Facility Registry Service Parser and Standardization Processes

Environmental Protection Agency (EPA)
Facility Registry Service (FRS)

Delivery Order # GS00Q09BGD0022
Task Order # 47QFCA-18-F-0009
Version 1.0
July 18, 2022

Table of Contents

List of Exhibits

Revision Log

Date	Version No.	Description	Author	Reviewer	Review Date
04/01/20	0.01	Draft version	K. Hoke	C.P. Urbano J. Chilton K. Sanagapaty	04/01/20
07/18/22	0.02	Delivery of Final Facility Registry Service Parser and Standardization Processes	N. Spaeth	K. Hoke C.P. Urbano	05/19/22

1 Introduction

Facility data is at the core of federal, and state, local, and tribal (SLT) environmental regulatory processes. Knowing a facility's name, ownership, location, and characteristics are key to a comprehensive picture of past, current, future, and potential environmental impacts. Linked to other critical environmental data such as ambient air and water quality data, census figures, and other demographic information, facility data has the capacity to provide a comprehensive picture that can enable co-regulators to make more informed decisions so they can better protect human health and the environment.

The Facility Registry Service (FRS) is Environmental Protection Agency's (EPA's) source for integrated facility information, managing integrated data for more than 5.2 million facilities. FRS provides information for public viewing via queries and in other EPA platforms such as Envirofacts, Cleanups in My Community (CIMC), and Enforcement and Compliance History Online (ECHO).

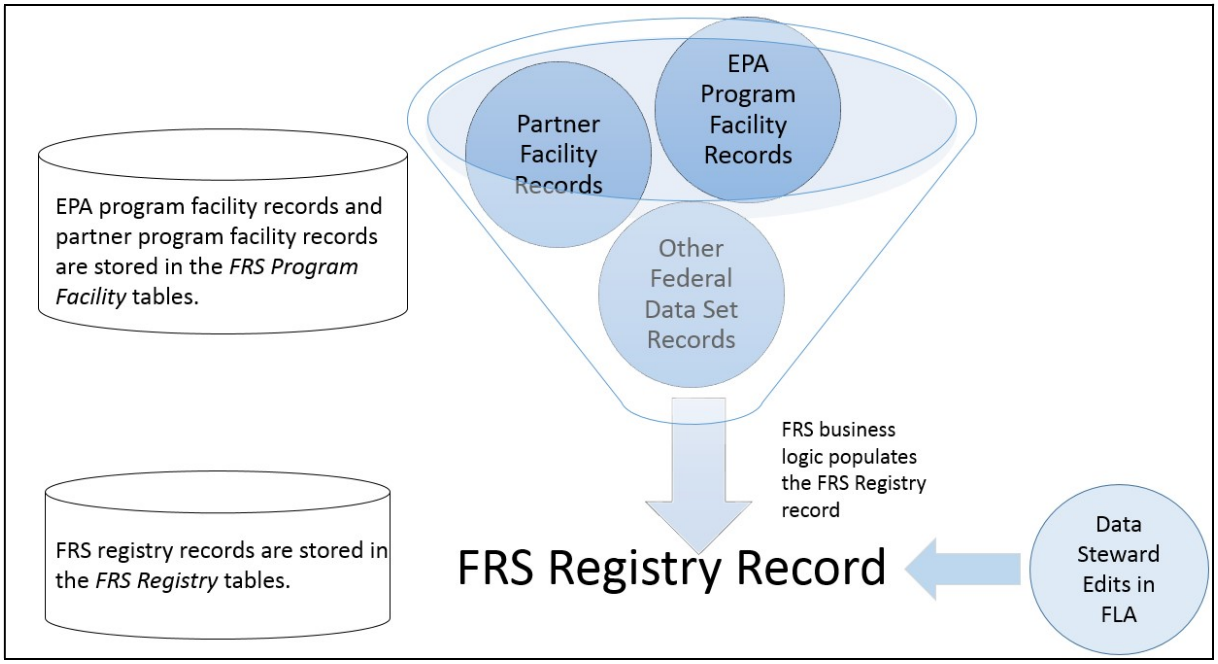
FRS links facility, geospatial, and reference data across EPA programs, states, tribes, and other federal agencies based on a consistent set of data elements such as address, facility name, and location. The main goals of FRS are to integrate data, improve data quality, and provide data to EPA programs, partners, and the public. FRS allows users to identify the permits or regulations that apply to a particular facility and the location of a regulated facility within a specific sector. Program management, enforcement, rulemaking, reporting, analysis, and emergency response activities use FRS and its data.

By integrating disparate sources of facility information together, FRS provides critical comprehensive information about a facility's regulatory processes and environmental impact. FRS also enhances facility data by applying data validation, standardization, and mapping capabilities, further enabling the use of this critical data.

1.1 FRS Registry and Program Records

FRS links facility data from a variety of sources into a single record, called the FRS registry record. The FRS registry record is linked to one or more program records provided by EPA program systems and partners (states, tribes, and local agencies). Each program record contains its own attributes for facility name, address, and location coordinates, among others. FRS uses a set of processes to determine how to populate the FRS registry record. FRS also has a tool, the Facility Linkage Application (FLA), which allows FRS data stewards to make updates to a FRS registry record. **Exhibit 1-1** illustrates the creation of FRS registry records from EPA program facility records, partner facility records, as well as from edits within the FLA.

Exhibit 1-1: Development of FRS Registry Records from EPA Program Facility Records, Partner Facility Records, and FLA Use



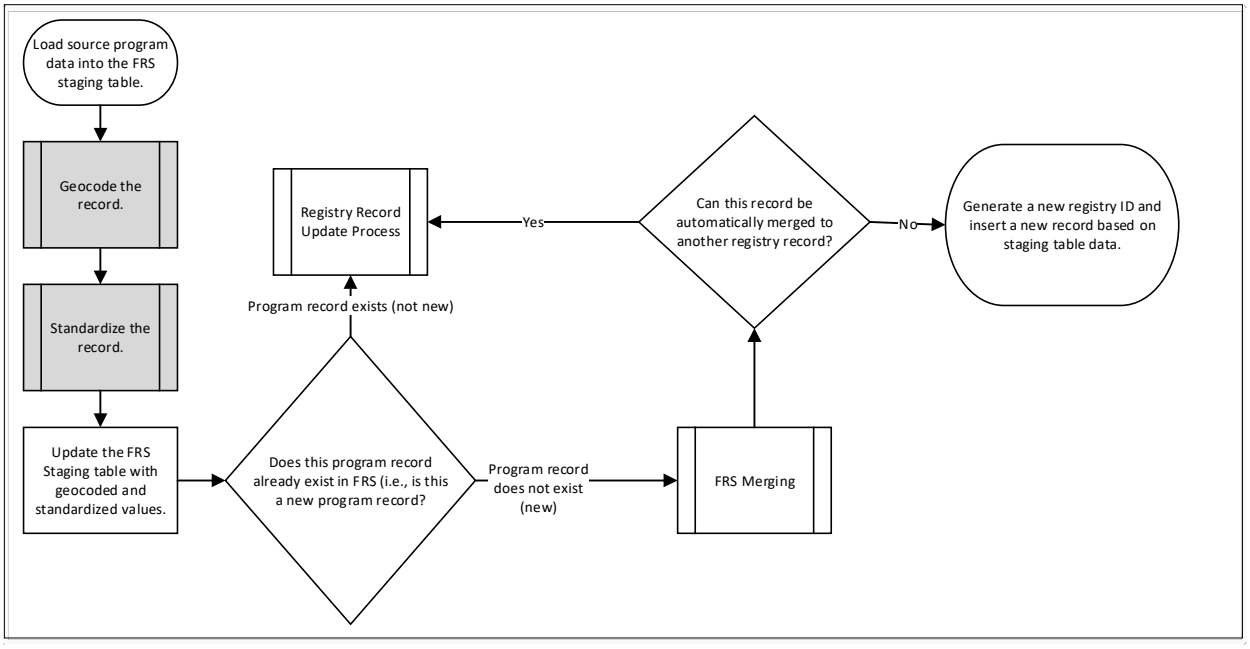
1.2 Document Purpose

As **Exhibit 1-1** illustrates, a FRS record is composed of individual program records that together build a registry record. The registry record represents a summary view of the facility and includes standardized data and data from the related program records. FRS populates the registry record by standardizing and integrating program facility records and by applying business logic, geocoding, and data validation tools to incoming data. **Exhibit 1-2** illustrates an overview of the process used to populate FRS registry records.

The first two steps in this process are geocoding and standardization. The purpose of this document is to describe the geocoding and standardization processes used as part of the core FRS business logic.

Exhibit 1-2: Overview of the FRS Registry Record Population Process

The shaded boxes refer to processes discussed in this document.



2 Process Overview

FRS uses geocoding, parsing, and standardization to more accurately identify which records should be linked together under the same registry record. The goal of these processes is, among other things, to account for differences in how different source systems populate their data. FRS first attempts to geocode the data and derive a standardized address using the geocoder software. If geocoding is not successful, FRS runs a customized parsing and standardization routine.

These processes result in a standardized address and, if geocoding is successful, a geocoded address. The geocoded/standardized address is stored in separate fields than the original program record data. FRS does not change the original data provided by the source.

These procedures are critical to FRS because they enable FRS logic to match data values accurately to link incoming program records together and to existing registry records. The linking and merging process is fundamental to FRS and is necessary to prevent duplicate record creation.

The FRS data tables used for the parsing and standardization routine are listed below in **Exhibit 2-1**.

