



Facility Registry Service (FRS)

Automated and Manual Facility Merging

UPDATED DECEMBER, 2016

Contents

A.1 Purpose & Overview.....**Error! Bookmark not defined.**

A.1 PURPOSE & OVERVIEW

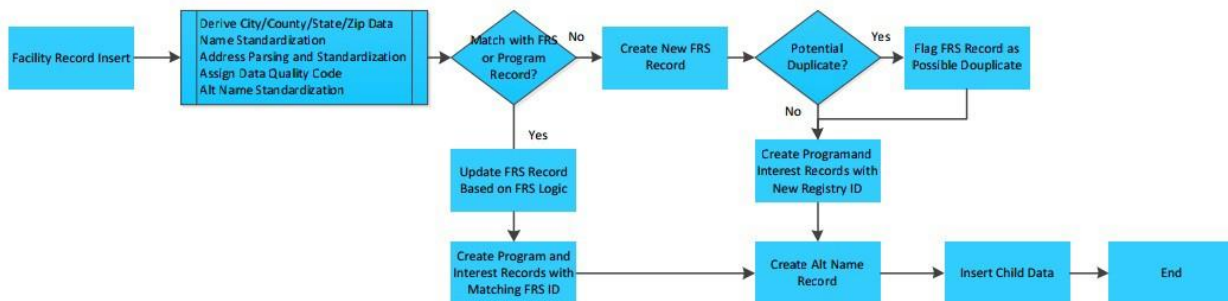
The purpose of this document is to describe the automated and the manual merging process utilized by Data Stewards in the Facility Linkage Application.

Program records in the FRS database are associated together automatically during integration and refreshes of Program/State data into the FRS Database. The FRS database matches the program records that have varying data quality in order attempt to identify unique facilities which are utilized by other programs to allow for searching for unique facilities using a single standardized address with improved data quality.

As FRS receives data flows from its partners, there are times when the incoming facility is already listed in the FRS database. In this case, if the facility is an exact match or meets a certain matching criteria threshold, it is automatically linked to an existing facility. (Automatic Merge) If the facility is deemed as a potential duplicate, but not close enough to automatically match, a flag is raised, it is then referred to the data stewards for further analysis and resolution (Manual Merge). This merging capability and function reduces redundancy within the system, making the service more concise, and reduces burden of data management in the long run.

A.2 AUTOMATED MERGING

The automated merging is done by PL/SQL logic in the Oracle database. When a new record is inserted into FRS it follows the logic shown in the flowchart below:



In the matching process, the incoming records are first standardized, parsed, and evaluated for data quality. These processes makes the data uniform and facilitates in the matching process, as it makes comparing the records easier and faster. FRS performs dynamic searches of the existing database using name and address matching algorithms, which results in a score, as shown below. If the record is matched with a FRS or program record with a score of 85 or above, the incoming record is linked to the FRS record and then goes through the update process. If the matching score for the incoming record is 84 or below, a new FRS record is created. If the score is between 35 and 84, FRS flags the record as a potential duplicate for a steward to review. If the score is 34 or below, no flags are raised and no linking takes places, it is treated as a completely new record in FRS.

The data elements are given scores and the sum of their scores determines whether the facility will be merged or flagged. The scoring is based on the following standardized data elements:

Scoring Table	
Street name and House Number	50
Facility Name	25
City	15
County	5
Zip	5

In the three scenarios below, a match on the data elements shown would add up to a score of 85 and above, in which case the source record is linked with the FRS record.

Merge Combinations		
Scenario 1	Scenario 2	Scenario 3
City	City	County
County	State Code	Zip Code
Zip Code	House Number	State Code
State Code	Street Name	House Number
House Number	Facility Name	Street Name
Street Name		Facility Name
Facility Name		

A.3 MANUAL MERGING

Manual merging of duplicate facilities is done through the Facility Linkage Application (FLA). When the automated process detects close potential matches to facilities, but not exact duplicates, the facilities are flagged as possible duplicates. The possible duplicates then must be

resolved by a data steward to confirm if these facilities are the same through research. If duplicates are present the data steward then merges the duplicate facilities, consolidating the multiple records into one FRS record. If through research it is deemed that the given facility is a new facility, and not a duplicate; then the data steward simply clears the flag. Facilities can also be merged directly by the data steward through FLA, if they are known to be duplicates without being flagged.

