Facility Registry System (FRS)

Facility Best Practices

NOVEMBER, 2012

Prepared for: ENVIRONMENTAL PROTECTION AGENCY

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

TABLE OF CONTENTS

REVISION HISTORY

Version Number	Date	Description of Changes
1.0	11/29/2012	Initial creation.

A. General Facility Conventions and Best Practices

A.1 Facility naming Conventions

- Wherever possible, provide actual facility name, as opposed to owner/operator name
    - Corporations often own facilities, but facilities are not, themselves, corporations. Leave out corporate designations and abbreviations such as 'Corporation', 'Inc' or 'LLP' from the facility name field. These are appropriate in a field that captures owner or operator name, but are not appropriate for the name of the physical object of environmental interest. Nearly every non-public sector facility is owned by some sort of corporation. The word, its variants and abbreviations just take up valuable space in a facility name field (as illustrated in the example above).

- We usually don't need to have the city name in the facility name field unless it's a municipal facility. For example "Haverhill Wastewater Treatment Facility"
- Names should be as descriptive as possible. Descriptive naming example: "Harvard Blackstone Steam Plant" instead of "The President & Fellows of Harvard College"
- Provide adequate detail on what the facility actually is and leave out what it is not
    - For example "Colstrip" by itself as a facility name is not adequately descriptive, and could refer to Colstrip Airport (as reported to EIS) or Colstrip Steam Electric Plant (as reported to E-GGRT) both located in Colstrip, MT.

- Common descriptive terms
- *Need consistent approach on how to deal with WWTP/STP, et cetera (discussion on how to handle various types of facilities - should these be done in concert with some type of parsing on the database backend?)*
- Examples of improvable facility ID data (FRS improved data in bold):

10024355945	**CABOT INKJET COLORANTS**	**50 ROGERS RD**	**HAVERHILL**	**ESSEX**	**MA**
(MA-EPICS-373274)	CABOT CORPORATION	50 ROGERS RD	HAVERHILL	ESSEX	MA
(TSCA-200000727)	CABOT CORPORATION, HAVERHILL MANUFACTURI	50 ROGERS ROAD	HAVERHILL, ESSEX		MA
(TSCA-200000728)	CABOT CORPORATION - INKJET COLORANTS DIV	50 RODGERS ROAD	HAVERHILL, ESSEX		MA
(RCRAINFO-MAC300010444)	CABOT CORPORATION	50 ROGERS RD	HAVERHILL	ESSEX	MA
(MA-EPICS-377196)	CABOT CORPORATION	50 ROGERS RD	HAVERHILL	ESSEX	MA

A.2 FACILITY LOCATION

- Physical location should actually be where the facility is actually located, not the responsible office location, if that is different. Additional locations can be stored in mailing address fields designed for that purpose.
    - Example: POTW location should NOT be listed as at the municipal town hall, except in the extremely unlikely event that it is actually there.
- Address should be an actual street address, never a PO Box. A small and decreasing, but persistent number of facilities still contain PO Box information in the physical address field. A problem with PO Boxes is that geocoding algorithms cannot locate them to anywhere other than the ZIP centroid. Addresses should follow FGDC/URISA/NENA United States Thoroughfare, Landmark, and Postal Address Standard - http://www.urisa.org/about/initiatives/addressstandard
- When looking up or validating street addresses, consider that EPA uses the NAVTEQ streets database, which is consistent with Bing, Yahoo, MapQuest and Nokia maps - Google may provide slightly different coverage results, address formatting and naming conventions, as it uses TeleAtlas as its' backend streets data provider. Google provides excellent resources for searching for businesses, as well as additional resources for visual verification via aerial imagery and its' StreetView capability, but if there are inconsistencies in street naming, it is suggested that the NAVTEQ-based street naming be used for consistency.
- Tools such as Google StreetView and Bing Maps birdseye view may sometimes provide visual verification of facility locations via street-level and oblique aerial photos
- Failing the ability to provide an actual street address:
    - lat/long values should be provided - using EPA Lat Long standard. The Lat/Long data standard lays out how coordinates should be formatted, i.e. latitude preceding longitude, provide values with decimal degrees preceded by +/- for direction indicator, e.g. northern or southern hemisphere. http://iaspub.epa.gov/sor_internet/registry/datastds/findadatastandard/epaapproved/latitudelongitude/LatLongStandard_08112006.pdf
    - If a coordinate is provided, unless otherwise specified, it will be assumed that the point was collected via most common method available – that being GPS Coordinate, WGS84 datum as is available on most mobile devices and recreational GPS receivers.
    - (For discussion) Failing address and lat/long, is there some hierarchy of precedence that can provide contextually accurate locations?
    - PLSS Township/Range/Section - how to format for parseability and what minimum elements are needed (e.g. Meridian where applicable), and use in BLM Geocommunicator geocoder
    - How to handle things like "14 miles northeast of Gunstock WY"
    - How to handle areas with unique addressing problems, like Puerto Rico
    - Other cases?
- Portable sources (e.g. portable asphalt plants)
- How to deal with facilities or permitted systems with a broad geographic scope, e.g. municipal sewer system?
- Others?

