

Appendix 1: HUD

PUERTO RICO ADDRESS DATA WORKING GROUP - PROVEN PRACTICES TEMPLATE

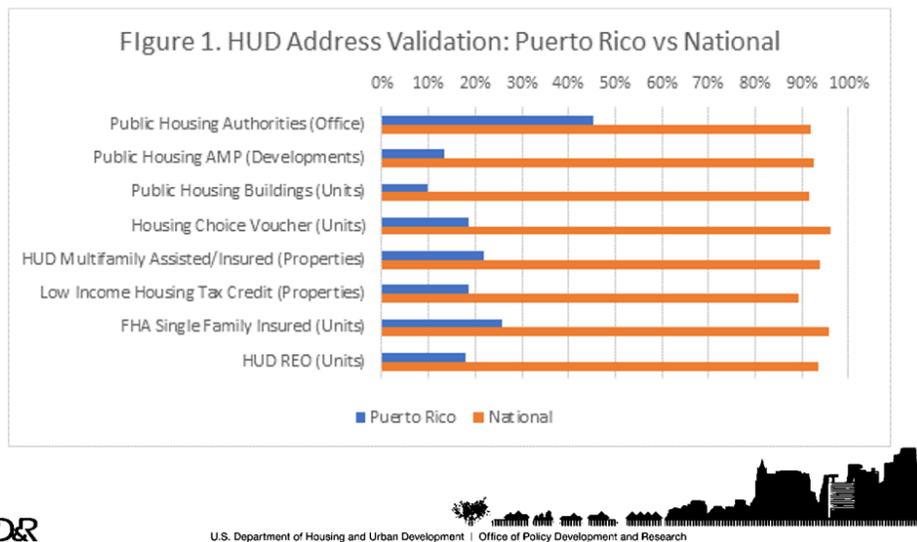
Organization	Housing and Urban Development
Point of Contact:	John Huggins Social Science Analyst, Research Geographer John.C.Huggins@hud.gov (202) 402-4153
Proven Practice in a Tweet	HUD utilizes Puerto Rico address data as given, but on occasion has contracted out for address cleansing.
Submitted On	5/13/2019
Challenge or Opportunity	<ul style="list-style-type: none"> • <i>What problem are you trying to solve?</i> HUD is able to validate only 20% of the assisted, insured or owned properties in Puerto Rico. The goal was to improve the quality of addresses in our administrative data on Puerto Rico and to get recommendations on how to improve the processes and systems HUD uses to collect addresses on the island. • <i>How is this problem related to the mission of your agency or agencies?</i> HUD's Office of Community Planning and Development had an immediate need for improved geocoded data in Puerto Rico to support planning and disaster recovery efforts For HUD, location intelligence does not exist without good address data. <ul style="list-style-type: none"> ○ Successful address geocoding depends on the quality of the input address. The geocoding success rate and precision goes down when input addresses are unable to be validated. ○ HUD programs operations and regulations require the use of demographic data that is only available at the Census tract or block group level. If you can't locate the property or the tenant in the correct geography, you are unable to relate the required demographic data to it. ○ Inadequate data on the location of assets and tenants inhibits HUD's ability to conduct research, evaluation, and monitoring activities on its programs. ○ The lack of complete data on the location of HUD's investments in Puerto Rico means the agency is unable to provide the level of transparency expected by our government and its constituents. ○ Poor data quality for HUD investments in Puerto Rico prevents the commonwealth from fully leveraging information technology solutions that HUD provides for its program participants and customers. • <i>Include project or product history that is relevant, data that demonstrates the scope of the problem, and/or images, including legacy product images (subject to governing disclosure requirements), which help explain the problem.</i> • There are several major challenges with Puerto Rico Addresses:

Appendix 1: HUD

- Use of an additional line in the address data structure called **Urbanization Name** which the USPS defines as “an area, sector or residential development within a geographic area” and is an essential part of the address format.
 - Unlike CONUS, street names and address ranges are repeated even within the same ZIP code. The urbanization name is often the only address element that makes the address unique.
- **Language and cultural differences** result in address data input that is extremely difficult to match with USPS databases.
- The **absence of a centralized source of addressing approval or street naming conventions**, either at the centralized government or at the local municipality level, may be at the root of the lack of standardization found in local mailing addresses.

How bad is the problem?

HUD is able to validate only 20% of the assisted, insured or owned properties in Puerto Rico.



Solution or Output

Please describe your Proven Practice in detail here.

- *Be clear on the status: implemented or proposed*
- *Solution description, including:*
 - *Data model/framework and data methods, new or already in use*
 - *Data access mechanism*
 - *Datasets involved, both acquired and generated*
 - *Data platforms or systems, analytics, visualization tools and practices used/developed*

HUD, for the most part, accepts addresses as they come. Recently, we have hired a contractor to improve the quality of a subset of addresses:

- PD&R supported CPD in acquiring the services of an on-island firm called Digital Media Creations (DMC) to assist in improving the quality of

Appendix 1: HUD

addresses in its administrative data on Puerto Rico and to make recommendations on how to improve the processes and systems HUD uses to collect addresses on the island.

- The goal of the project was not to try to fix every single address that HUD has collected in PR. We focused on a sample of ~100,000 addresses from the rental assistance stock.
 - Public Housing units
 - Housing Choice Voucher units
 - Multifamily assisted properties
 - Low Income Housing Tax Credit properties
 - HUD Single Family REO properties

Address cleansing was a largely manual effort



Address Standardization for HUD Summary Report

Date: April 6, 2018

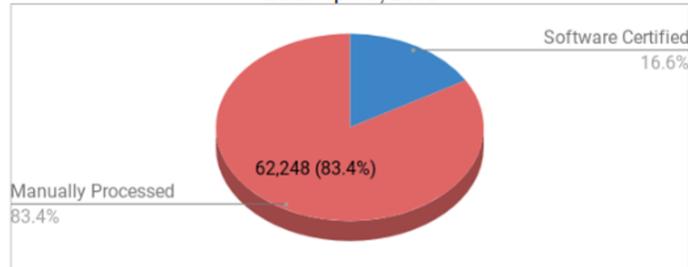


Table 3. CASS Results

Type *	Quantity	Percentage	Notes
Software Certified	12,415	16.6%	CASS software was able to process.
Manually Processed	62,248	83.4%	Modification to client data needed.
Total	74,663		* Values from "CASS_MatchType" data field



Appendix 1: HUD

A good address does not always translate into locational precision



Address Standardization for HUD Summary Report

Date: April 6, 2018

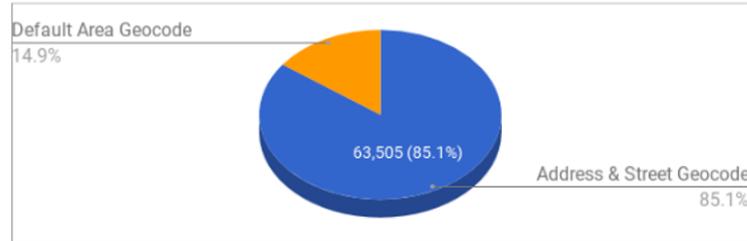


Table 7. Coded Addresses Geocoding

Type *	Quantity	Percentage	Notes
Address & Street Geocode	63,505	85.1%	Geocoded to approximate location
Default Area Geocode	11,158	14.9%	Geocoded to area (Urbanization, ZIP, etc)
Not Geocoded	0	0.0%	Addresses not geocoded
TOTAL	74,663		<i>* Values from "LOC_GeoArea" data field</i>