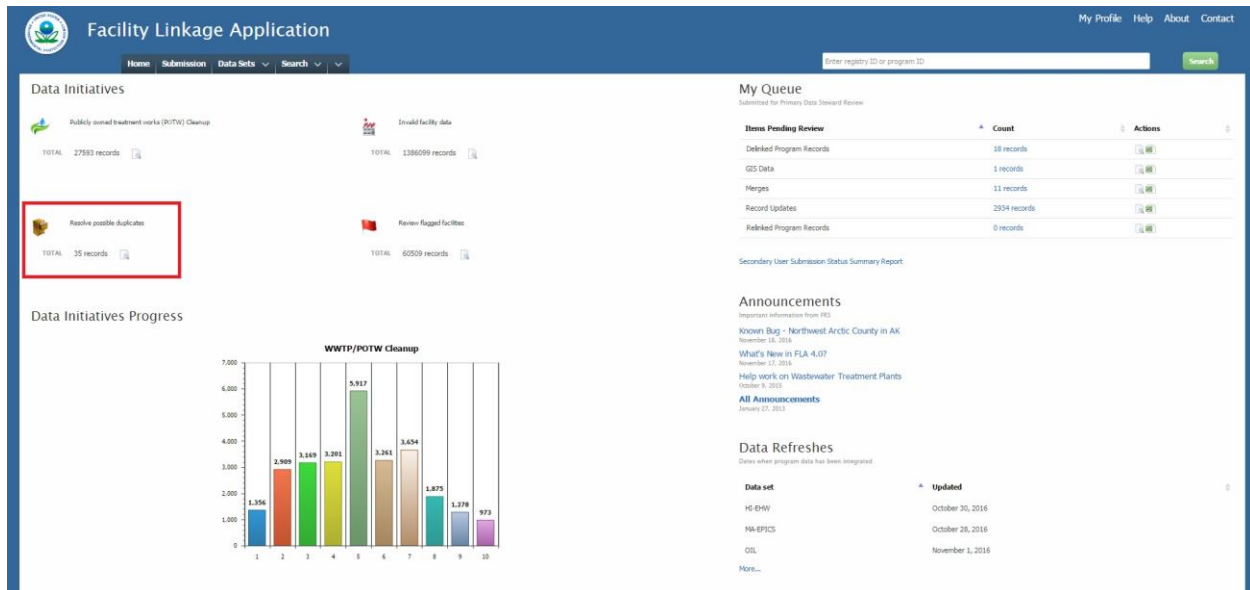
Manual merging is performed by FRS Primary Data Stewards. A primary data steward acts as a point of contact for regional data needs and data quality issues. Assists with the dissemination of information about FRS, FRS data improvement projects, and coordinates with State Data Stewards. The primary data stewards make data corrections in FLA, which are incorporated directly into the FRS production layer, and approve data corrections submitted by secondary or emergency response stewards within their region. Each EPA region is represented by a primary data steward; a list is available from the FRS website or by clicking [here](#). Secondary data stewards consist of staff from regions, program offices, states and trusted partners not identified as primary stewards. Secondary data stewards are delegated responsibility by the primary stewards. These stewards recommend changes that are later approved by their regional primary data steward.

There are multiple ways to identify facilities to merge in FLA

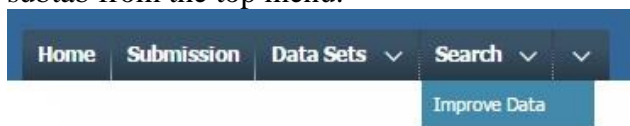
- Via the possible duplicates or resolve duplicates report
- Merging facilities not in the duplicates reports (if known, through other reports, searching FLA, etc.)

There are multiple ways to access facilities flagged as possible duplicates

- 1) Through home page quick access under the “Data Initiatives” at the bottom of the FLA page. The “Review” button highlighted by the redx in the image below will return a report of all possible duplicates for the Data Stewards Region.



2) Another way to access the records flagged as duplicates is to select the Improve Data subtab from the top menu.



From the Improve Data page you can create and refine a query to identify the data you are focusing on. When on this page select the Resolve Duplicates option to have potential duplicates returned using the criteria you specify. Click the Review button to evaluate potential duplicate candidates.

<input type="checkbox"/>	<input type="button" value="Review"/>	110055662607	1-DAY PAINT & BODY #127	1100 COMMERCIAL AVE.	OXNARD	CA	93033	VENTURA	1100 COMMERCIAL AVE, OXNARD, CA 93030 US	V	09
--------------------------	---------------------------------------	--------------	-------------------------	----------------------	--------	----	-------	---------	---	---	----

3) From the Registry Details page you can select the “Find Duplicates” option and a list of potential duplicates will be presented.



After selecting a record

Clicking on the Review button or the Find Duplicates button returns a list of the facilities that have been identified by the application as potential duplicate(s).(see image below) Note that the first record is already checked, because that is the record that is the selected record, all potential matches will appear below it.

Registry Matching Results

Analyze the duplicate candidates and check the checkbox beside any duplicate records.

- Select any registry IDs to merge by checking the check box.
- Additional registry IDs can be added manually by entering a registry ID and selecting the add button.
- Click the Cancel button to not merge any records and return to the previous page.
- Please click the Remove button in the same row as the Registry ID you do not want included in the merge (except the First Row).