A.3 FACILITY SITE IDENTIFICATION STANDARD

- Facility Site Identification Standard – Relevant sections providing guidance for stewards:
  - Facility Site Identification Standard Reference: http://iaspub.epa.gov/sor_internet/registry/datastds/findadatastandard/epaapproved/facilitysiteidentification/FacilitySite_01062006.pdf

A.4 FEDERAL FACILITIES

- Reference: FGDC Facilities Workgroup - http://www.fgdc.gov/standards/projects/FGDC-standards-projects/FederalBldgsFacilities

A.5 TRIBAL

FRS currently maintains two tribal indicators which are aggregated from various programs:

- Tribal Flag (Yes/No value)
- Tribal Reservation Land

These are currently not cross-checked against other available sources. The FRS team is considering using the IND-3 Tribal Boundary as a QA point to flag potential anomalies and discrepancies between what was reported and the spatial extent in IND-3.

- Tribal Data Standard Reference: http://iaspub.epa.gov/sor_internet/registry/datastds/findadatastandard/epaapproved/tribalidentifier/

A.6 INDUSTRIAL CLASSIFICATION

- If translating from SIC to NAICS or vice versa, particular care must be taken, due to differences in granularity.
- Industrial Classification Standard: http://iaspub.epa.gov/sor_internet/registry/datastds/findadatastandard/epaapproved/sicnaics/SICNAICS_01062006.pdf

A.7 CONTACT INFORMATION

- Contact Information Standard - maybe provide insights on how to deal with incorrect POCs, via Error Tracker et cetera?
- http://iaspub.epa.gov/sor_internet/registry/datastds/findadatastandard/epaapproved/contactinformation/ContactInfo_01062006.pdf

A.8 OTHERS?

- Any other applicable standards, guidance and best practices?
- EPA Approved Standards: http://iaspub.epa.gov/sor_internet/registry/datastds/findadatastandard/epaapproved/

B. BEST PRACTICES FOR FACILITY RECONCILIATION USING FLA

B.1 RESOLVE POSSIBLE DUPLICATE FACILITIES

The "Resolve Possible Duplicate Facilities" option allows data stewards to select FRS records that have been created by the automated process and marked as possible duplicates. Users first search for records in their home state or region and then select facilities, one at a time, from the query results. Once a facility is selected, FRS builds a candidate list of possible matches (i.e., facilities with similar name and addresses), ordered by degree of similarity (with the highest first). The user reviews the candidate list and checks any facility or facilities that are duplicates. The user must then decide which FRS record to retain. All linkages under the duplicate records are re-linked under the retained record.

This section should describe:

- how to create the FLA possible duplicates report
- steps to resolve a duplicate using FLA (show 2-3 examples, including at least one that merges the facilities and one that does not)
- how to decide if the records should be merged
- how to decide which record should be retained
- how to choose/enter comments

Explain how to cross-check the existing facility identification data (that is, affiliated organizations and contacts, alternative names, alternative IDs, mailing addresses, SIC Codes, NAICS Codes, and geospatial data) from within the FRS application to aid in decision making. Additionally, explain how to cross-check the wealth of environmental information available through Envirofacts from within the FRS application. Mention using TRI reported releases and expected links to air, water, or RCRA facility records as an example and other external links (e.g., State systems, ECHO, My Environment).