Exhibit 2-1: FRS Tables Used During the Parsing and Standardization Process

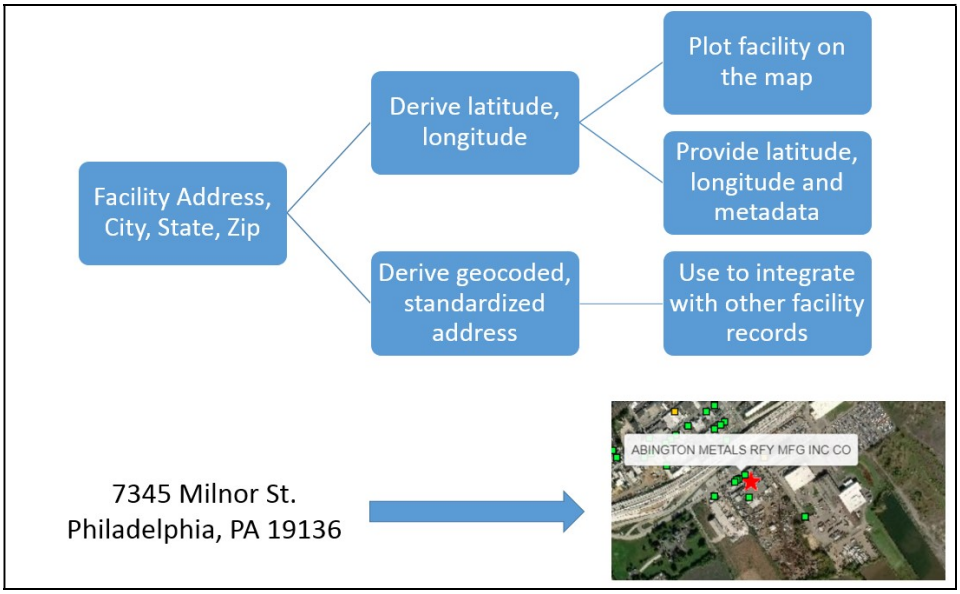
Table Name	Definition
FRS_RES_ALT_NAME	Table used for standardizing alternative names
FRS_STATE_REF	Reference table mapping state code, state name and EPA region codes
FRS_RES_REGISTRY	Staging table to load geocoding and standardization values of facility name and address
FRS_STD_CITY_CNTY	Table used in the standardization process to replace names with suggested abbreviations
FRS_ZIP_CODE_REF	Reference table mapping state code, county code, state name, county name, city name, zip code, and region codes
FRS_DQ_REF	Reference table mapping state code, state Federal Information Processing Standard (FIPS) code, county name, county FIPS code, city name, and postal code
FRS_COUNTRY_REF	Reference table mapping country code, country name, and country FIPS code
FRS_STD_ADDRESS	Table used in the standardization process to replace address names with suggested abbreviations

2.1 Geocoding

Geocoding is a process used to translate a physical address, such as a street address, to a set of coordinates (latitude and longitude values). FRS attempts to geocode addresses after source program data is loaded into FRS.

Geocoding uses a piece of software (geocoder) that evaluates physical addresses and determines the appropriate latitude and longitude values that correspond to that address. The ability to translate a street address into a point that can be plotted on a map is especially important when location (latitude/longitude) data was not widely available for most data sets. **Exhibit 2-2** illustrates the geocoding process.

Exhibit 2-2: Schematic of FRS Geocoding Process



The geocoder standardizes FRS program facility spatial coordinates to the United States Standard North American Datum of 1983 (NAD83) horizontal datum using Oracle Spatial geocoder routines along with HERE Navteq Point Address data. HERE is a location data and technology platform; the HERE data is refreshed quarterly. Historically, the FRS team has updated this dataset, within FRS, annually. The most recent update was in October 2018. The update begins with National Computer Center (NCC) loading the latest HERE data into a local database. FRS then pulls the latest dataset and re-geocodes all of the data in FRS.

Using the Oracle Spatial geocoder, all new FRS facilities, and all existing FRS facilities for which any part of the address has been modified or edited, are geocoded. The Oracle Spatial geocoder uses proprietary code for this process and follows these general steps:

1) Before the geocoder is used, 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," and "BORO" with "BOROUGH OF." Please note that step 1 and step 2 (below) reference the general standardization of parameter data before the data goes through the geocoder for further processing. The updates made in this step are not used elsewhere and may not be saved; it is used just for initial processing.
2) Identify each crossing street of intersections within the FRS location address.
3) Geocode a formatted address using the Oracle Spatial geocoder.
4) Validate the returned geocode object from the geocoder. That validation follows the steps described below:
    a) Reject returned geocodes that do not meet FRS accuracy standards and accept only reference point (e.g., rooftop), house number, street, intersection, and landmark place name geocodes. Geocoded data is given an accuracy score based on a reference point such as a house number, street, etc. The more specific the reference point, the better. For example, a house number reference point is better than a street reference point. If the

geocoded address does not receive an accuracy score, then it is rejected. The cause of this is usually incomplete or missing data.

b) For reference point and house number geocodes, validate that the returned house number matches the supplied house number. If not, convert the geocode to the "Street" geocode.
c) Using the Jaro-Winkler and Levenshtein Distance algorithms, validate the returned street name/place name against the supplied street name/place name. If they do not match, reject the geocode.
d) For street level geocodes, identify the number of blocks for the street and the length of the street. If the street is longer than two (2) kilometers, and if the street contains more than (two) 2 blocks, reject the geocode.
e) Validate the state returned from the geocoder. If it does not match the supplied state, reject the geocode.
f) Validate the Zone Improvement Plan (ZIP) code, county, and city returned from the geocode together using the United States Postal Service (USPS) dataset. This dataset is updated monthly. In certain conditions, Oracle will return the county name instead of the city name for rural areas. If the city, county, and ZIP code cannot be validated, reject the geocode.

Not all addresses can be geocoded. Some program records do not have a complete address or have values of "unknown" for the address. In other cases, the geocoder does not recognize the address in the record. During geocoding, addresses are evaluated, and if they can be geocoded, they are given a score to indicate the level of confidence in the geocoded location. The score is stored within FRS in the 'accuracy_value' column of the database. The lower the accuracy value, the better the score. These levels of confidence and their corresponding accuracy values, in decreasing order, are:

1) Point (e.g., rooftop, center of facility) – 30 meters
2) Street House Number – 150 meters
3) Intersection – 200 meters
4) Street of only one 1 block – 500 meters
5) Street of only (two) 2 blocks or less than two (2) kilometers in length – 2,000 meters
6) Place name – 4,500 meters

If geocoding was successful, the record now has a geocode (latitude and longitude value) and a standardized address. FRS uses the standardized address as part of the linking and merging processes.

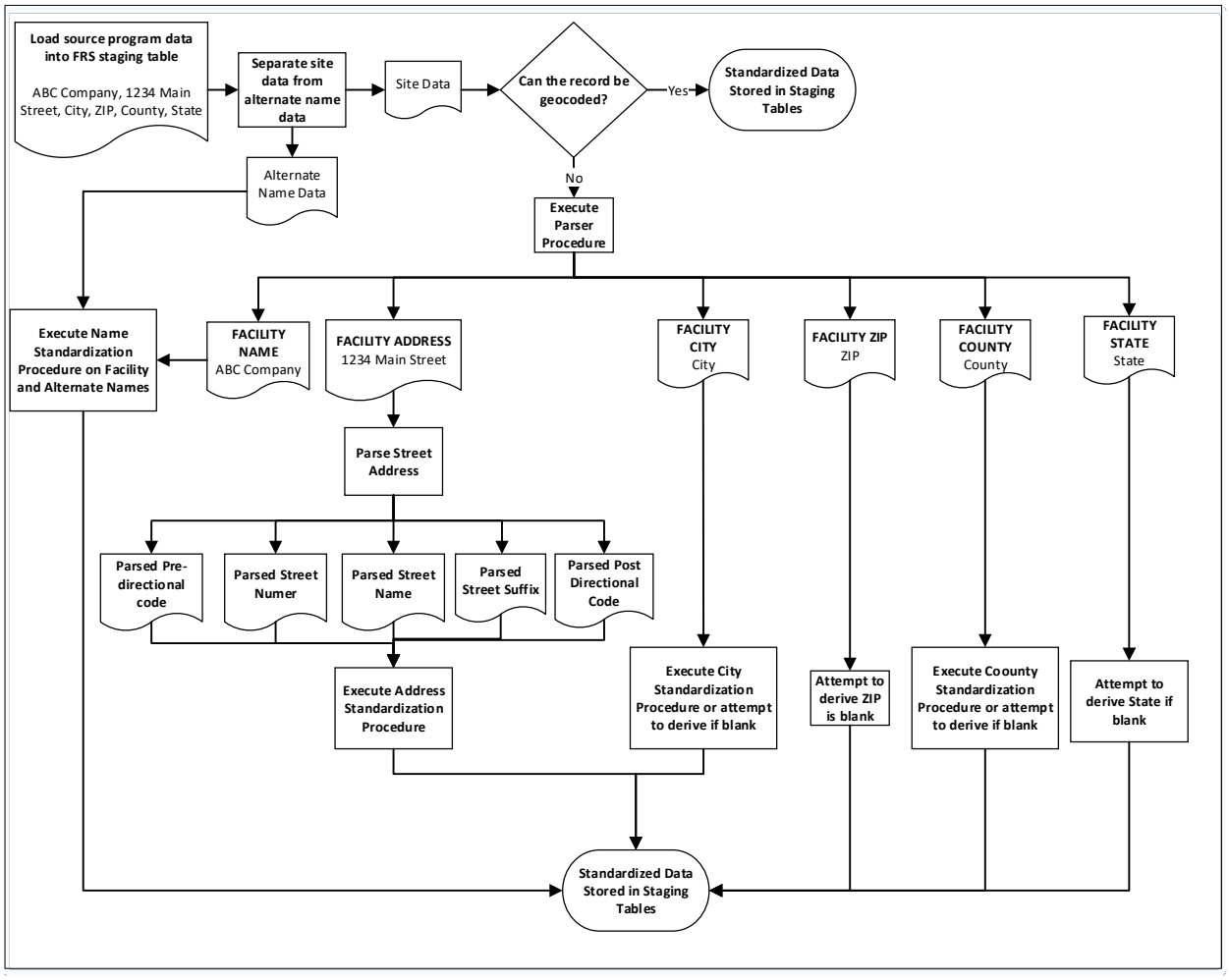
2.2 Parser and Standardization

If the geocoding fails, FRS executes parsing and standardization procedures. Geocoding fails on about forty-five percent of the data, as of 2022. The purpose of these procedures are to:

- Break up data values into distinct parts and the smallest possible data units.
- Derive missing data using the reference data noted in the exhibits within this document.
- Standardize characters where possible.