U.S. Department of Housing and Urban Development | Office of Policy Development and Research

Impacts – Focus on Outcomes

- Long term solutions:
 1. Change the data collection forms used in Puerto Rico to include a field for the “urbanization” name
 2. Change systems to do point of entry address validation
 - Integration of DMC’s API can help program participants or HUD staff find and enter the correctly formatted addresses
- Short term solutions:
 1. If the administrative system only allows as single address line: Provide guidance so that users from Puerto Rico use a delimiter (such as a "comma") to separate street information from “urbanization”
 2. If an address line 2 or address 3 are available there should be information to tell the user what is expected to provide in each line. For example: Address Line 2 (for PR addresses, use for Urbanization names)
 3. DMC produced a Guidance Document that can help HUD personnel and the general public format Puerto Rican addresses
 4. DMC also offers on-site technical training sessions for HUD staff on how to convert Puerto Rican addresses to USPS standards

Appendix 1: HUD

5. Use output provided by DMC processes to reach out to property owners and program participants to modify existing addresses when it comes time to recertify their participation in HUD programs.

Impact Measurements – Qualitative and Quantitative

Overall Results for Address Cleansing



Program	Count	Addresses Validated Before Cleaning	% Addresses Validated Before Cleaning	Addresses Validated After Cleaning	% Addresses Validated After Cleaning	% Point Change
Public Housing	59,074	1,323	2%	50,745	86%	84%
HCV	33,326	5,967	18%	23,008	69%	51%
Multifamily	216	39	18%	45	21%	3%
LIHTC	206	40	19%	56	27%	8%
HUD REO	513	114	22%	223	43%	21%
Total	93,335	7,483	8%	74,077	79%	71%




U.S. Department of Housing and Urban Development | Office of Policy Development and Research

Lessons Learned

The goal is to share and use knowledge derived from experience, both positive and negative, to promote the recurrence of desirable outcomes and preclude the recurrence of undesirable outcomes.

Tips on how to Replicate

Actions, resources or access required to replicate the work

Additional information

Please feel free to attach any relevant diagrams or images

Appendix 2: FEMA

PUERTO RICO ADDRESS DATA WORKING GROUP - PROVEN PRACTICES TEMPLATE

Organization	<i>Name of organization:</i> DHS/Federal Emergency Management Agency
Point of Contact:	Julia W. O'Brien Region II Geospatial Coordinator Julia.obrien@fema.dhs.gov 212-680-3698
Proven Practice in a Tweet	Enter everything in the street address field and sort it out later.
Submitted On	5/13/2019
Challenge or Opportunity	<ul style="list-style-type: none"> • <i>What problem are you trying to solve?</i> Collecting physical addresses of people applying for Individual Assistance (IA) in Puerto Rico and the USVI • <i>How is this problem related to the mission of your agency or agencies?</i> We are required to certify that homes of applicants have received sufficient physical damage to qualify for IA in order to provide financial assistance. • <i>Include project or product history that is relevant, data that demonstrates the scope of the problem, and/or images, including legacy product images (subject to governing disclosure requirements), which help explain the problem.</i> During Hurricanes Irma and Maria, thousands of applicants registered with the addresses they use. These could not be geocoded, as there are no standardized addressing systems in PR, and many of them did not relate to the physical home. Some examples are addresses that are highway markers: CARR 173 KM 7.1 BARRIO SUMIDERO SECTOR LA JACANA CARR 3 R 753 PAR 676 BUENA VISTA SECT PUNTA BRAVA Or addresses that are on numbered streets without an urbanization code: 7 CALLE 5, CANOVANAS, PR Or just building names: EDIF 5 APART 28 RES JARDINES DE CATANO, CATANO, PR Or addresses that make no sense: 15 3, GUARABO, PR 3 1, GUYNABO, PR Spreadsheet of samples from housing mission attached. • <i>Do other federal data practitioners face similar/related challenges or opportunities?</i> Yes, USACE also had this issue when identifying blue-roof mission homes. Spreadsheet from them also attached.
Solution or Output	<p><i>Please describe your Proven Practice in detail here.</i></p> <ul style="list-style-type: none"> • <i>Be clear on the status: implemented or proposed:</i> Implemented • <i>Solution description, including:</i> <ul style="list-style-type: none"> ○ <i>Data model/framework and data methods, new or already in use:</i> We entered the data, as received, into the street address field ○ <i>Data access mechanism:</i> The addresses are attached to the applicant information, and only accessed by applicant number. ○ <i>Datasets involved, both acquired and generated:</i> Internal IA databases ○ <i>Data platforms or systems, analytics, visualization tools</i>

Appendix 2: FEMA

	<p><i>and practices used/developed:</i> This is what we implemented, but it required using local guides to go into the field and find the homes for physical inspections. We can, in most states, use our remotely sensed damage assessments to speed the process. In PR, we were unable to tie that data to the ground without field work due to the lack of physical addresses or standard addressing system.</p> <ul style="list-style-type: none"> • <i>Outputs, directly produced from this project or product, including documentation, etc.</i> Nothing documented on methodology.
<p>Impacts – Focus on Outcomes</p>	<ul style="list-style-type: none"> • Internal improvements: <i>How this solution, if fully implemented, successfully benefits the sponsor? The federal government as a whole? The targeted user groups? n/a</i> • Public-facing: <i>Beyond the value to your agency, help us understand how the impact of this project will be seen and felt by affected policy-makers and residents of Puerto Rico? The lack of an ability to find physical locations tied to addresses slows delivery of needed services and aid.</i>
<p>Impact Measurements – Qualitative and Quantitative</p>	<ul style="list-style-type: none"> • <i>Include user stories, relevant research, customer citations, and</i> • <i>Identify key metrics (realized or for future) including market indicators etc.</i>
<p>Lessons Learned</p>	<p><i>The goal is to share and use knowledge derived from experience, both positive and negative, to promote the recurrence of desirable outcomes and preclude the recurrence of undesirable outcomes.: n/a</i></p>
<p>Tips on how to Replicate</p>	<p><i>Actions, resources or access required to replicate the work: As an agency, we would argue against replicating this if possible.</i></p>
<p>Additional information</p>	<p><i>Please feel free to attach any relevant diagrams or images: Spreadsheets of example addresses are attached.</i></p>