B.2 IMPROVE DATA QUALITY

The "Improve Data Quality" option allows users to select FRS records that have been assigned a data quality code that indicates missing or erroneous information. Users first search for records in their home state or region and then select facilities from the query results to update or improve

This section should describe:

- how to create the FLA report of records that have data quality problems
- how to research a facility using web resources to verify and obtain information
- helpful web resources (see table 1)
- how and when to use the "Unlist Option"

Commonly Used Facility Research Web Sites By Category

Business Search/Address

http://smallbusiness.dnb.com (Dun & Bradstreet)

http://www.melissadata.com/Lookups/index.htm (Melissa Data)

http://www.google.com (Google)

http://www.superpages.com/ (SuperPages Lookup)

http://www.switchboard.com/ (Switchboard Lookup)

http://www.usps.com/ (USPS Lookup)

Additional ZIP Code Listings

http://www.zipinfo.com/search/zipcode.htm

http://www.mongabay.com/igapo/zip_codes/

Additional County Listings

http://www.census.gov/geo/www/gazetteer/gazetteer2010.html (2010 Census)

http://www.csac.counties.org/default.asp (California Counties)

Latitude / Longitude

http://geocoder.us/

http://www.maporama.com/share/

http://www.ajmsoft.com/ac/geocode.php

http://www.zipmgr.com/geocodeo.aspx

http://www.geocode.com/modules.php?name=TestDrive_Eagle

http://worldatlas.com/aatlas/imageg.htm

http://www.hmssurprise.org/Resources/whereami.html

Other

http://www.circlek.com/CircleK/FindAStore.aspx (Circle K Find A Store)

http://www.7-eleven.com/Drinks/Cold/Iced-Coffee/Default.aspx (7-Eleven Find A Store)

http://www.chevron.com/products/locator/locmap_query.asp (Chevron Find A Station)

http://www.texaco.com/findatexaco/ (Texaco Find A Station)

http://ccionline.org/counties_map.htm (Gas Station Locations)

http://www.automotive.com/gas-prices/index.html (Gas Station Locations)

Anytime the quality of an existing FRS record is improved by filling in data gaps and/or correcting values for the core facility site data elements, the new data is processed through the Data Standardization and Parsing and Data Quality Evaluation routines. If the data quality evaluation results in inconsistencies, the user is notified on-line. In addition, to prevent the creation of duplicate facilities, the updated FRS record is also run through the Automated Integration process.

If the automated integration processes result in one or more potential matches, the list of candidate matches is displayed on the screen. The data steward has the option of either linking to one of the existing candidates, proceeding with updating the existing FRS record, or creating a new FRS record. If there are no duplicate facility listings, the data steward continues to process the facility record. The record is finished processing once a screen appears stating that the update was successful.

B.3 FACILITIES FLAGGED FOR MANUAL REVIEW

The "Facilities Flagged for Manual Review" option allows users to select FRS records that have updated program system records that do not match the FRS information. Users first search for records in their home state or region and then select facilities, one at a time, from the query results. The Facility Detail Screen displays the specific reason the facility was flagged. Users must then determine the course of action to take to resolve the discrepancy.

This section should describe:

- how to create the FLA report of updated program records that do not match their parent FRS records
- how to determine if the FRS record should be updated with the new program data
- how to update the FRS record or clear the flag

B.4 TIPS ON HOW TO SEARCH FOR FACILITIES

Describe expected results with the following types of Searches:

- Registry ID, Program ID
- Name
- ZIP Code
- City, County
- System
- Multiple criteria

B.5 BEST PRACTICES FOR SITE LOCATOR TOOL

When working with the Site Locator Tool, the Bing Maps aerial photography can provide the best perspective on facility location, for example allowing the steward to distinguish between different buildings which may be on a site. The Bing Maps Birdseye view can also provide additional perspective, by showing an oblique view of the facility, as well as allowing the user to rotate around the facility to view it from different perspectives. Note however that point features overlaid on the Bing Maps birdseye view may appear offset relative to the feature, and after ascertaining the correct location, the user should ideally shift back to the orthophoto view for adjusting and updating facility locations. The Site Locator Tool calculates an accuracy estimate calculated from the aerial photo view. The accuracy estimation routine calculates a value based on RMS distance in pixels As such, more accurate results will be obtained by zooming in on the aerial photo.

Retrieved from "https://wiki.epa.gov/frs/index.php/Facility_Best_Practices"