Exhibit 2-3 illustrates an overview of the standardization and parsing processes.

Exhibit 2-3: Overview of Parsing and Standardization Processes



Once source data is stored in FRS staging tables, FRS separates the site data from alternative name data. The alternative name data is later standardized along with the site name data. FRS attempts to geocode the record as described in **Section 2.1**. If geocoding is successful, FRS stores the geocoded and standardized data in the staging tables.

If geocoding is not successful, FRS parses the data into name and individual address fields (e.g., street number, street name, and street suffix) and attempts to standardize the data. FRS also attempts to derive some data that may be missing from the source record. FRS stores standardized data in FRS staging tables for use in the linking and merging processes. The following sections describe these processes.

3 Name Standardization

FRS standardizes the facility name (also called the primary name), and any alternate names, by calling the name standardization procedure. The name standardization procedure looks for special characters in the name fields and replaces them in the steps described below.

1. The standardization routine replaces special characters as follows:
   a. '#' is replaced with "NUMBER"
   b. '$' is replaced with "DOLLAR"
   c. '&' is replaced with "AND"
   d. All of the following special characters are removed and are not replaced with anything: '-,' '_,' '.,' '@,' '/,' '\,' ',' '(,' '),' ';,' ':,' '%,' '*,' '+,' '',' '",' '{,' '},' '[,' ']'
2. The standardization routine uses a FRS reference table to replace full names with abbreviated names. A copy of this reference table, as of November 2021, is in **Appendix A**.
3. The standardization routine runs additional logic on the abbreviated names and replaces those values with those indicated in **Appendix B**. If the last word in the primary name field is 'AND,' it is removed from the standard name field.
4. Once completed, FRS stores the results in the 'STD_PRIMARY_NAME' or 'STD_ALT_NAME' fields in the database.

3.1 Alternate Name Standardization

The alternative name standardization procedure takes in one parameter, and is used to standardize the alternative name field. The result is then assigned to another variable, called 'STD_ALT_NAME,' or standardized alternative name.

4 Address Standardization

4.1 Address Type

FRS assigns each address an address type. FRS uses address types to categorize addresses so specific logic can execute on specific address types rather than all logic running on all records. This speeds processing. FRS address types are:

- Direction
- Regular Urban
- Hwy
- PO Box
- Irregular
- No Address

FRS can assign multiple address types to a single address such as 'Regular Urban, PO Box.'

The FRS address routine executes the following steps:

1. If the address is null (e.g., there is no address for the record), flag the address type as 'No Address.'
2. If the address contains any of the following: 'North,' 'South,' 'East,' 'West,' 'N,' 'S,' 'E,' 'W,' 'To,' 'From,' 'And,' 'MI,' or 'Behind,' then flag the address type as 'Direction.'
3. If the address contains a street number and street name, flag the address type as 'Regular Urban.'
4. If there is no route number in the address, flag it as 'Hwy.'
5. If 'PO Box' is in the address, flag it as 'PO Box.'
6. If the address starts with 'Restricted,' then flag the address as 'Irregular.'
7. Flag anything else as 'Irregular.'

Exhibit 4-1 lists some examples of various address types.

Exhibit 4-1: Example Address Types in FRS

Example Addresses	Address Type
SWSE Sec 29 T04N R68W	Irregular
(null)	No Address
4.2 Mi E of City Hall	Direction
Unknown	Irregular
112 Center St.	Regular Urban
P.O. Box 907	PO Box

FRS also replaces some values if they are found in the parsed street name:

- Replace '-' with a space
- Replace 'South' with 'S'
- Replace 'North' with 'N'

4.2 Address Standardization

The address standardization procedure populates the following standardized address fields by evaluating the provided address and running word pattern procedures:

- STD_LOC_ADDRESS
- STD_PARSED_STREET_NAME
- STD_PARSED_STREET_NUMBER
- STD_PARSED_PRE_DIR_CODE
- STD_PARSED_STREET_SUFFIX
- STD_PARSED_POST_DIR_CODE

The initial phase of this process is to standardize the raw address passed into the address standardization procedure, which looks for any of the following words and symbols listed under the 'Original Value' heading and, if found, is replaced by the word listed under the 'New Data Value' heading – see **Exhibit 4-2**. The idea behind this process is to feed more meaningful address components to the procedures in the next phase.

The address standardization procedure looks for any of the following words and symbols listed under the 'Original Value' heading and, if found, it is replaced by the word listed under the 'New Data Value' heading. Please refer to the table below that shows the hard-coded data values that are in the current address location script. Refer to **Appendix C** for a list of address location standardization values.

Exhibit 4-2: Hard-coded Location Address Part A

Original Value	New Data Value	Original Value	New Data Value
#	*(space)*	TWENTY FOURTH	24TH
RR*(number)*	RR *(number)*	TWENTY FIFTH	25TH
RR	ROUTE	TWENTY SIXTH	26TH
R	ROUTE	TWENTY SEVENTH	27TH
RTE*(number)*	RTE *(number)*	TWENTY EIGHTH	28TH
RT*(number)*	RT *(number)*	TWENTY NINTH	29TH
ROUTE*(number)*	ROUTE *(number)*	THIRTY FIRST	31ST
RTS*(number)*	RTS *(number)*	THIRTY SECOND	32ND
P.R.	PR	THIRTY THIRD	33RD
K.M.	KM	THIRTY FOURTH	34TH
PR-	PR	THIRTY FIFTH	35TH
R.D.	ROAD	THIRTY SIXTH	36TH
RD.	ROAD	THIRTY SEVENTH	37TH
R. D.	ROAD	THIRTY EIGHTH	38TH
junkdot	*(space)*	THIRTY NINTH	39TH

Original Value	New Data Value	Original Value	New Data Value
(Note: a "junkdot" is basically a period. Based on the logic in the code, whenever it finds a period after a non-numeric character (e.g., St.), it replaces the period with a space. (e.g., St).)			
I-*(number)*	INTERSTATE *(number)*	FORTY FIRST	41ST
I*(number)*	INTERSTATE *(number)*	FORTY SECOND	42ND
I *(number)*	INTERSTATE *(number)*	FORTY THIRD	43RD
N/A	NOT APPLICABLE	FORTY FOURTH	44TH
-	*(space)*	FORTY FIFTH	45TH
.	*(space)*	FORTY SIXTH	46TH
_	*(space)*	FORTY SEVENTH	47TH
@	*(space)*	FORTY EIGHTH	48TH
/	*(space)*	FORTY NINTH	49TH
\	*(space)*	FIFTY FIRST	51ST
,	*(space)*	FIFTY SECOND	52ND
&	AND	FIFTY THIRD	53RD
(	*(space)*	FIFTY FOURTH	54TH
)	*(space)*	FIFTY FIFTH	55TH
;	*(space)*	FIFTY SIXTH	56TH
:	*(space)*	FIFTY SEVENTH	57TH
%	*(space)*	FIFTY EIGHTH	58TH
*	*(space)*	FIFTY NINTH	59TH
NO	*(space)*	SIXTY FIRST	61ST
No.	*(space)*	SIXTY SECOND	62ND
$	*(space)*	SIXTY THIRD	63RD
+	*(space)*	SIXTY FOURTH	64TH
"	*(space)*	SIXTY FIFTH	65TH
' '	*(space)*	SIXTY SIXTH	66TH
{	*(space)*	SIXTY SEVENTH	67TH
}	*(space)*	SIXTY EIGHTH	68TH
[	*(space)*	SIXTY NINTH	69TH
]	*(space)*	SEVENTY FIRST	71ST

Original Value	New Data Value	Original Value	New Data Value
`	*(space)*	SEVENTY SECOND	72ND
~	*(space)*	SEVENTY THIRD	73RD
^	*(space)*	SEVENTY FOURTH	74TH
S E	SE	SEVENTY FIFTH	75TH
S W	SW	SEVENTY SIXTH	76TH
N W	NW	SEVENTY SEVENTH	77TH
N E	NE	SEVENTY EIGHTH	78TH
E W	EW	SEVENTY NINTH	79TH
NORTH WEST	NW	EIGHTY FIRST	81ST
EAST WEST	EW	EIGHTY SECOND	82ND
NORTH EAST	NE	EIGHTY THIRD	83RD
SOUTH WEST	SW	EIGHTY FOURTH	84TH
SOUTH EAST	SE	EIGHTY FIFTH	85TH
P O BOX	POBOX	EIGHTY SIXTH	86TH
PO BOX	POBOX	EIGHTY SEVENTH	87TH
POST OFFICE BOX	POBOX	EIGHTY EIGHTH	88TH
- -	-	EIGHTY NINTH	89TH
OUTER	*(space)*	NINETY FIRST	91ST
EXTENDED	*(space)*	NINETY SECOND	92ND
BUSINESS	*(space)*	NINETY THIRD	93RD
OF	FROM	NINETY FOURTH	94TH
TO	FROM	NINETY FIFTH	95TH
OFF	FROM	NINETY SIXTH	96TH
TWENTY FIRST	21ST	NINETY SEVENTH	97TH
TWENTY SECOND	22ND	NINETY EIGHTH	98TH
TWENTY THIRD	23RD	NINETY NINTH	99TH

Other values that are checked for in the address standardization procedure are as follows:

Exhibit 4-3: Hard-Coded Location Address Part B

Original Value	New Data Value	Original Value	New Data Value
(space)	*(space)*	PR ROAD	ROAD
RURAL ROUTE	ROUTE	HCR	ROAD
OLD ROUTE	ROUTE	U.S.	UNITED STATES

Original Value	New Data Value	Original Value	New Data Value
R ROUTE	ROUTE	US.	UNITED STATES
STATE ROUTE	ROUTE	US.	UNITED STATES
ST ROUTE	ROUTE	U S	UNITED STATES
ROUTE NO	ROUTE	UNITED STATE	UNITED STATES
SR	ROUTE	UNITED STATES HWY	HWY
S R	ROUTE	UNITED STATES ROUTE	ROUTE
COUNTY	CNTY	STATE HWY	HWY
CNTY ROUTE	ROUTE	ST. HWY	HWY
CNTY	COUNTY	NATIONAL HWY	HWY
STATE ROAD	ROAD	NATL. HWY	HWY
STATE ROAD*(number)*	ROAD *(number)*	NATL HWY	HWY
STATE ROAD *(number)*	ROAD *(number)*	NATNL. HWY	HWY
STATE	*(space)*	NATNL HWY	HWY
ST ROAD	ROAD	OLD HWY	HWY
R D	ROAD	NEW HWY	HWY
ACCESS ROAD	ROAD	HWY	HWY

Consider the following example: Assuming 'RR4' as the input phrase, the explanation below illustrates how it is converted into a more meaningful format.