Add Registry ID:

Registry ID	Registry Name	Location Address	City	State	Postal Code	County	DQ Code	Alternative Names	Last Update	Remove
<input checked="" type="checkbox"/> 110069458759	POPLAR BLUFF SENIOR HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	MO	63901-3300	BUTLER	Valid	-	16-AUG-30 - FLA	<input type="button" value="x"/>
<input type="checkbox"/> 110022017921	POPLAR BLUFF HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	MO	63901	BUTLER	Valid	POPLAR BLUFF HIGH	16-NOV-08 - ICIS	<input type="button" value="x"/>

row(s) 1 - 2 of 2

Through research, the data steward chooses to either merge facilities or clear the flag. One or more facilities can be merged to selected facility if they are verified as the same facility.

Proper research includes looking at the facility detailed view(s) to look at associated data such as, alternate name(s), interests, SIC, NAICS codes and geospatial data to see if they are related. Search engines such as Google, Bing along with map searches should also be employed in order to ensure that the facilities are the same.

In order to merge the facilities click the box (to add a check) to the facility being merged and then click on the blue “Merge” button. The data steward will be taken to the page shown below, which lists the facilities and its associated interests.

Facility Linkage Application My Profile Help About Contact

Home Submission Data Sets Search

Enter registry ID or program ID

Merge Registries

Identify the record that you would like to keep as the FRS record.

- To include another facility to merge, enter the Registry ID and select the Add button.
- Select the record to retain from the list of records to merge.
- Enter comments.
- To edit the merged facility record, select the Merge and Update button.
- If no duplicates were found for this facility, select the Clear Flag button to remove this facility from the list of Possible Duplicates.

Add Registry ID:

Registry ID	Registry Name	Location Address	City	County	State	Postal Code	Last Update	Data Quality Code
<input checked="" type="radio"/> 110022017921	POPLAR BLUFF HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	BUTLER	MO	63901	16-NOV-08 - ICIS	Valid
<input type="radio"/> (ICIS-600009442)	POPLAR BLUFF HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	BUTLER	MO	63901	16-NOV-08 - ICIS	Valid
<input type="radio"/> 110069458759	POPLAR BLUFF SENIOR HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	BUTLER	MO	63901	16-AUG-30 - FLA	Valid
<input type="radio"/> (RCRANFO-MOR000524173)	POPLAR BLUFF SENIOR HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	BUTLER	MO	63901	16-AUG-30 - FLA	Valid

row(s) 1 - 4 of 4

Comments

Additional Comments:

Here the data steward will be choosing one record and FRS Registry ID to keep, and the environmental interests from the other facility records will be merged into the one that is retained.

After researching, if it is deemed that the facility data needs to be updated, the data steward can choose the “Merge and Update” button. This will allow the data steward to update the chosen facility before merging them.

If a known duplicate is not listed as a possible duplicate.

Another method for merging is manually selecting and merging facilities that are known to be duplicates. An example would be to merge two facilities that the data steward knows for a fact are duplicate, but have not made it to the possible duplicates report or have had the flag incorrectly removed. There are various reasons why a facility would not show up in the possible duplicates report, one example is if the naming is incorrect or address has been wrongly entered – In which case you would update the facility data first. After updating any facility registry information a possible duplicate check is automatically performed.

The on the Merge Registries page follow the step below describing how to merge a facility that is not listed as a “possible duplicate.” Duplicates in the system can also be found and merged with facilities found to be the same while searching FLA.

From the Resolve Possible Duplicates page add in the Registry ID that should be linked with the selected facility and click the add button. The page will refresh containing the record you added. Then complete the merge as described above.

Facility Linkage Application My Profile Help About Contact

Home Submission Data Sets Search

Enter registry ID or program ID Search

Merge Registries

Identify the record that you would like to keep as the FRS record.

- To include another facility to merge, enter the Registry ID and select the Add button.
- Select the record to retain from the list of records to merge.
- Enter comments.
- To edit the merged facility record, select the Merge and Update button.
- If no duplicates were found for this facility, select the Clear Flag button to remove this facility from the list of Possible Duplicates.

Add Registry ID:

Registry ID	Registry Name	Location Address	City	County	State	Postal Code	Last Update	Data Quality Code
110022017921	POPLAR BLUFF HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	BUTLER	MO	63901	16-NOV-08 - ICIS	Valid
(ICIS-60009442)	POPLAR BLUFF HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	BUTLER	MO	63901	16-NOV-08 - ICIS	Valid
110069458759	POPLAR BLUFF SENIOR HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	BUTLER	MO	63901	16-AUG-30 - FLA	Valid
(RCRANFO-MOR000524173)	POPLAR BLUFF SENIOR HIGH SCHOOL	1300 VICTORY LN	POPLAR BLUFF	BUTLER	MO	63901	16-AUG-30 - FLA	Valid

row(s) 1 - 4 of 4

Comments

Additional Comments: