Facility Registry Service (FRS)

FRS Representative Point Processing

JULY, 2013

Prepared for: ENVIRONMENTAL PROTECTION AGENCY

Prepared by: INDUS Corporation
1951 Kidwell Drive
Vienna, Virginia 22182

TABLE OF CONTENTS

PREPARED FOR: I
TABLE OF CONTENTS II
REVISION HISTORY III
A. INTRODUCTION – FRS REPRESENTATIVE POINT PROCESSING 4
B. REPRESENTATIVE POINT PROCESSING STEPS 4
B.1 Representative Point Processing Steps 4
C. APPENDIXES 7
C.1 Appendix A, Collection Method Default Accuracies 7
C.2 Appendix B Maximum NAD27 to NAD83 Datum Shift By State 10

REVISION HISTORY

Version Number	Date	Description of Changes
1.0	04/08/2013	Initial creation.
1.1	07/01/2013	Updated to include Program as part of EPA verification group.

A. INTRODUCTION – FRS REPRESENTATIVE POINT PROCESSING

The FRS Representative Point Processing occurs on the Envirodb (Envirofacts) Oracle database server at RTP within the ENVIROI (Intranet) database instance. Co-located with FRS in the ENVIROI database instance, is the Office of Environmental Information's (OEI) Integrated GeoDatabase (IGD). The IGD hosts spatial reference tables like Zip Code, State, County spatial boundary tables that are used to both quality assure FRS Program spatial coordinates and to produce derived spatial information like US Census Block code, watershed and Congressional District. At a high level, this processing performs the following:

- Identify changes in facility addresses, facility linkages, and program coordinate spatial data from the last processing run, so only the changed or new facilities will need to be reprocessed
- Geocodes all FRS Facility Site records
- Standardizes all FRS Program Facility spatial coordinates to the US Standard NAD83 horizontal datum and stores the standardized values in an Oracle Spatial geometry column
- Creates a standard set of spatially derived data for each FRS Program Facility spatial coordinate
- Creates an accuracy score for each FRS Program spatial coordinate based on supplied metadata and quality assurance checks
- Based on the accuracy score, identifies the Representative Point for each FRS Facility based on the related (linked) FRS Program Facility spatial coordinates

B. REPRESENTATIVE POINT PROCESSING STEPS

B.1 REPRESENTATIVE POINT PROCESSING STEPS

These are the steps, executed in the following order, that encompass FRS Representative Point Processing. Prior to beginning this step, all new FRS Program Facility Coordinates from FRS Program and State refreshes and new FRS Facility coordinates entered through the FLA have been added to the GEO Program Coordinate table.

1. Create a "DELETE FACILITIES" table by identifying all FRS Facility Site records where the facility site address has changed using the facility address stored on the Facility Geocodes table. The Facility Geocodes table represents a baseline of FRS facilities as of the last successful FRS Representative Point Processing run.
2. Add to the "DELETE FACILITIES" table all facilities on the Facility Geocodes table that have been deleted or merged out of existence from FRS.
3. Delete from the Facility Geocodes table all facilities that are listed in the "DELETE FACILITIES" table.
4. Delete from the Program Coordinate table all FRS geocodes that are associated with Facilities listed in the "DELETE FACILITIES" table.
5. Delete from the Program Coordinate QA table and Program Coordinate QA Flags table all rows associated with those facilities in the "DELETE FACILITIES" table (address has changed or deleted from FRS).
6. Add to the "DELETE FACILIITES" table all facilities for which a new FRS Program spatial coordinate has been added or modified.

7. Add to the "DELETE FACILITIES" table all facilities for which the current representative program coordinate has been deleted from FRS through the refresh process.
8. Add to the "DELETE FACILITIES" table all facilities for which a new "Verified" location has been identified in the FLA.
9. Add to the "DELETE FACILITIES" table all facilities for which a new NPL Program Facility coordinate has been added.
10. Delete from the Facility Representative Point table and the Facility Buffer QA Flags table all records identified in the "DELETE FACILTIES" table.
11. Create a "DELETE COORDINATES" table for all existing GEO Program Facility Coordinates that have been deleted or modified.
12. Delete from the Program Coordinate QA Flags table all records identified in the "DELETE COORDINATES" table.
13. Delete from the Program Coordinate Spatially Derived table all records identified in the "DELETE COORDINATES" table.
14. Delete from the Facility Representative Point table all records identified in the "DELETE COORDINATES" table.
15. Due to license restrictions with the dissemination of geocoded coordinates returned from Google's, Bing's or ESRI's geocoders, FRS uses Oracle Spatial's Geocoder routines along wih NAVTEQ Point Address data to geocode all FRS Facilities. Using the Oracle Spatial Geocoder, all new FRS Facilities and all FRS Facilities where any part of the address has been modified are geocoded. The geocoding process follows these steps:
    a. Modify facility address replacing certain abbreviations and address components with values that the Oracle Geocoder can more readily recognize. Examples include replacing "TWP" with "TOWNSHIP OF", "BORO" with "BOROUGH OF".
    b. Identify Intersections within the FRS location address and identify each intersecting street.
    c. Geocode a formatted address using Oracle Spatial's geocoder.
    d. Validate the returned geocode object from the geocoder
        i. Reject returned geocodes that do not meet FRS accuracy standards. Accept only Point (aka rooftop), House Number, Street, Intersection, and Landmark Placename Geocodes.
        ii. For Point and Housenumber geocodes validate that the returned Housenumber matches the supplied housenumber. If not, convert the geocode to "Street" geocode.
        iii. Using the Jaro-Winkler and Levenshtein Distance alogrithms validate the returned street name/placename against the supplied street name/placename. If they don't match, reject the geocode.
        iv. For street level geocodes, identify the number of blocks for the street and the length of the street. If the street is longer than 2km and if the street contains more than 2 blocks, reject the geocode.
        v. Validate state returned from the geocode. If it doesn't match the supplied state, reject the geocode.
        vi. Validate the zip code, county and city together returned from the geocode using USPS dataset. In certain conditions, Oracle will return the county name instead of the city name for rural areas. If the city, county, and zip can not be validated, reject the geocode.
    e. For accepted geocodes, set the level of geocode returned:
        i. Point (rooftop)

ii. Housenumber
iii. Intersection
iv. Street with 1 block
v. Street with 2 blocks or Street Length < 2km
vi. Landmark Placename

16. Create new Program Coordinate records for all new valid FRS geocodes. According to the level of geocode that was performed assign the following accuracy values to the corresponding geocode level:
    a. Point (Rooftop/Centroid) – 30 meters
    b. Street House Number – 150 meters
    c. Intersection – 200 meters
    d. Street of only 1 block – 500 meters
    e. Street of only 2 blocks or < 2km in length – 2000 meters
    f. Placename – 4500 meters
17. Standardize all new and modified Program Coordinate records standardizing to the NAD83 horizontal datum and creating the standardized Oracle Spatial Geometry.
18. Create new Program Coordinate Spatially Derived records for new and modified Program Coordinates.
19. Using the Program Coordinate Spatially Derived results, create new Program Coordinate QA Flag records for new and modified Program Coordinates. QA Flags indicate whether or not the Program Coordinate is within the Zip Code, City, County and State boundaries of the FRS Facility.
20. Using Program Coordinate metadata for Collection Method, Horizontal Datum, Accuracy Value, and Reference Point, calculate the Accuracy Score for all new and modified Program Coordinate records. The Accuracy Score is a value, expressed in meters, estimating the accuracy of the collected Program Coordinates. The following process is used to calculate the Accuracy Score:
    a. If the collection method is missing then set the accuracy score to 17,400. This represents the square root of the average area of all zip codes. So a “missing” collection method defaults to the default accuracy of a zip code centroid. If the accuracy value is present it is ignored because it can not be validated against the collection method.
    b. If the accuracy value is null, then calculate the accuracy score based on the default accuracy for the collection method. If the accuracy value is not null, verify the maximum accuracy allowed for the collection method. If the reported accuracy value is greater than the maximum accuracy allowed for the collection method, then the accuracy score is equal to the accuracy value, else the accuracy score is equal to the default accuracy based on the collection method. Appendix A, Collection Method Default Accuracies lists the Default Accuracy Score and Maximum Allowed Accuracy for each collection method. This prevents situations where there is a reported accuracy value of “5 meters” but the coordinate was obtained from map interpolation of a 1:100000 scale map.
    c. If the horizontal datum value is null and the collection method is one where NAD27 could be used as the horizontal datum (example map interpolation), then add to the accuracy score the maximum shift in meters between NAD27 and NAD83 coordinates, for the state where the coordinate is located. Appendix B, Maximum NAD27 Datum Shift By State, identifies the datum shift values by state.
    d. Next, using the qa_flags for the program coordinate, the accuracy score is adjusted if the program coordinate flunked any of the spatial boundary checks. To the accuracy score the following values are added:
        i. Flunk State Boundary – add 999999999.

ii. Flunk County Boundary – add 64228.
iii. Flunk City Boundary – add 18500.
iv. Flunk Zip Code Boundary – add 17400.

e. If the reference point is null or if the reference point is a facility boundary or a location on the street then add 30 to the accuracy score.