'RR4' → 'RR 4' → 'Route 4'

In the above example, the standardization process takes the 'RR (number)' and replaces this value with 'Route 4' found in **Exhibit 4-2**. FRS uses these standardized address fields as part of the matching process in a future data integration step. The FRS standardization procedure executes the following steps:

1. Copy data values from address fields into the standardized location address fields.
2. Execute word pattern procedures on the data in order to populate temporary values for two fields: 'Out Word' and Chopped Word.' See **Exhibit 4-4** for more information about the word pattern procedures.
    - A "chopped" word is parts of a sentence or word that are parsed out for further process comparison. An "out" word is part of the word/sentence that was parsed out.
3. Word pattern procedures continue until both the out words and chopped words are null, because it cycles through the data to pull out specific keywords to identify data such as street name, number, and direction to make matches against the keywords noted below. Out and chopped words get populated when they are both null until there is a proper, standardized address available in the location key.

- The purpose is to eliminate any inconsistencies in the address field. For example, the procedure checks to see if there is a number after “Highway” in an attempt to make the address more complete.

The word pattern procedures look for the following terms (see list below). If any of them are found, it executes the steps explained below.

- HWY
- ROUTE
- POBOX
- ROAD
- KM
- SR
- STE
- CR
- CNTY ROAD
- BLDG
- PR
- INTERSTATE
- UNITED STATES
- UNIT00
- BLK
- LOT
- BOX

If it finds any of these terms, it executes other steps as follows:

1. If it finds HWY, then it will look to see if it also finds ‘COUNTY HWY.’
    - If it does not find HWY or COUNTY HWY (or other key words noted in this document), the word pattern procedure will continue to cycle through to find other patterns/values that match. However, it may not find matches if the location data is incomplete. In this case, the data is not modified and the data will go into the standardized fields as it was received from the source.
2. The new format phrases are fed into the next set of procedures. Please refer to Error! Reference source not found., which shows the hard-coded data values that in the current address location script. Refer to **Appendix C** for a list of address location standardization values.

If the field contains any of these keywords, a word pattern procedure is used to check the word pattern in the field. For example, a value of 2095 COUNTY HWY PARK HWY EAST is entered into the address location field. The value is copied and put into the standardized location address for modification. Then the procedure checks for any of the keywords listed previously. In this case, the keyword HWY is found. Then the procedure runs another check to see whether

COUNTY HWY is entered. Finally, the procedure checks for any of the following words that come after the keyword HWY:

- NORTH
- SOUTH
- EAST
- WEST
- POBOX
- ROUTE
- STE
- BLDG
- UNIT
- INDL
- /BOX
- N
- S
- E
- W
- AND
- HC
- UNITED

If any of these keywords are found, the value repeats the word pattern procedure. Once the word pattern procedure is finished, FRS continues with the first procedure. After the procedures are finished executing, two values are returned. One value stored in the FRS location key table is called the 'out word.' The other value, 'chopped word,' becomes the input to the same word pattern procedure, which is then executed again through the word pattern loop to get truncated even more. This procedure is true for all other keywords.

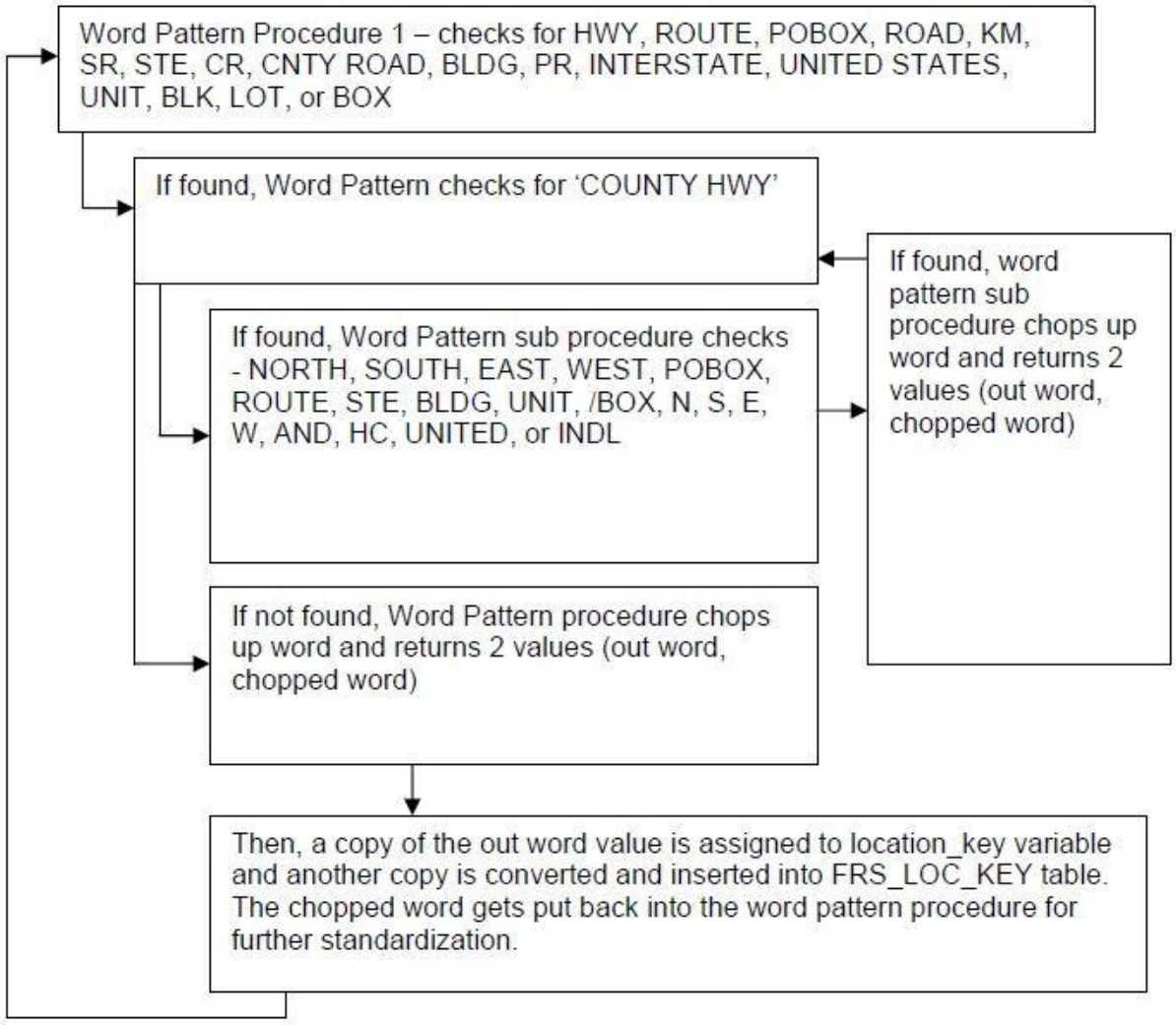
The last chopped word and key are passed into a parser, which splits the total address into different fields, namely street name, street number, pre-direction, and post-direction. A similar process is followed when other keywords are found and subtle code changes happen for different keyword types, but overall, the concept stays the same. The out word and chopped word help obtain patterns so the address can be split into different standardization fields. Using the example above, the word pattern procedure iterations would be as follows:

- Out word: COUNTY HWY PARK
- Chopped word: 2095 HWY EAST
- Out word: HWY PARK
- Chopped word: 2095 HWY EAST
- Out word: (null)

- Chopped word: 2095 EAST

For a visual interpretation of how the procedures work, please refer to the diagram below.

Exhibit 4-4: Word Pattern Procedure for Data that cannot be Geocoded



4.3 City and County Standardization

FRS standardizes city and county names by calling the city and county standardization procedure one by one, using the fields as parameter values. Once the procedure finishes, FRS stores the values in the standard city name and standard county name fields 'STD_CITY_NAME' and 'STD_COUNTY_NAME,' respectively. If the procedure finds any of the following special characters or keywords within the city and county name fields, it will be replaced by the value under the 'New Data Values' heading. Refer to **Appendix D** for a list of city and county name standardization values.