Appendix 3: USACE - NSI

PUERTO RICO ADDRESS DATA WORKING GROUP - PROVEN PRACTICES TEMPLATE

Organization	<i>US Army Corp of Engineers</i>
Point of Contact:	<p><i>William Lehman</i> <i>Senior Economist</i> William.P.Lehman@usace.army.mil 530.302.3736</p> <p><i>Nicholas Lutz</i> <i>Regional Economist</i> Nicholas.J.Lutz@usace.army.mil 502 315-6874</p>
Proven Practice in a Tweet	USACE uses an array of data sources to establish pseudo structure locations, structure characteristics, and population at-risk for risk assessment purposes.
Submitted On	<i>4/18/2019</i>
Challenge or Opportunity	<ul style="list-style-type: none"> <i>What problem are you trying to solve?</i> <p>Availability of Structure Inventories to support catastrophic modeling for all federal and state governmental agencies.</p> <p>The structure inventories should represent location, population, classification and valuation for each structure in the inventory.</p> <ul style="list-style-type: none"> <i>How is this problem related to the mission of your agency or agencies?</i> <p>USACE makes investments to reduce flood risk to persons and property across the United States. USACE uses risk assessments to help prioritize investments in new and existing projects and analyzing the potential consequences of hazardous flood events is an essential part of this effort. As a part of this process, USACE must identify the persons and property at-risk and determine how vulnerable this inventory may be to flooding. In coordination with FEMA, USACE created the National Structure Inventory to help solve this problem.</p> <p>Flood Risk Management is a priority mission area for USACE Civil Works efforts. An important element of the FRM mission is reducing risk of existing dams and levees; however, reducing risk is often expensive and available funds to reduce risk are finite. A risk informed framework is essential in order to prioritize investments and ensure public safety.</p> <ul style="list-style-type: none"> <i>Include project or product history that is relevant, data that demonstrates the scope of the problem, and/or images, including legacy product images (subject to governing disclosure requirements), which help explain the problem.</i> <p>In 2011, USACE created the National Structure Inventory. The NSI was heavily based on data released by FEMA that supported the HAZUS software. The NSI provided users with a portal wherein they may download point shapefiles for specific flood</p>

Appendix 3: USACE - NSI

analyses. However, structure location was only accurate to a subset of census blocks classified as “developed” based on remote sensing data, wherein they were distributed uniformly. Structure characteristics were also estimated from data over wide geographic areas.



The data has been updated three times now 2011, 2014, and the most recent release was in 2019. The data processing and access has improved in a variety of ways.

- *Do other federal data practitioners face similar/related challenges or opportunities?*

The Bureau of Reclamation, FERC and TVA all conduct risk assessments on their existing projects and FEMA routinely conducts flood consequence assessments. All risk assessments must consider at-risk persons and/or property.

Any agency interested in the impacts of catastrophic events on the structures has a related challenge or opportunity.

Solution or Output

Please describe your Proven Practice in detail here.

- *Be clear on the status: implemented or proposed*

In 2019, USACE implemented a major update to the National Structure Inventory. It is now less reliant on HAZUS/Census data and has vastly improved the XY locations of structures included in the inventory. However, many of the data sources used in this effort were not available for Puerto Rico, instead only the 50 states and the District of Columbia.



- *Solution description, including:*
 - *Data model/framework and data methods, new or already in use*
 - *Data access mechanism*
 - *Datasets involved, both acquired and generated*
 - *Data platforms or systems, analytics, visualization tools and practices used/developed*

The 2019 version of the NSI relies on numerous different data sources to inform structure counts, placement, characteristics, and population at-risk. The NSI is currently available internally to USACE analysts, however, efforts are underway to make the NSI portal accessible to Federal Agencies and other partners that have signed Data Use agreements with HIFLD.

A Data Use agreement with HIFLD is currently necessary as parcel data from CoreLogic is a key data sources for the NSI generation process. Parcel data is used to identify residential structures, and along with Microsoft footprint data, is used to refine structure placement. ESRI Business layer data is used to identify commercial structures, and structure placements are refined with building footprint and parcel data as well.

Appendix 3: USACE - NSI



Appendix 3: USACE - NSI

Source	Database	Dataset	Description
HAZUS	Bndrygbs.mdb	hzMeansCountyLocations	Provides county level price adjustments.
		hzExposureOccupB	Informs estimated dollar per square foot used in structure valuation.
		hzCensusBlock	Provides the structure building schemes and block type.
		flSchemeCoastal, flSchemeRiverine, flSchemeGLakes	Provides information on foundation type and height.
	MSH.mdb	flGenBldgScheme	Provides the construction type distributions and NFIP entry year for structures.
USACE	NSI 2015	Base layer	Used in any Census Block that lacks ESRI or CoreLogic data.
Homeland Infrastructure Foundation-Level Data	CoreLogic	County Level Data	Parcel polygons and associated data tables; used for initial spatial location and Occupancy Type.
Esri	Business Layer	InfoGroup	Provides initial structure location; NAICS code informs occupancy type, number of employee field informs square footage estimate and population weighting.
Microsoft	Building Footprints	State level polygons	Paired with parcel polygons to improve structure location and to inform structure aggregation.
U. S. Census Bureau	American Community Survey	Population, Demographics	Informs population growth estimates, disability rates, and age distribution.
	Characteristics of New Housing	Annual, Various	Provide structure characteristic data such as number of stories and square feet.
	Longitudinal Employer-Household Dynamic Database	Population Data	Contains worker counts by origin and destination census blocks.
NCES	Schools Database	School Data	Contains the locations of schools, number of teachers and students per school by census block.
U. S. Geological Survey	National Elevation Dataset	10 Meter Dataset	Provides raster ground elevation (in feet) data.

Appendix 3: USACE - NSI

NSI NAME	Description	Attribute Type	Limits
FD_ID	A number that should be unique for all structures.	INT 64	
CBFips	Census Block that contains the structure.	String	15 Characters
OccType	Damage Function or Occupancy Type of the structure. This field relates the structures depth-damage relationships, number of stories, number of households, and other generic characteristics to the structure location.	String	9char
YrBuilt	When available, records the parcel data assessment of the actual year built of the structure	Integer	
Num_Story	Estimated number of stories the structure has.	Integer	
RESUnits	Parcel informed number of residential units estimate.	Integer	
Stacked	Indicates whether multiple structures were stacked on top of each other during the generation process.	String	1 char
Source	Indicates the main source for the original xy and OccType data (P = Parcel, E = ESRI, N = NCES, H = HAZUS).	String	1 char
X	X coordinate of each structure; it is in the Geographic Coordinate System (GCS) WGS84.	Double	
Y	Y coordinate for each structure in GCS WGS84.	Double	
EmpNum	Source estimate of number of employees at a commercial structure	Integer	
Teachers	Source estimate of number of teachers at an educational structure	Integer	
Students	Source estimate of number of students at an educational structure.	Integer	
SqFt	Estimated total square footage for the structure.	Double	
Pop2amU65	Population at night for the structure of people under the age of 65.	Integer	
Pop2amO65	Population at night for the structure of people over the age of 65.	Integer	
Pop2pmU65	Population during the day for the structure of people under the age of 65.	Integer	
Pop2pmO65	Population during the day for the structure of people over the age of 65.	Integer	
St_DamCat	Damage category of the structure. Damage categories are a larger aggregation than occupancy type (Res = Residential, Com = Commercial, Ind = Industrial, and Pub = Public).	String	3 char
Basement	Defines if the structure has a basement or not (1 or 0).	Int 16	1 or 0
BldgType	Construction type of the structure. This is informative for structural stability criteria. From the HAZUS database this comes out of the MSH.mdb (W = Wood, H = Manufactured, S= Steel, C = Concrete, M = Masonry).	String	1 char
Found_Ht	Describes the foundation height of the structure in feet from the ground elevation.	Double	
Found_Type	Describes the type of foundation on the structure.	String	4 char
Val_Struct	Value in dollars of the structure. The analyst should document what value is represented (e.g., depreciated replacement cost, replacement cost, market value, etc.).	Double	
Val_Cont	Value in dollars of the contents of the structure.	Double	
Val_Vehic	Value in dollars of the vehicles at the structure.	Double	
Med_Yr_Blt	Describing a best estimate for the year the building was built.	Int 16	
FipsEntry	Value stating estimated year that the house entered the National Flood Insurance Program.	Int 16	
FIRMZone	When available, states the FEMA Flood Insurance Rate Map flood zone for the structure.	String	2 char
O65Disable	The estimated disability rate for the structure among those over the age of 65.	Double	
U65Disable	The estimated disability rate for the structure among those under the age of 65.	Double	
Ground_Elv	Ground elevation (in feet) at the structure.	Double	

The structures in the NSI are accessible through API access which supports a variety of queries and data formats. The current API is behind the USACE firewall. We are planning to migrate to a publically accessible API access with the new update to the NSI base data layer described above.

Since FEMA cannot currently access our API or website, we provide a copy of the data for them to use for their needs. They host that on their own internal FTP servers.

- *Outputs, directly produced from this project or product, including documentation, etc.*

The 2019 version of the National Structure Inventory contains millions of individual structures for the 50 states and the District of Columbia. Data for Puerto Rico is only available using the 2014 process.

Appendix 3: USACE - NSI

<p>Impacts – Focus on Outcomes</p>	<ul style="list-style-type: none"> • Internal improvements: <i>How this solution, if fully implemented, successfully benefits the sponsor? The federal government as a whole? The targeted user groups?</i> <p>The National Structure Inventory would need to expand its data sources to create a product for Puerto Rico that is comparable to what was generated for the states. This would aid decision-making for flood risk mitigation investments within the territory by improving the efficiency and effectiveness of flood risk management studies.</p> <ul style="list-style-type: none"> • Public-facing: <i>Beyond the value to your agency, help us understand how the impact of this project will be seen and felt by affected policy-makers and residents of Puerto Rico?</i> <p>If improved, it is likely that NSI would improve the quality of flood risk management projects within the territory. This would reduce damages from floods and reduce the likelihood of future flood related fatalities within the community.</p>
<p>Impact Measurements – Qualitative and Quantitative</p>	<ul style="list-style-type: none"> • <i>Include user stories, relevant research, customer citations, and</i> • <i>Identify key metrics (realized or for future) including market indicators etc.</i> <p>The National Structure Inventory is currently used by dozens of analysts and study teams across USACE. The 2014 version was made available to HAZUS users within FEMA and partnered state and local agencies.</p> <p>Analyses that utilized the NSI supported decision-making that led to hundreds of millions of dollars in dam and levee safety investments.</p>
<p>Lessons Learned</p>	<p><i>The goal is to share and use knowledge derived from experience, both positive and negative, to promote the recurrence of desirable outcomes and preclude the recurrence of undesirable outcomes.</i></p> <p>Data gaps and inconsistencies exist at many levels. Some data sources are unavailable for certain jurisdictions and others have inconsistent attribute classifications between jurisdictions. Special care must be made so that consistent quality of data is available across all areas of interest.</p>
<p>Tips on how to Replicate</p>	<p><i>Actions, resources or access required to replicate the work</i></p> <p>USACE possess code to replicate the process for existing jurisdictions; however, if others wish to adapt the process for other jurisdiction that may be feasible. For instance, using building footprint centroids to update structure counts and locations is possible for any area that has sufficient imagery or LIDAR data.</p>
<p>Additional information</p>	<p><i>Please feel free to attach any relevant diagrams or images</i></p> <p>Documentation and user guides are available upon request.</p>

Appendix 4: U.S. Census Bureau

PUERTO RICO ADDRESS DATA WORKING GROUP - PROVEN PRACTICES TEMPLATE

Organization	<i>US Census Bureau Geography Division</i>
Point of Contact:	Gregory Hanks Jr. Deputy Division Chief Gregory.F.Hanks.Jr@census.gov 301-763-3093
Proven Practice in a Tweet	The Census Bureau utilizes in-house tools and open source algorithms to cleanse and validate Puerto Rico address data collected from various sources.
Submitted On	<i>5/9/2019</i>
Challenge or Opportunity	<ul style="list-style-type: none"> <i>What problem are you trying to solve?</i> <p>The largest categories of address types in the Census Bureau’s Puerto Rico Address inventory – <i>Urbanizacion</i> and <i>Area Name</i> – illustrate the fact that the majority of Puerto Rico addresses do not match the conventional General/City-Style format, which makes it difficult for Census IT systems to match data unless they have some customizations in place.</p> <ul style="list-style-type: none"> <i>How is this problem related to the mission of your agency or agencies?</i> <p>These proven practices were developed to help the Census Bureau collect and maintain Puerto Rico addresses, which are needed to support the Decennial Census and many surveys.</p> <ul style="list-style-type: none"> <i>Include project or product history that is relevant, data that demonstrates the scope of the problem, and/or images, including legacy product images (subject to governing disclosure requirements), which help explain the problem.</i> <p>The Census Bureau maintains five (5) different varieties of Puerto Rico addresses in an internal address list, called the Master Address File, or MAF. Each of these addresses is residential in nature, representing a single or multi-family structure, an institutional or group quarters facility, or military housing. The Census Bureau maintains a separate inventory of commercial addresses for the U.S. and Puerto Rico, called the Business Register, but this inventory was not considered as part of this use case.</p> <p>The five common residential address types in PR and a fictional example of each:</p> <ol style="list-style-type: none"> General: 1234 Calle Juan Doe San Juan, PR 00926 Apartment Complex: 8 Vista Suites APT 8910 San Juan, PR 00926

Appendix 4: U.S. Census Bureau

3. Urbanizacion:

Urb Smith
567 Calle A
San Juan, PR 00926

4. Area Name:

Barrio Juan Sectore Smith
Carretera 12
San Juan, PR 00926

5. And three (3) kinds of 'Other' – Rural Routes and Highway Contract delivery route addresses, P.O. Boxes, and addresses with only descriptions of their locations.

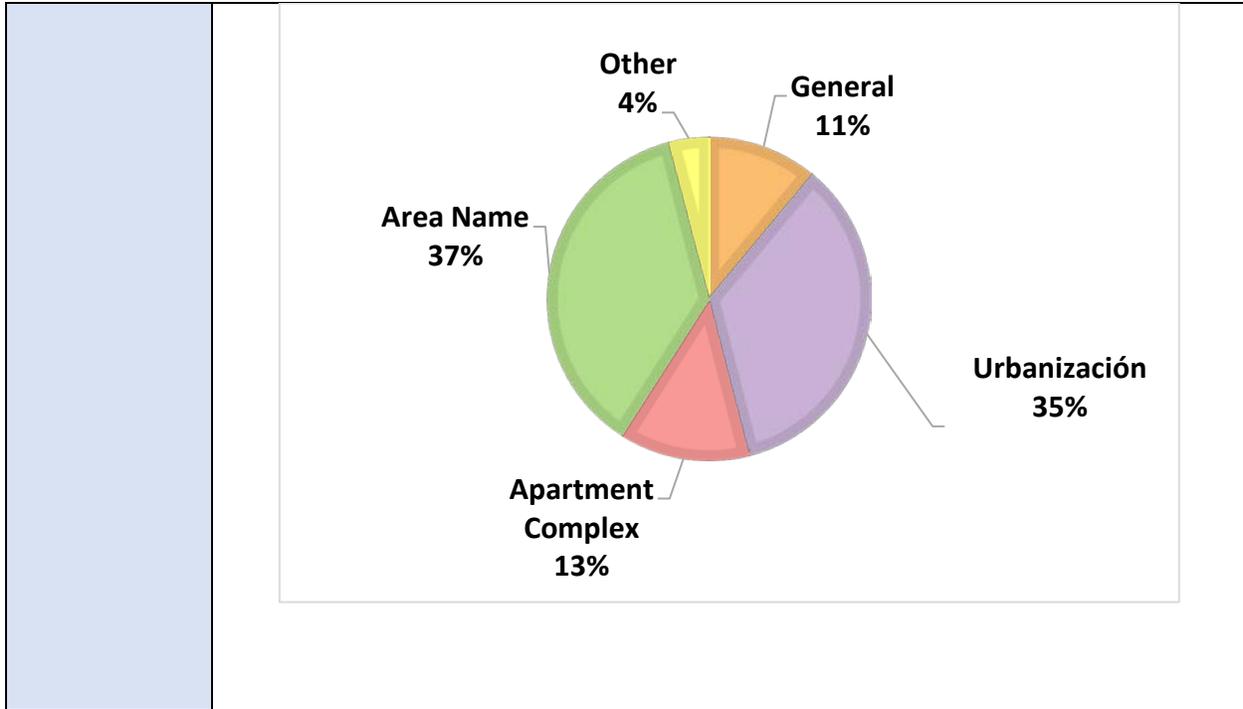
Examples:

- Some areas in PR do not have street names.
- Other areas have repetitive street names and house numbers within a ZIP Code, which are distinguishable only by the Urbanizacion name.
- There are also public housing projects (residenciales) without street names or repetitive apartment numbers.

These categories are helpful for understanding some of the variance encountered in the addresses the Census Bureau has collected, but they should not be wholly regarded as authoritative from a local government, community, or utility perspective. An exception is those addresses that are recognized and deliverable by the USPS, whose published standards (Pub. 28) have informed the definition of these categories.

Coverage for each type of address in Puerto Rico:

Appendix 4: U.S. Census Bureau



Solution or Output

Please describe your proven practice in detail here.

- *Be clear on the status: implemented or proposed*

The five components of the Census Bureau address management proven practices have been implemented and are improving our Puerto Rico address coverage.

- *Solution description, including:*
 - *Data model/framework and data methods, new or already in use*
 - *Data access mechanism*
 - *Datasets involved, both acquired and generated*
 - *Data platforms or systems, analytics, visualization tools and practices used/developed*

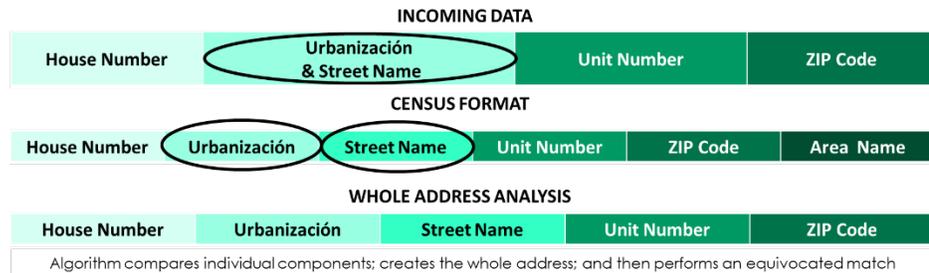
1. Standardize the Data - To account for various fields and words used in Puerto Rico addresses, the Census Bureau utilizes standardized naming conventions to capture and organize data.

ADDRESS ELEMENT	SOURCE VERSION	STANDARDIZED VERSION
Street Name	Calle 1	CLL 1
Unit Descriptor	Apartment	APT
Building Descriptor	Edificio	EDIF
Urbanización	Urbanización	URB
Apartment Complex Name	Residencial	RES
Area Name 1/Area Name 2	Barrio	BO
KNMHM	KM 01.5	1.5
Non-City Style	Rural Route 3	RR 3

2. Clean the Data - Recognizing and rearranging different fields and formats allows software to recognize more variety in how Puerto Rico addresses are represented.

Appendix 4: U.S. Census Bureau

In addition to standardizing language, the Census Bureau developed a method that enables it to “clean up” address data formatting and sequencing, enabling analysis using the whole address or individual address components.



- Match Data - The addition of a whole address matching solution bolsters the typical ‘exact’ and ‘equivocated’ match formulas that are typically used on incoming data.

The Census Bureau match process uses both individual address components as well as the **whole address**.



NOTE: This example is a fictional address that is not protected from disclosure by the Data Stewardship and Confidentiality requirements of 13 U.S.C. §9.

24

In Practice: Standardize, Clean and Match Data

A combination of the standardization, cleaning and matching data steps

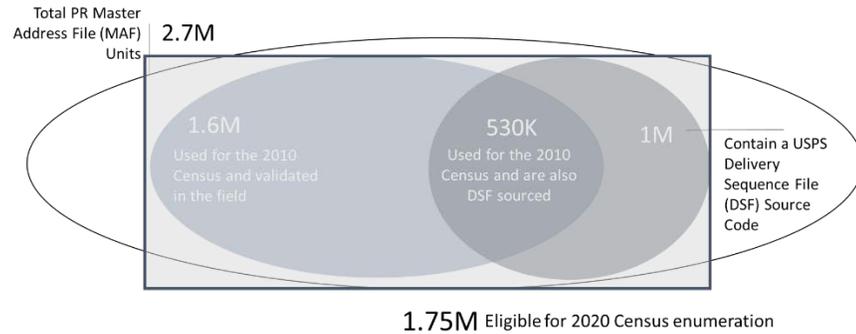
Phase	Example 1	Example 2
Incoming Address	Calle Main 11A Urbanizacion Coastal Gardens	Condominio San Juan Garden Edificio A Apart 100
Standardized Address	CLL MAIN 11A URB COASTAL GARDENS	COND SAN JUAN GARDEN EDIF A APT 100
Equivocated Match	CLL MAIN A11 URB COASTAL GRDNS	COND SAN JUAN GRDN BLDG A UNIT 100

NOTE: These examples are fictional addresses that are not protected from disclosure by the Data Stewardship and Confidentiality requirements of 13 U.S.C. §9.

Appendix 4: U.S. Census Bureau

4. Maintain Data Source and History – Source and history information enables the Census Bureau to develop a longitudinal pattern for an address that can be compared to and validated by other authoritative address sources.

This proven practice enables the Census Bureau to establish a longitudinal pattern that can be compared to other sources and helps validate that an address should be included in the Decennial Census and other surveys.



The more successful sources you can attach to an address, the higher the confidence that it's an accurate address and should be included in Census enumeration efforts

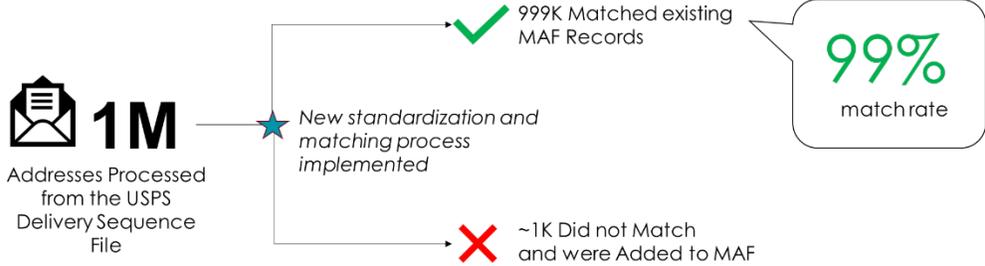
5. Gather Data Regularly - Collecting data regularly allows the Census Bureau to maintain records that capture change in addresses, enabling the agency to manage information requests and deliver on mission objectives.

The Census Bureau's MAF is built from multiple sources and partners in an attempt to maintain a current, comprehensive inventory of accurate data – the United States Postal Service, the Commonwealth of Puerto Rico, and Census field operations.

The Census Bureau performed an island-wide refresh of our Puerto Rico address inventory from the 2010 Census using the Postal Service's Delivery Sequence File in 2016. As a result of new and upgraded standardization and matching processes implemented after 2010, nearly 60% of the DSF matched to an address in the inventory. This is a very successful outcome, and added confidence in the quality and usability of our existing addresses. The 40% that did not match were added to the MAF as prospective new records.

- *Outputs, directly produced from this project or product, including documentation, etc.*

Appendix 4: U.S. Census Bureau

	<p>After five (5) "refreshes" of the MAF with the Delivery Sequence File, the rate of new addresses has stabilized at approximately 1% every six (6) months. Our upgraded processes enable the Census Bureau to track change and build data sets with a high level of confidence.</p>  <p>Addresses Processed from the USPS Delivery Sequence File: 1M</p> <p>New standardization and matching process implemented</p> <p>999K Matched existing MAF Records</p> <p>~1K Did not Match and were Added to MAF</p> <p>99% match rate</p>
<p>Impact Measurements – Qualitative and Quantitative</p>	<ul style="list-style-type: none"> As a result of these proven practices, the Census Bureau is able to track address activity (inclusion in Censuses) and measure Census universe growth. (See above Venn Diagram) Of the total MAF addresses, nearly 1.75M addresses are validated and in good condition to be used for enumeration in Census 2020. Ninety-four percent of these validated addresses were successfully used as part of the 2010 Census. This represents an increase of approximately 150,000 new addresses for the 2020 Census. Of the 1M addresses from our 2020 Census enumeration universe that have been validated by the USPS Delivery Sequence File, 30% of the total (530,000) of these were validated prior to the 2010 Census nine years ago.
<p>Lessons Learned</p>	<ul style="list-style-type: none"> Dedicate time to understand the address data. Adapt process to manage data as knowledge evolves. Communicate with others about knowledge and usage of PR address data. Evolve practices with data. If database contains both stateside and PR addresses then be consistent where possible. Develop and maintain documentation of practices. Maintain source information.
<p>Tips on how to Replicate</p>	<ul style="list-style-type: none"> Prepare data repository with necessary fields to handle PR values, e.g. address components, address types, source information, etc. Classify the addresses into the five different address types. Standardize the address data. Clean incoming and existing data. Acquire a string comparator and run against incoming address data. Record source information for incoming address data. Gather data from reliable sources and continually update address database.

Appendix 5: USDA

PUERTO RICO ADDRESS DATA WORKING GROUP - PROVEN PRACTICES TEMPLATE

Organization	<i>United States Department of Agriculture – Farm Service Agency</i>
Point of Contact:	<i>Kerry Sefton Agricultural Program Specialist kerry.sefton@wdc.usda.gov (202) 720-6120</i>
Proven Practice in a Tweet	<i>FSA utilizes Modernize and Innovate the Delivery of Agricultural Systems Customer Relationship Management (MIDAS CRM) as the system of record/system of entry for address data.</i>
Submitted On	<i>6/12/2019</i>
Challenge or Opportunity	Data collection and validation helps ensure that the correct producers are identified on a farming interest and also helps ensure the eligibility of producers requesting FSA program benefits.
Solution or Output	<p>Solution:</p> <ul style="list-style-type: none"> • Modernize and Innovate the Delivery of Agricultural Systems Customer Relationship Management (MIDAS CRM) • Implemented by FSA in 2013 as the system of record/system of entry for master data (producers and farms) • MIDAS CRM resides on a customized SAP CRM solution. <p>The two components of MIDAS CRM:</p> <ul style="list-style-type: none"> • Business Partner – producer names, Tax IDs, addresses, phone numbers, email addresses, demographic and social statistics, dates of birth and death, relationships and signature authority, and links to FSA and NRCS field offices. • Farm Records - customer farming interest data including the geospatial delineation of farm, tract, & field boundaries; operators, owners, and other producers with interest in the farm; land usage including cropland, farmland and associated acreages, conservation compliance determinations. <p>The Business Partner component includes the following data validations:</p> <ul style="list-style-type: none"> • Tax ID verification, via interface with IRS • Date of Death updates, via weekly file load from SSA • USPS Address validation, via SAP Data Services DQM • Duplicate customer check, via SAP Data Services DQM <p>Service provided by partnership between SAP and USPS</p> <ul style="list-style-type: none"> • USPS sends address directories to SAP • MIDAS CRM downloads the USPS address directories from SAP Market Place - refreshed every 3 months • As addresses are loaded or edited in MIDAS CRM/Business Partner, data is validated against the most current directory

Appendix 5: USDA

Address Component	Original Address	Validated Address
Address Line 1	487 CALLE SIERA MORNA	487 CALLE SIERRA MORENA
Secondary Number		
Street 2		
Address Line 2		
DPV/Carr./Cong.dist		679 C048 7201
Building Code		
Room Number		
Floor		
City	SAN JUAN	SAN JUAN
ZIP Code	00926	00926-5574
State	PR	PR
Country	US	US

Accept **Use Original Address** **Cancel**

USPS Partnership Cont:

- DQM also has a batch process feature that facilitates address validation of the entire dataset in a batch mode
- When USPS address directories are refreshed quarterly from SAP Market Place, the batch mode validation is run to update/correct all applicable addresses

Impact Measurements – Qualitative and Quantitative

FSA currently has a 79% USPS-validation rate for active customer addresses in Puerto Rico.