21. For Facilities that don't have a representative point identified, pick the most accurate Program Coordinate as the representative point. The following method is used:
    a. Identify if an NPL Superfund Program Coordinate exists for the Facility. If so, the latest coordinate obtained from the Superfund system becomes the representative point.
    b. If no NPL Superfund Program Coordinates exists, identify if an existing Program Coordinate has been verified by the EPA (Region, FLA Data Steward, or Program). If one exists, then the latest coordinate verified becomes the representative point.
    c. If no verified coordinates exists, then select from the entire relevant program coordinates the one with the best accuracy score. This becomes the representative point.
22. For those facilities where the representative point has flunked either the state or county boundary checks, recheck the boundaries with a one-mile buffer. For those that still flunk the buffered boundary, delete the representative points.
23. Identify from the representative points, which coordinates are from a restrictive system, like LANDFILL, and if another non-restrictive program interest exists. On the public web site and the Envirop instance, these coordinates will have their origin modified from the restricted program system to the "FRS" program.
24. Using a database link to the public ENVIROP database, create tables that identify deleted program coordinates, new and modified program coordinates, deleted representative points, and new/modified representative points.

C. APPENDIXES

C.1 APPENDIX A, COLLECTION METHOD DEFAULT ACCURACIES

Of the selection criteria listed below, those that have an "Equal" value under the Search Operator column will provide the best performance. Examples are provided at the end of the section to illustrate how selection criteria can be combined.

Collect Method Code	Collect Method Description	Scale Lowerbound	Scale Upperbound	Default Accuracy	Maximum Accuracy	WGS84/ NAD83 Default Datum
001	ADDRESS MATCHING-HOUSE NUMBER			150	30	Y
002	ADDRESS MATCHING-BLOCK FACE			2000	1000	Y
003	ADDRESS MATCHING-STREET CENTERLINE			4500	500	Y
004	ADDRESS MATCHING-NEAREST INTERSECTION			1000	100	Y

Collect Method Code	Collect Method Description	Scale Lowerbound	Scale Upperbound	Default Accuracy	Maximum Accuracy	WGS84/ NAD83 Default Datum
005	ADDRESS MATCHING-PRIMARY NAME			4500	1000	Y
006	ADDRESS MATCHING-DIGITIZED			2000	30	Y
007	ADDRESS MATCHING-OTHER			5000	150	Y
008	CENSUS BLOCK-1990-CENTROID			2000	2000	
009	CENSUS BLOCK/GROUP-1990-CENTROID			4000	4000	
010	CENSUS BLOCK/TRACT-1990-CENTROID			5000	5000	
011	CENSUS-OTHER			17400	1000	
012	GPS CARRIER PHASE STATIC RELATIVE POSITION			1	1	Y
013	GPS CARRIER PHASE KINEMATIC RELATIVE POSITION			5	1	Y
014	GPS CODE (PSEUDO RANGE) DIFFERENTIAL			5	1	Y
015	GPS CODE (PSEUDO RANGE) PRECISE POSITION			10	1	Y
016	GPS CODE (PSEUDO RANGE) STANDARD POSITION (SA OFF)			10	1	Y
017	GPS CODE (PSEUDO RANGE) STANDARD POSITION (SA ON)			100	1	Y
018	INTERPOLATION-MAP	20001	24000	50	15	
018	INTERPOLATION-MAP			300	50	
019	INTERPOLATION-PHOTO	20001	25000	25	15	Y
019	INTERPOLATION-PHOTO			50	10	Y
020	INTERPOLATION-SATELLITE			200	100	
021	INTERPOLATION-OTHER			300	100	
022	LORAN C			500	185	Y
023	PUBLIC LAND SURVEY-QUARTER SECTION			805	805	
024	PUBLIC LAND SURVEY-SECTION			2000	1612	
025	CLASSICAL SURVEYING			1	0	Y