Exhibit 4-5: Hard-coded City and County Name Values

Special Characters and Data Values for City and County Names		Special Characters and Data Values for City and County Names	
Original Values	**New Data Values**	**Original Values**	**New Data Values**
/C/	*(space)*	&	*(space)*
/T/	*(space)*	(	,
/V/	*(space)*	)	,
/O/	*(space)*	;	*(space)*
-	*(space)*	:	*(space)*
_	*(space)*	%	*(space)*
.	*(space)*	*	*(space)*
@	*(space)*	#	*(space)*
/	*(space)*	$	*(space)*
\	*(space)*	+	*(space)*
' '	*(space)*	^	*(space)*
"	*(space)*	*(double space)*	*(space)*
{	*(space)*	CITY OF	*(null)*
}	*(space)*	BOROUGH OF	*(null)*
[	*(space)*	BORO OF	*(null)*
]	*(space)*	TOWNSHIP OF	*(null)*
`	*(space)*	TWNSP OF	*(null)*
~	*(space)*	TWP OF	*(null)*
,	*(space)*		

Next, any commas and entries after the comma are removed as well. The next step in the procedure is to retrieve the word and construct a new primary name; specifically, the program retrieves each word in the field, separates them using quotes, and standardizes them. For example, if the word is 'CO,' then 'Colorado' is appended to the standard city name or standard county name field. Additionally, the standard name from the FRS standard city and county table can be appended to the standard city name or standard county name field for that particular name.

After standardization, the following items, if found, are replaced with corresponding values listed under the 'New Data Values' heading. For example, "Town of Crozet" would be changed to only "Crozet." Please refer to the list below for a complete hard-coded list of these values and their new data values.

Exhibit 4-6: Hard-coded Standardized City and County Names

After Standardization Data Values for City and County Names		After Standardization Data Values for City and County Names	
Original Data Values	**New Data Values**	**Original Data Values**	**New Data Values**
TOWN OF	*(null)*	NORTH EAST OF	*(null)*
TWN OF	*(null)*	SOUTH EAST OF	*(null)*
VILLAGE OF	*(null)*	NORTH OF	*(null)*
SOUTH WEST OF TOWN	*(null)*	SOUTH OF	*(null)*
NORTH WEST OF TOWN	*(null)*	EAST OF	*(null)*
NORTH EAST OF TOWN	*(null)*	WEST OF	*(null)*
SOUTH EAST OF TOWN	*(null)*	VICINITY OF	*(null)*
EAST OF TOWN	*(null)*	IS NEARBY	*(null)*
WEST OF TOWN	*(null)*	BURGH	BURG
NORTH OF TOWN	*(null)*	A F B	*(null)*
SOUTH OF TOWN	*(null)*	AIR FORCE BASE	*(null)*
SOUTH WEST OF	*(null)*	*(space)*	*(space)*
NORTH WEST OF	*(null)*	*(double space)*	*(space)*

Finally, if the final word after standardization is null, then the original value is going to be assigned to standard city name or standard county name. In other words, if a standardized value is not found, the original value is kept.

4.4 State and ZIP Standardization

Missing states and ZIP codes are both derived using the ZIP code reference table. The procedure takes the data values in the address, city, and county columns and uses the ZIP code reference table to match up the missing states and ZIP code data. Once the information is found, the state and ZIP code for the record is stored in the 'FRS RES Registry' table. Otherwise, if there is no match, then the data fields stay empty.

The 'frs_zip_code_ref' reference table is not recreated in this document due to its size; however, a copy of the table is available upon request.

5 Data Quality Standardization

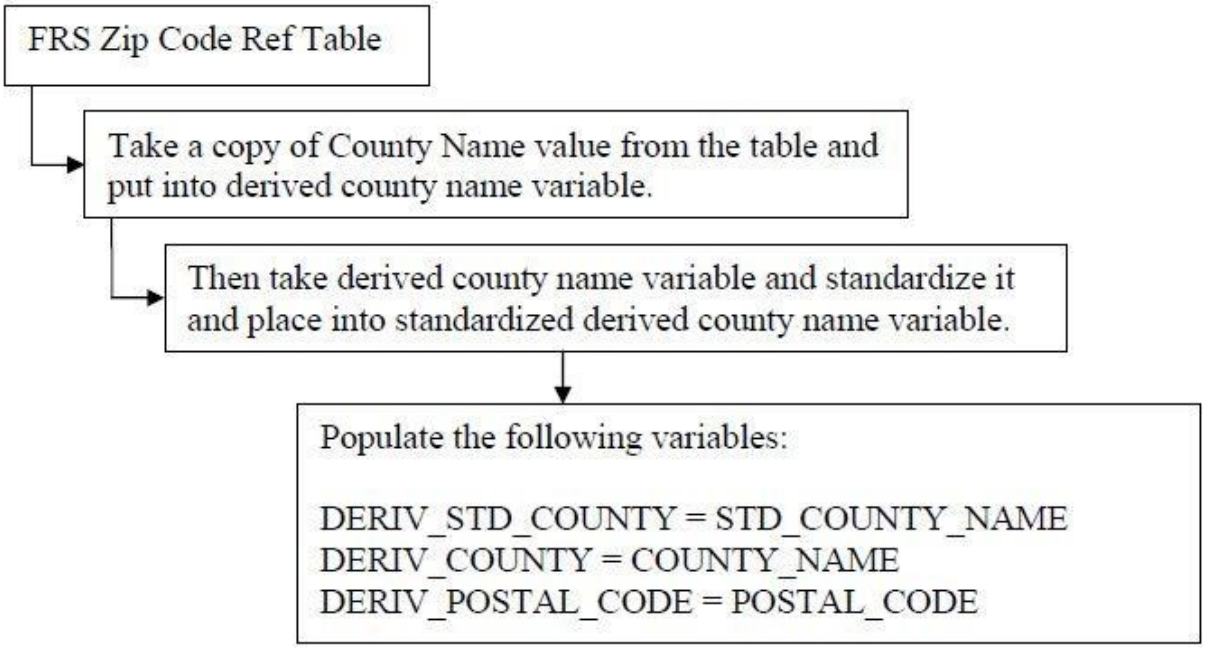
Data quality uses the city, county, state, and ZIP code fields to standardize and determine the quality of the data. If any of the data values in those four fields are missing or invalid, that data value is marked in the data quality code field. The lowercase letters identify what problem has been identified, and the uppercase letters identify which data element has been identified with having a problem. Please refer to the chart below for a list of the data quality codes that are used in the FRS database.

Exhibit 5-1: Data Quality Codes Used in FRS

Code	Description
i	Invalid: used for ZIP codes, County names and City names that do not correspond with one another.
e	Erroneous: Used for facility names and addresses that contains anomalies.
m	Missing: Used for any core data element containing no information (Facility Name, Address, City, County, State or ZIP Code).
A	Address
N	Facility Name
C	City Name
O	County Name
Z	ZIP Code
M	Combination of city and county
V	Valid

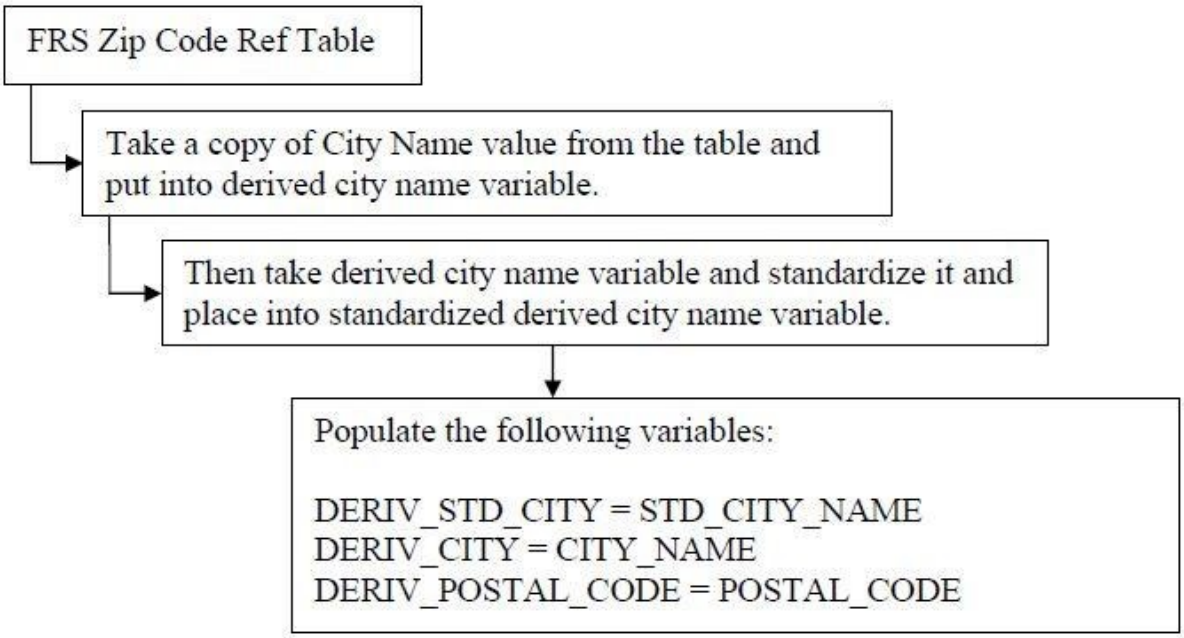
The data quality code field is determined by calling the FRS data quality code procedure. There are three different cases that determine the outcome of this procedure. The first case is concerned with missing or invalid county name. The county name is derived from the FRS ZIP code reference table from the county name column for a particular record. The county name is then standardized using the city and county name procedure and is stored into a variable called the derived standard county variable. Also, the derived postal code variable is populated with the value in the postal code field. Please refer to the following for a graphical representation of the first case.

Exhibit 5-2: Case Concerning Missing or Invalid County Name



The second case involves missing or invalid city name. The city name value is derived from the FRS ZIP code reference table as in the first case and then is put into the derived city name variable for standardization. The city and county name standardization procedure then occurs and creates a standardized city name variable and populates it with the result. Also, the derived standard county variable populates the standard county name field. The derived county name variable populates the county name field. The derived postal code variable populates the postal code field. Please refer to the following for a graphical representation of the second case.

Exhibit 5-3: Case Concerning Missing or Invalid City Name



The third case involves missing or invalid ZIP codes. The postal code is derived from the FRS zip code reference table and populates the following fields:

- DERIV_STD_COUNTY = STD_COUNTY_NAME
- DERIV_STD_CITY = STD_CITY_NAME
- DERIV_COUNTY = COUNTY_NAME
- DERIV_CITY = CITY_NAME

All above populated values are updated in the FRS RES Registry table.

6 Summary

As seen throughout this document, the parser and standardization procedures contain multiple processes for the data to go through. First the data is geocoded; if the record cannot be geocoded (i.e., the geocoder fails), the data is divided using the parser procedure into different fields corresponding to the type of data it is. Then these values are standardized by their respective procedures to allow for consistency, such as the standardized name procedure that will standardize the primary name field. These values are then validated using the data quality code procedure to make sure the data is not null or inconsistent. Once this code is assigned to each record, the data is integrated into the FRS database.

The parser and standardization procedures are important because they ensure the accuracy of the data. If any data is found to be inaccurate, the procedures will flag the data records for manual review, and an analyst will review the record for inconsistencies and correct any data that is incorrect. These procedures also improve the matching purpose by making sure that the data is accurate and uniform. Parsed and standardized values for each record are compared to parsed and standardized values of other records and used in the merging and linking process. Without accuracy and uniformity, there will be multiple records within FRS that could be the same facility; however, the difference in data between the records may not suggest that they are the same and therefore, the records will be left unlinked. The automation of these procedures saves time and money compared to when the parsing and standardization process is not used.

Appendix A: FRS Reference Table for Standardized Facility Names

The FRS standardization routine replaces any values in the 'Original Value' column below with the value in the 'Standardized Value' column. See **Section 3** for a description of this process.

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
ONE	1	CPTAL	CPTL
TWO	2	CPTOL	CPTL
THREE	3	CORRECTION	CRRCTN
FOUR	4	CORRECTIONAL	CRRCTN
FIVE	5	COURT	CT
SIX	6	CRT	CT
SEVEN	7	CEN	CTR
EIGHT	8	CENT	CTR
NINE	9	CENTER	CTR
TEN	10	CENTERS	CTR
ELEVEN	11	CENTR	CTR
TWELVE	12	CENTRE	CTR
THIRTEEN	13	CNTR	CTR
FOURTEEN	14	DEPART	DEPT
FIFTEEN	15	DEPARTM	DEPT
SIXTEEN	16	DEPARTMENT	DEPT
SEVENTEEN	17	DEPARTMNT	DEPT
EIGHTEEN	18	DEVEL	DEV
NINETEEN	19	DEVELOP	DEV
TWENTY	20	DEVELOPM	DEV
ACAD	ACDMY	DEVELOPMEN	DEV
ACADEM	ACDMY	DEVELOPMENT	DEV
ACADEMY	ACDMY	DEVELOPMENTAL	DEV
ADHESIVE	ADHSV	DEVELOPMNT	DEV
ADHESIVES	ADHSV	DEVELOPMT	DEV
ADMIN	ADMN	DEVELP	DEV
ADMINI	ADMN	DEVELPMT	DEV
ADMINIST	ADMN	DEVLMNT	DEV
ADMINISTER	ADMN	DEVLPMNT	DEV
ADMINISTR	ADMN	DEVMT	DEV

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
ADMINISTRA	ADMN	DVLOPMT	DEV
ADMINISTRATI	ADMN	DVLPMNT	DEV
ADMINISTRATION	ADMN	DVLPMT	DEV
ADMINISTRATIVE	ADMN	DISTILLER	DISTLR
ADMINISTRATN	ADMN	DISTILLERY	DISTLR
ADMINISTRATOR	ADMN	DISTILLING	DISTLR
ADMINISTRATV	ADMN	DIVISION	DIV
ADMINSTR	ADMN	DIVISIONAL	DIV
ADMNSTRN	ADMN	DIVISIONS	DIV
ADMR	ADMN	DIVSN	DIV
ADMSTR	ADMN	ECOM	ECON
ADMSTRN	ADMN	ECONOMIC	ECON
ADVERT	ADVT	ECONOMIST	ECON
ADVERTISE	ADVT	ECONOMY	ECON
ADVERTISER	ADVT	ELECT	ELEC
ADVERTISERS	ADVT	ELECTRIC	ELEC
ADVERTISIN	ADVT	ELECTRICAL	ELEC
ADVERTISING	ADVT	ELECTRICIAN	ELEC
ADVERTISNG	ADVT	ELECTRICITY	ELEC
ADVG	ADVT	ELECTRICS	ELEC
ADVR	ADVT	ELECTRONIC	ELEC
ADVTG	ADVT	ELECTRONICS	ELEC
ADVTNG	ADVT	ENERGY	ENGRY
AFFILIATE	AFFL	ENGY	ENGRY
AFFILIATED	AFFL	ENRG	ENGRY
AFFILIATES	AFFL	ENTERP	ENTRPRS
AFFILIATION	AFFL	ENTERPRISE	ENTRPRS
AFFILIATIONS	AFFL	ENTERPRISES	ENTRPRS
AGENC	AGCY	ENTPR	ENTRPRS
AGENCIES	AGCY	ENTRPR	ENTRPRS
AGENCY	AGCY	ENTS	ENTRPRS
AGNCY	AGCY	ENVIRON	ENVIR
AGRCLTRL	AGRI	ENVIRONMENT	ENVIR

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
AGRICULTURAL	AGRI	ENVIRONMENTAL	ENVIR
AGRICULTURE	AGRI	ENVRMT	ENVIR
ALGNMNT	ALIGN	ENVRMTL	ENVIR
ALGNMT	ALIGN	ENVRNMTL	ENVIR
ALIG	ALIGN	ENVRONMEN	ENVIR
ALIGNING	ALIGN	ESTATES	ESTATE
ALIGNMENT	ALIGN	FABR	FAB
ALIGNMENTS	ALIGN	FABRIC	FAB
ALIGNMNT	ALIGN	FAC	FACLTY
ALIGNMT	ALIGN	FACILITIES	FACLTY
ALIMENT	ALIGN	FACILITY	FACLTY
ALLIANCE	ALLNCE	FIBERGLASS	FBRGLS
ALARM	ALRM	FDRL	FED
ALARMS	ALRM	FEDERAL	FED
ALUM	ALUMN	FEDL	FED
ALUMINUM	ALUMN	FDRY	FNDRY
AMERICA	AMER	FNDY	FNDRY
AMERICAN	AMER	FOUNDRY	FNDRY
AMMONIA	AMMN	FREIGHT	FRGHT
NH3	AMMN	FRGT	FRGHT
AMMUNITION	AMMO	FRT	FRGHT
&	AND	FISHERY	FSHRY
&&	AND	FORT	FT
+	AND	GASOLINE	GAS
ANESTHESIA	ANESTHES	GEOLOGICAL	GEO
ANESTHESIOLOGY	ANESTHES	GEOLOGY	GEO
ANAL	ANLYST	GEOPHYSICAL	GEO
ANALY	ANLYST	GENERAL	GNRL
ANALYS	ANLYST	GENL	GNRL
ANALYSIS	ANLYST	GEN	GNRTNG
ANALYST	ANLYST	GENERATING	GNRTNG
ANALYTIC	ANLYST	GENERATION	GNRTNG
ANALYTICAL	ANLYST	GENERATOR	GNRTNG

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
ANLST	ANLYST	GYMNASIUM	GYM
ANLYS	ANLYST	GYMNASTIC	GYM
ANIMAL	ANML	HARDWARE	HDWR
ANIML	ANML	HOSPIT	HOSP
ANMAL	ANML	HOSPITAL	HOSP
ANNEX	ANX	HSPTL	HOSP
ANNX	ANX	HDQS	HQ
APART	APT	HDQTRS	HQ
APARTMENT	APT	HEADQUARTERS	HQ
ARCHITECT	ARCH	HQS	HQ
ARCHITECTURAL	ARCH	HQTS	HQ
ARCHITECTURE	ARCH	IMPORT	IMPRT
ARCHTCT	ARCH	IMPORTERS	IMPRT
AIRP	ARPRT	IMPORTING	IMPRT
AIRPORT	ARPRT	IMPORTS	IMPRT
AIRPORTS	ARPRT	INCOR	INC
AIRPT	ARPRT	INCORP	INC
ARPT	ARPRT	INCORPORATED	INC
ARSENAL	ARSNL	INDL	IND
ASSEM	ASMBLY	INDS	IND
ASSEMBLE	ASMBLY	INDSTRL	IND
ASSEMBLER	ASMBLY	INDTRY	IND
ASSEMBLY	ASMBLY	INDUS	IND
ASO	ASSOC	INDUSTR	IND
ASOC	ASSOC	INDUSTRIA	IND
ASS	ASSOC	INDUSTRIAL	IND
ASSC	ASSOC	INDUSTRIES	IND
ASSCD	ASSOC	INDUSTRL	IND
ASSCE	ASSOC	INDUSTRY	IND
ASSCO	ASSOC	INFIRM	INFRM
ASSO	ASSOC	INFIRMARY	INFRM
ASSOCATE	ASSOC	INSTIT	INST
ASSOCATED	ASSOC	INSTITUE	INST