Category	Count
USPS-Validated	17,842
non-validated	4,692

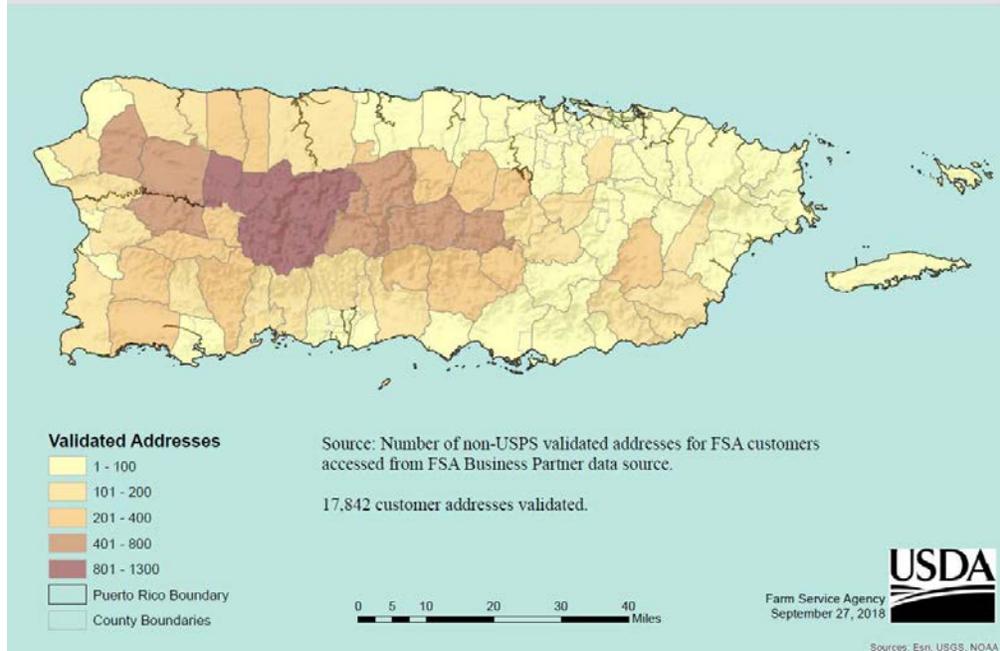
Addresses that validate successfully usually include:

- PO Boxes

Appendix 5: USDA

- HC (Highway Contracts) in range
- Simple urban addresses
 - i.e.: 15 Calle Coral, 261 Ave Ponce De Leon, etc.
- Urbanization Codes along with a street number and name

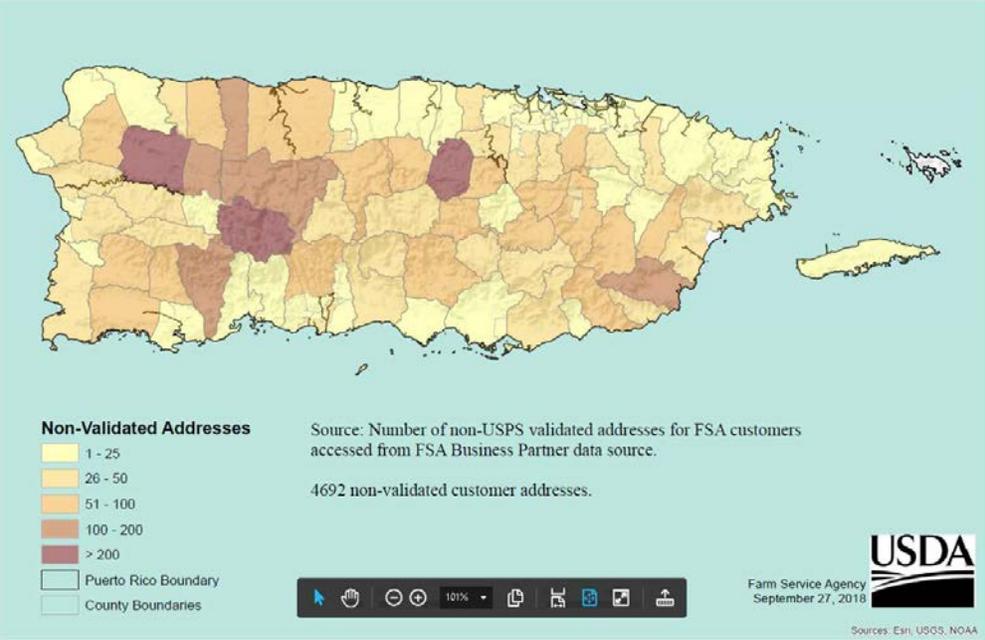
Number of FSA Customers by Zipcode with USPS Validated Addresses



Addresses that fail validation usually include:

- Urbanization Codes without a street number or name
 - i.e.: Urb El Maestro
- HC (Highway Contracts) out of range
 - The HC is valid, but provided street number is invalid
- PMB (Private Mail Boxes)
 - USPS validation cuts off the PMB #, workaround is to add it to Information line

Appendix 5: USDA

	<p>Number of FSA Customers by Zipcode with Non-USPS Validated Addresses</p>  <p>Non-Validated Addresses</p> <ul style="list-style-type: none"> 1 - 25 26 - 50 51 - 100 100 - 200 > 200 <p>Source: Number of non-USPS validated addresses for FSA customers accessed from FSA Business Partner data source. 4692 non-validated customer addresses.</p> <p>USDA Farm Service Agency September 27, 2018 Sources: Esri, USGS, NOAA</p>
<p>Lessons Learned</p>	<p>FSA Response to 2017 Hurricanes</p> <ul style="list-style-type: none"> • Non-validated addresses were not a major pain point for FSA following the 2017 hurricanes (Irma, Maria) <ul style="list-style-type: none"> ○ Producers came to FSA Offices for assistance ○ Newsletters/announcements were not mailed • Dairy Assistance Program (DAP-PR) and Wildfires and Hurricanes Indemnity Program (WHIP) were both successfully implemented following the 2017 hurricanes to provide assistance to affected producers regardless of a USPS validated address • FSA staff from county offices in continental US were detailed to PR to assist with disaster program operations
<p>Tips on how to Replicate</p>	<p>The MIDAS CRM solution has worked well for USDA/FSA business, but is a COTS package that was customized for the business need. It is not open source. Licensing fees and development were required, but they system has greatly benefited the agency</p>
<p>Additional information</p>	<p><i>Please feel free to attach any relevant diagrams or images</i></p>

