equivalents
- Counties and Equivalents (Name)
  - Mandatory
  - No unknown values
  - County equivalents:
    - Parishes (LA), Boroughs (AK), Municipios (PR)
    - Independent Cities
    - Fictitious County Equivalents (DC and Guam)
- Postal City Name
  - Mandatory
- Census Designated Place (CDP) Name
  - Optional
  - Null Values Allowed

- Extensive information on and TIGER/Line Shapefiles with CDPs available from [www.census.gov](http://www.census.gov)
- Unincorporated Community Name
  - Optional
  - Null Values Allowed
- Neighborhood Name
  - Optional
  - Null Values Allowed

**Action Items**

- Continue to review and comment on the NAD Content Recommendations – Address Subcommittee Members.

**Next meeting:** Wednesday, September 16, 2020 at 11am ET.