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
ASSOCI	ASSOC	INSTITUT	INST
ASSOCIA	ASSOC	INSTITUTE	INST
ASSOCIAT	ASSOC	INSTITUTION	INST
ASSOCIATE	ASSOC	INSTITUTIONAL	INST
ASSOCIATED	ASSOC	INTERNATI	INTL
ASSOCIATES	ASSOC	INTERNATIO	INTL
ASSOCIATION	ASSOC	INTERNATION	INTL
ASSOCIATIONS	ASSOC	INTERNATIONA	INTL
ASSOD	ASSOC	INTERNATIONAL	INTL
ASSRNC	ASSURNC	INTERNATL	INTL
ASSUR	ASSURNC	INTNL	INTL
ASSURANCE	ASSURNC	INTRNTNL	INTL
ASSURE	ASSURNC	INVEST	INVSTMNT
AUTH	ATHRTY	INVESTMENT	INVSTMNT
AUTHORI	ATHRTY	INVESTMNT	INVSTMNT
AUTHORITIES	ATHRTY	INVESTMT	INVSTMNT
AUTHORITY	ATHRTY	INVST	INVSTMNT
AUTHY	ATHRTY	JCTION	JCT
ATN	ATTN	JCTN	JCT
ATTENTION	ATTN	JUNCTION	JCT
ATTNTN	ATTN	JUNCTN	JCT
AUTOMOBILE	AUTO	JUNCTON	JCT
AUTOMOTIVE	AUTO	LABORATORIES	LAB
AUXIL	AUX	LABORATORY	LAB
AUXILARY	AUX	LABS	LAB
AUXILIARY	AUX	LANDFILL	LNDFLL
AUXILRY	AUX	LDFL	LNDFLL
AV	AVE	LNDFL	LNDFLL
AVENUE	AVE	LNFL	LNDFLL
BAPTIST	BAPT	LAUNDROMAT	LNDRY
BPTST	BAPT	LAUNDRY	LNDRY
BAT	BATT	LANDSCAPE	LNDSCP
BATTERIES	BATT	LIMITED	LTD

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
BATTERY	BATT	LMTD	LTD
BTRY	BATT	LUBRICANT	LUBE
BARBEQUE	BBQ	LUBRICANTS	LUBE
BARBQUE	BBQ	LUBRICATION	LUBE
BEACH	BCH	LUBRICNT	LUBE
BEVERAGE	BEV	LIVESTOCK	LVSTCK
BEVERAGES	BEV	LVSTK	LVSTCK
BICYCLE	BIKE	MACHINE	MACH
BIOTECHNOLOGY	BIOTECH	MACHINER	MACH
BUILDING	BLDG	MACHINERY	MACH
BUILDINGS	BLDG	MACHINING	MACH
BUILDER	BLDR	MACHINIST	MACH
BOULEVARD	BLVD	MCH	MACH
BANK	BNK	MCHINE	MACH
BD	BOARD	MCHY	MACH
BRD	BOARD	MATERIAL	MATL
BOROUGH	BORO	MATERIALS	MATL
BREWERY	BREW	MATLS	MATL
BREWING	BREW	MANF	MFG
BRDGE	BRG	MANUF	MFG
BRIDGE	BRG	MANUFACTURE	MFG
BARREL	BRL	MANUFACTURER	MFG
BROTHER	BROS	MANUFACTURI	MFG
BROTHERS	BROS	MANUFACTURING	MFG
BOTLER	BTLG	MFGNG	MFG
BOTLNG	BTLG	MFGR	MFG
BOTTLING	BTLG	MFR	MFG
BTLNG	BTLG	MANAGE	MGMT
BTLR	BTLG	MANAGEMENT	MGMT
BTTLR	BTLG	MANAGER	MGMT
BSNS	BUSN	MANGMNT	MGMT
BUSINES	BUSN	MGMENT	MGMT
BUSINESS	BUSN	MNAGER	MGMT

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
BUSN	BUSN	MNGMNT	MGMT
CAPTAIN	CAPT	MNGMT	MGMT
CPT	CAPT	MNGN	MGMT
CATHOLIC	CATH	MNGR	MGMT
CTHLC	CATH	MINES	MINE
CABINET	CBNT	MINING	MINE
CABINETMAKER	CBNT	MINNG	MINE
CABINETRY	CBNT	MARKET	MKT
CABINETS	CBNT	MRKT	MKT
CEMENT	CEM	MARKETING	MKTG
CERTD	CERT	MKTING	MKTG
CERTIF	CERT	MKTNG	MKTG
CERTIFIE	CERT	MRKTG	MKTG
CERTIFIED	CERT	MAINT	MNTNC
CHEMCAL	CHEM	MAINTENANCE	MNTNC
CHEMICAL	CHEM	MTNCE	MNTNC
CHEMICALS	CHEM	MNTN	MTN
CHEMIST	CHEM	MOUNTAIN	MTN
CHEMISTS	CHEM	NAT	NATL
CHEMS	CHEM	NATIONAL	NATL
CHMST	CHEM	OPER	OPR
CHEV	CHEVY	OPERATING	OPR
CHEVROLET	CHEVY	OPERATION	OPR
CHCKN	CHICK	OPERATIONAL	OPR
CHICKEN	CHICK	OPERATIONS	OPR
CHKN	CHICK	OPERATOR	OPR
CHILD	CHLD	OPRTR	OPR
CHILDHOOD	CHLD	PETROLEUM	PETRO
CHILDREN	CHLD	PHARMACEUTICAL	PHARM
CHILDRENS	CHLD	PHARMACIST	PHARM
CHILDS	CHLD	PHARMACY	PHARM
CHLDHD	CHLD	FOTO	PHOTO
CHLDRN	CHLD	PHOTOGRAPH	PHOTO

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
CHRCH	CHURCH	PHOTOGRAPHER	PHOTO
CHUR	CHURCH	PHOTOGRAPHY	PHOTO
CHURC	CHURCH	PACKAGE	PKG
CLEANER	CLNR	PACKAGING	PKG
CLEANERS	CLNR	PACKG	PKG
CLEANG	CLNR	PASTICS	PLAS
CLEANING	CLNR	PLASTIC	PLAS
CLEANSER	CLNR	PLANT	PLT
CLOTHES	CLTHS	PLANTS	PLT
CLOTHING	CLTHS	PLNT	PLT
CMMNTY	CMNTY	PRODS	PROD
CMTY	CMNTY	PRODUCE	PROD
COMMUNITY	CMNTY	PRODUCING	PROD
COMNTY	CMNTY	PRODUCT	PROD
CMPR	CMPSR	PRODUCTS	PROD
CMPRSR	CMPSR	PSYCHIATRIC	PSYCH
CMPRSSR	CMPSR	PSYCHIATRIST	PSYCH
COMPRESSOR	CMPSR	PSYCHIATRY	PSYCH
CAMPGROUND	CMPST	PSYCHOLOGICAL	PSYCH
CAMPSITE	CMPST	PSYCHOLOGIST	PSYCH
COMPTR	CMPTR	PSYCHOLOGY	PSYCH
COMPU	CMPTR	PUBLICATION	PUB
COMPUTER	CMPTR	PUBLICATIONS	PUB
COMPUTING	CMPTR	PUBLISHER	PUB
CNTRCTR	CNTRCT	PUBLISHERS	PUB
CONTG	CNTRCT	PUBLISHING	PUB
CONTR	CNTRCT	POWER	PWR
CONTRAC	CNTRCT	REGION	REGN
CONTRACT	CNTRCT	REGIONAL	REGN
CONTRACTIN	CNTRCT	REMANUFACTURER	REMFG
CONTRACTING	CNTRCT	REMANUFACTURERS	REMFG
CONTRACTOR	CNTRCT	REMANUFACTURING	REMFG
CONTRACTORS	CNTRCT	REF	RFNRY

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
CONTRG	CNTRCT	REFINERY	RFNRY
COUNTRY	CNTRY	REFINING	RFNRY
COUNTRYSIDE	CNTRY	REFY	RFNRY
COUNTY	CNTY	RAIL	RR
COMP	CO	RAILROAD	RR
COMPAN	CO	RAILWAY	RR
COMPANIES	CO	SCHO	SCH
COMPANY	CO	SCHOOL	SCH
COMPNAY	CO	SCHOOLS	SCH
COMPNY	CO	SPEC	SPCLTY
CLCTN	COLLECT	SPECIALIST	SPCLTY
COLLECTION	COLLECT	SPECIALTIES	SPCLTY
COLLECTIONS	COLLECT	SPECIALTY	SPCLTY
COLLECTN	COLLECT	STAT	STA
COLLECTOR	COLLECT	STATION	STA
COLLECTORS	COLLECT	STATN	STA
COMMUN	COMM	STN	STA
COMMUNICATI	COMM	SER	SVC
COMMUNICATION	COMM	SERV	SVC
COMMUNICATIONS	COMM	SERVIC	SVC
COMMUNICTN	COMM	SERVICE	SVC
CNDMNM	CONDO	SERVICES	SVC
CONDOMINIUM	CONDO	SRVCS	SVC
CONDOMINIUMS	CONDO	SVCS	SVC
CONDOS	CONDO	SYSTEM	SYS
CONFERENCE	CONF	SYSTEMS	SYS
CONFRENCE	CONF	TECHNICAL	TECH
CNSRVTN	CONSERVE	TECHNICIAN	TECH
CNSVTN	CONSERVE	TECHNOLOGICAL	TECH
CONSER	CONSERVE	TECHNOLOGIES	TECH
CONSERV	CONSERVE	TECHNOLOGIST	TECH
CONSERVATION	CONSERVE	TECHNOLOGY	TECH
CNST	CONSTRCTN	TECHS	TECH

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
CNSTR	CONSTRCTN	TELCOMMN	TELE
CNSTRCTN	CONSTRCTN	TELECOMMUNICATION	TELE
CONST	CONSTRCTN	TELEPHONE	TELE
CONSTN	CONSTRCTN	TERM	TERML
CONSTR	CONSTRCTN	TERMIAL	TERML
CONSTRN	CONSTRCTN	TERMINAL	TERML
CONSTRTN	CONSTRCTN	TERMINALING	TERML
CONSTRUCT	CONSTRCTN	TERMINALS	TERML
CONSTRUCTION	CONSTRCTN	TRANSMISSION	TRANS
CONSTRUCTN	CONSTRCTN	TRANSMSSN	TRANS
CONSTRUCTOR	CONSTRCTN	TRANSPORT	TRANSP
CONTRL	CONTRL	TRANSPORTATION	TRANSP
CONTROL	CONTRL	TRANSPORTER	TRANSP
CTL	CONTRL	TRANSPORTERS	TRANSP
CTRL	CONTRL	TRNSPRT	TRANSP
COOPERATIVE	COOP	TRNSPRTN	TRANSP
COOPERATIVES	COOP	TOWNSHIP	TWP
CORPORATE	CORP	TWNSHP	TWP
CORPORATIN	CORP	UNIVERSITIES	UNIV
CORPORATIO	CORP	UNIVERSITY	UNIV
CORPORATION	CORP	UNITED STATES	US
CORPORT	CORP	UNITED STATES OF AMERICA	US
COATER	COTG	USA	US
COATERS	COTG	UTILITIES	UTIL
COATING	COTG	UTILITY	UTIL
COATINGS	COTG	VETERINARIAN	VET
CAPITAL	CPTL	WAREHOUSE	WHSE
CAPITOL	CPTL	WHS	WHSE

Appendix B: FRS Reference Table for Abbreviated Names Replacement

The FRS standardization routine replaces standard name abbreviations in the 'Name Abbreviations' columns with the value in the 'New Data Value' columns below. See **Section 3** for a description of this process.