Appendix 6: USPS

PUERTO RICO ADDRESS DATA WORKING GROUP - PROVEN PRACTICES TEMPLATE

Organization	<i>United States Postal Service</i>
Point of Contact:	<i>Aisha Ahmad Address Management Support Aisha.Ahmad@usps.gov (202) 268-4813</i>
Proven Practice in a Tweet	The United States Postal Services leverages carrier knowledge, work with local addressing authorities, and analyze delivery patterns to and validate delivery addresses nationally, including Puerto Rico.
Submitted On	
Challenge or Opportunity	<ul style="list-style-type: none"> • What problem are you trying to solve? Keep a current and accurate database of all mailing addresses (business and residential) in the United States for the purpose of mailing. • How is this problem related to the mission of your agency or agencies? Our mission is given to us in Section 101(a) of Title 39 of the U.S. Code, also known as the Postal Reorganization Act: The Postal Service shall have as its basic function the obligation to provide postal services to bind the Nation together through the personal, educational, literary, and business correspondence of the people. It shall provide prompt, reliable, and efficient services to patrons in all areas and shall render postal services to all communities. Scope: <ul style="list-style-type: none"> • The USPS address database contains 157M address records with 1M new address records added annually.
Solution or Output	<p><i>Please describe your proven practice in detail here.</i></p> <ul style="list-style-type: none"> • <i>Provide address matching software to customers and list processors</i> • <i>Offer Address correct services to inform mailers of potential address changes</i> • <i>Provide Delivery Point Validation tools which provide insight on deliverable or none deliverable addresses by the USPS</i>
Impacts – Focus on Outcomes	<ul style="list-style-type: none"> • Internal improvements: <i>Allows for expedited processing of mail across mail sorting equipment and faster delivery to the customers.</i> • Public-facing: <i>Customer can make better business decisions on when to mail or provide services to an address. Business or residence can be assured proper and prompt delivery.</i>
Impact Measurements – Qualitative and Quantitative	<ul style="list-style-type: none"> • <i>N/A</i>

Appendix 6: USPS

Lessons Learned	<i>Address hygiene has improved significantly over the years through collaboration with local addressing authorities, software vendors and direct correspondence with the customers.</i>
Tips on how to Replicate	<i>USPS has made address products and services available for purchase for stakeholders to use as part of their address data management practices.</i>
Additional information	

Available Resources, Activities, and Standards for Managing Puerto Rico Address Data

The following information represents available resources, activities, and standards that the PRADWG and its subgroups discussed or referenced for consideration by agencies who manage address data for Puerto Rico.

1. The National Address Database and the FGDC Address Subcommittee – The National Geospatial Advisory Committee (NGAC) 2012 report ‘The Need for a National Address Database’ set forth the vision of the NAD as “an authoritative and publicly available resource that provides accurate address location information”¹. As an open data resource aggregated by the Department of Transportation (DOT), the NAD supports many needs for address data across all levels of government, for the public and private industry. The expected main use case of the NAD is to serve as a publically available nationwide geocoding data source (NAD Pilot Project Final Report, 2016). Puerto Rico address data needs to be included in the NAD in order to meet its full potential. The NAD is a National Geospatial Data Asset (NGDA) of the FGDC Address Theme. The lead agencies for the Address Theme and the Address Subcommittee are the Department of Transportation and the Census Bureau. The Address Subcommittee consists of address data stakeholders from all levels of government and meets monthly to support ongoing efforts to develop the NAD.
 - a. Adding Puerto Rico Data to the NAD. Currently there are addresses from 22 states, with 3 more in progress. There are 47 million address points. Puerto Rico has not yet contributed addresses to the NAD. This is a goal going forward, and Puerto Rico address data stakeholders are taking action to add data for Puerto Rico/
 - b. Adding Data Attributes Specific to Puerto Rico Addresses to the NAD: The Content Subgroup of the Address Subcommittee has recommended modifications to the minimum content requirements for the NAD, to accommodate the unique requirements of Puerto Rico addresses. Additional place name fields are proposed to hold the urbanizacion attribute.
2. Open Data Ecosystem for Puerto Rico Address Data— The Federal Evidence-Based Policymaking (FEBP/Evidence Act), and the Open, Public, Electronic, and Necessary (OPEN) Government Data Act will help sustain the open data ecosystem by requiring access to federal government data. In addition, the Federal Data Strategy is being designed “to fully leverage the value of federal data for mission, service, and the public good.” In addition to the NAD, open sources for address data that should be explored include:
 - a. Openaddress.io – this has open addresses by state and there are 135,000 addresses for Puerto Rico
 - b. Puerto Rico Open Data Interconnection Portal (data.pr.gov) – this is the source of the openaddress.io data and there are many other datasets here including cadastral and transportation.
 - c. Open Street Map – OSM address point coverage for Puerto Rico should be analyzed.
 - d. USA Structures database (Puerto Rico dataset to be released in the summer of 2019)— Oak Ridge National Lab (ORNL) is working with FEMA to map the entire structural landscape of the United States using satellite imagery to create models to identify and characterize residences, businesses, schools, and hospitals in the United States. These

structural maps are paired with damage reports to direct emergency response and recovery efforts (<https://disasters.geoplatform.gov/publicdata/Partners/ORNL/>).

3. Data Sharing Opportunities and Tools – Organizing around cross-cutting challenges means the sharing of procedures and tools will be important.
 - a. The Opportunity Project (TOP) is a process for engaging government, communities, and the technology industry to create digital tools using federal open data to address the nation’s greatest challenges. Census Bureau participated in the recent TOP geo-cohort sprint to develop address data collection tools that help local governments including those in Puerto Rico, to create and maintain open address point data. Three private sector tech teams accepted the challenge and created address data collection tools based upon the National Address Database (NAD) schema, and in one case, utilized the FGDC United States Thoroughfare, Landmark, and Postal Address Data Standard (FGDC Address Standard), to improve the quality of the data entered. A summary report for this effort is available on the Census Bureau website: <https://opportunity.census.gov/>.
 - b. The U.S. Data Federation is a General Services Administration effort to curate reusable tools and repeatable processes that support combining or exchanging data across disparate systems. Recently funded for a fourth phase of work through the 10x program, the U.S. Data Federation will contribute to the establishment of a repository of tools, standards, guides, and other resources as required by the Federal Data Strategy and the Foundations for Evidence-Based Policymaking Act of 2018. More information can be found on GitHub: <https://github.com/GSA/us-data-federation> and at their website: <https://federation.data.gov>.

4. Address Standards –The FGDC Address Standard, the National Emergency Number Association (NENA) United States Civic Location Data Exchange Format (CLDXF) Standard (NENA CLDXF Standard), and the National Address Database (NAD) schema can also assist local governments in Puerto Rico. The NAD schema is largely based on the NENA CLDXF Standard. Federal agencies, state and local governments, and private industry with a critical mission need for nationwide authoritative address points have benefitted from the use of these standards. Address data collection tools developed for Puerto Rico from these addressing standards will also increase data quality and facilitate efficient data exchange. (See also The Opportunity Project, section 3a).