Collect Method Code	Collect Method Description	Scale Lowerbound	Scale Upperbound	Default Accuracy	Maximum Accuracy	WGS84/ NAD83 Default Datum
	TECHNIQUES					
026	ZIP CODE-CENTROID			17400	17400	
027	UNKNOWN			17400	17400	
028	GPS - UNSPECIFIED			100	5	Y
029	GPS, WITH CANADIAN ACTIVE CONTROL SYSTEM			25	1	Y
030	INTERPOLATION - DIGITAL MAP SRCE (TIGER)			400	20	Y
031	INTERPOLATION - SPOT			200	200	
032	INTERPOLATION -MSS			200	200	
033	INTERPOLATION -TM			200	200	
034	PUBLIC LAND SURVEY - EIGHTH SECTION			570	570	
035	PUBLIC LAND SURVEY - SIXTEENTH SECTION			402	402	
036	PUBLIC LAND SURVEY - FOOTING			15	15	
037	ZIP+4 CENTROID			1000	1000	
038	ZIP+2 CENTROID			6000	6000	
039	GDT-ADDRESS MATCHING (GEOCODING)			150	150	Y
101	THE GEOGRAPHIC COORDINATE DETERMINATION METHOD BASED ON ADDRESS MATCHING			150	50	Y
102	US BUREAU OF CENSUS BLOCK ESTABLISHED FOR YEAR NOTED			2000	2000	Y
103	THE GEOGRAPHIC COORDINATE DETERMINATION METHOD BASED ON GPS			10	1	Y
104	THE GEOGRAPHIC COORDINATE DETERMINATION METHOD BASED ON INTERPOLATION			300	1	
105	THE GEOGRAPHIC COORDINATE DETERMINATION METHOD BASED ON LORAN C			500	185	Y

Collect Method Code	Collect Method Description	Scale Lowerbound	Scale Upperbound	Default Accuracy	Maximum Accuracy	WGS84/ NAD83 Default Datum
106	THE GEOGRAPHIC COORDINATE DETERMINATION METHOD BASED ON PUBLIC LAND SURVEY			2000	15	
107	THE GEOGRAPHIC COORDINATE DETERMINATION METHOD BASED ON CLASSICAL SURVEYING TECHNIQUES			1	0	Y
108	THE GEOGRAPHIC COORDINATE DETERMINATION METHOD BASED ON ZIP CODE			17400	17400	
109	THE INFORMATION IS NOT KNOWN			17400	17400	
030	INTERPOLATION - DIGITAL MAP SRCE (TIGER)	0	1200	1	1	Y
030	INTERPOLATION - DIGITAL MAP SRCE (TIGER)	1201	24000	30	15	Y
030	INTERPOLATION - DIGITAL MAP SRCE (TIGER)	24001	100000	100	50	Y
018	INTERPOLATION-MAP	0	1200	1	1	
018	INTERPOLATION-MAP	1201	20000	15	15	
018	INTERPOLATION-MAP	20001	100000	100	50	
019	INTERPOLATION-PHOTO	0	20000	15	5	Y

C.2 APPENDIX B MAXIMUM NAD27 TO NAD83 DATUM SHIFT BY STATE

OBJECT NAME	Object Type
AK	214.91
AL	22.46
AR	23.1
AS	109.24
AZ	74.59
CA	100.9
CO	58.14
CT	42.24

OBJECT NAME	Object Type
DC	29.17
DE	34.93
FL	50.88
GA	31.44
GU	149.19
HI	466.15
IA	25.96
ID	80.87
IL	13.89
IN	8.52
KS	39.63
KY	17.88
LA	29.08
MA	47.3
MD	34.97
ME	46.47
MI	14.63
MN	27.69
MO	22.64
MS	22.47
MT	74.95
NC	41.65
ND	35.68
NE	42.37
NH	42.32
NJ	38.38
NM	58.61
NV	89.4
NY	42.48
OH	19.1
OK	41.44
OR	102.04
PA	35.81
PR	225.12
RI	44.17

OBJECT NAME	Object Type
SC	30.92
SD	38.81
TN	19.91
TX	52.99
UT	71.73
VA	35.24
VI	224.59
VT	39.31
WA	102.97
WI	16.85
WV	27.01
WY	63.2