Name Abbreviations	New Data Value	Name Abbreviations	New Data Value
CO OF	COUNTY OF	MICHIGAN	MI
CORP	;	MINNESOTA	MN
INC	;	MISSISSIPPI	MS
U S	US	MISSOURI	MO
U.S.	US	MONTANA	MT
U. S.	US	NEBRASKA	NE
L L P	;	NEVADA	NV
L L C	;	NEW HAMPSHIRE	NH
LTD	;	NEW JERSEY	NJ
;;	;	NEW MEXICO	NM
THE	" "	NEW YORK	NY
ALABAMA	AL	NORTH CAROLINA	NC
ALASKA	AK	NORTH DAKOTA	ND
AMERICAN SAMOA	AS	OHIO	OH
ARIZONA	AZ	OKLAHOMA	OK
ARKANSAS	AR	OREGON	OR
CALIFORNIA	CA	PENNSYLVANIA	PA
COLORADO	CO	PUERTO RICO	PR
CONNECTICUT	CT	RHODE ISLAND	RI
DISTRICT OF COLUMBIA	DC	SOUTH CAROLINA	SC
DELAWARE	DE	TENNESSEE	TN
FLORIDA	FL	TEXAS	TX
GEORGIA	GA	UTAH	UT
GUAM	GU	VERMONT	VT
HAWAII	HI	VIRGIN ISLANDS	VI
IDAHO	ID	VIRGINIA	VA
ILLINOIS	IL	WASHINGTON	WA
INDIANA	IN	WEST VIRGINIA	WV
IOWA	IA	WISCONSIN	WI

Name Abbreviations	New Data Value	Name Abbreviations	New Data Value
KANSAS	KS	WYOMING	WY
KENTUCKY	KY	PLT NUMBER	PLT
LOUISIANA	LA	BLDG NUMBER	BLDG
MAINE	ME	UNIT NUMBER	UNIT
MARYLAND	MD	DIV NUMBER	DIV
MASSACHUSETTS	MA		

Appendix C: FRS Reference Table for Standardized Address Data Values

The FRS standardization routine replaces the address data entered in the 'Original Value' columns with the value in the 'Standardized Value' columns below. See **Section 4** for a description of this process.

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
ONE	1	HIWAY	HWY
TWO	2	HWYS	HWY
APPROX	*(null)*	HIGHWY	HWY
ABOUT	*(null)*	HIWY	HWY
JCT	*(null)*	HWAY	HWY
INTERSECTION	*(null)*	HYWAY	HWY
INTERSEC	*(null)*	INDUSTRIAL	INDL
BETWEEN	*(null)*	IND	INDL
AROUND	*(null)*	INLET	INLT
ON	*(null)*	ISLND	ISLAND
LOCATED	*(null)*	ISLANDS	ISLAND
THE	*(null)*	ISLNDS	ISLAND
NEAR	*(null)*	ISS	ISLAND
AT	*(null)*	ISLES	ISLE
FIRST	1ST	JUNCTION	JCT
SECOND	2ND	JCTION	JCT
THIRD	3RD	JCTN	JCT
FOURTH	4TH	JUNCTN	JCT
FIFTH	5TH	JUNCTON	JCT
SIXTH	6TH	JCTNS	JCT
SEVENTH	7TH	JCTS	JCT
EIGHTH	8TH	JUNCTIONS	JCT
NINETH	9TH	KEYS	KEY
TENTH	10TH	KYS	KEY
ELEVENTH	11TH	LAKES	LAKE
TWELFTH	12TH	LKS	LAKE
THIRTEENTH	13TH	LN	LANE
FOURTEENTH	14TH	LA	LANE
FIFTEENTH	15TH	LANES	LANE

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
SIXTEENTH	16TH	LANDING	LNDG
SEVENTEENTH	17TH	LNDNG	LNDG
EIGHTEENTH	18TH	LODG	LODGE
NINETEENTH	19TH	LDGE	LODGE
TWENTIETH	20TH	LOOPS	LOOP
THIRTIETH	30TH	MNR	MANOR
FORTIETH	40TH	MANORS	MANOR
FIFTIETH	50TH	MNRS	MANOR
SIXTIETH	60TH	MDW	MEADOW
SEVENTIETH	70TH	MDWS	MEADOW
EIGHTIETH	80TH	MEADOWS	MEADOW
NINETIETH	90TH	MEDOWS	MEADOW
ALLEY	ALY	MILES	MI
ALLEE	ALY	MILE	MI
ALLY	ALY	MILLS	MILL
ANEX	ANX	MNT	MOUNT
ANNEX	ANX	MILEPOST	MP
ANNX	ANX	MNTAIN	MTN
ARCADE	ARC	MNTN	MTN
AVENUE	AVE	MOUNTAIN	MTN
AV	AVE	MOUNTIN	MTN
AVENIDA	AVE	MTIN	MTN
AVEN	AVE	MNTNS	MTN
AVENU	AVE	MOUNTAINS	MTN
AVN	AVE	MOTORWAY	MTWY
AVNUE	AVE	NORTHEAST	NE
BEACH	BCH	NORESTE	NE
BUILDING	BLDG	NORTE	NORTH
BLUF	BLF	NORTHWEST	NW
BLUFF	BLF	NOROESTE	NW
BLUFFS	BLF	ORCH	ORCHARD
BLOCK	BLK	ORCHRD	ORCHARD
BOULEVARD	BLVD	OPAS	OVERPASS

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
BL	BLVD	OVAL	OVL
BOUL	BLVD	PK	PARK
BOULV	BLVD	PRK	PARK
BEND	BND	PKWYS	PARKWAY
BR	BRANCH	PKWY	PARKWAY
BRNCH	BRANCH	PKY	PARKWAY
BRG	BRIDGE	PKWAY	PARKWAY
BRDGE	BRIDGE	PARKWY	PARKWAY
BRK	BROOK	PSGE	PASSAGE
BROOKS	BROOK	PNES	PINES
BOT	BTM	PLC	PLACE
BOTTM	BTM	PL	PLACE
BOTTOM	BTM	PLN	PLAIN
BYPA	BYP	PLNS	PLAINS
BYPAS	BYP	PLZ	PLAZA
BYPASS	BYP	POB	POBOX
BYPS	BYP	PO BOX	POBOX
BAYOO	BYU	P O BOX	POBOX
BAYOU	BYU	POST OFFICE BOX	POBOX
CIR	CIRCLE	DRAWER	POBOX
CIRC	CIRCLE	CALL BOX	POBOX
CIRCL	CIRCLE	PT	POINT
CRCL	CIRCLE	PTS	POINT
CRCLE	CIRCLE	PRARIE	PRAIRIE
CIRCLES	CIRCLE	RNCHS	RANCH
CLIFF	CLF	RNCH	RANCH
CLIFFS	CLF	RDGS	RIDGE
CLFS	CLF	RDG	RIDGE
CLB	CLUB	RDGE	RIDGE
CMN	COMMON	RVR	RIVER
COMMONS	COMMON	RIVR	RIVER
CORS	CORNER	RIV	RIVER
CORNERS	CORNER	RDS	ROAD

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
COR	CORNER	RD	ROAD
CAMP	CP	ROUTE	ROUTE
CMP	CP	RRTE	ROUTE
CAPE	CPE	RR	ROUTE
CRK	CREEK	RT	ROUTE
CK	CREEK	RRT	ROUTE
CRECENT	CRESCENT	ROUTES	ROUTE
CRES	CRESCENT	RFD	ROUTE
CRESENT	CRESCENT	RTS	ROUTE
CRSCNT	CRESCENT	RTE	ROUTE
CRSENT	CRESCENT	SOUTHEAST	SE
CRSNT	CRESCENT	SURESTE	SE
XING	CROSSING	SHLS	SHOAL
CRSSING	CROSSING	SHL	SHOAL
CRSSNG	CROSSING	SHRS	SHORE
XRD	CROSSROAD	SHOARS	SHORE
COURSE	CRSE	SHR	SHORE
CAUSEWAY	CSWY	SHOAR	SHORE
CAUSWAY	CSWY	SKWY	SKYWAY
COURT	CT	SO	SOUTH
CRT	CT	SUR	SOUTH
COURTS	CT	SPRNG	SPRING
CEN	CTR	SPNG	SPRING
CENT	CTR	SPG	SPRING
CENTER	CTR	SPRNGS	SPRINGS
CENTR	CTR	SPNGS	SPRINGS
CENTRE	CTR	SPGS	SPRINGS
CNTER	CTR	SQ	SQUARE
CNTR	CTR	SQRS	SQUARE
CENTERS	CTR	SQU	SQUARE
CURV	CURVE	SQRE	SQUARE
COVE	CV	SQR	SQUARE
COVES	CV	STREETS	ST

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
CANYN	CYN	STS	ST
CANYON	CYN	STRT	ST
CNYN	CYN	STR	ST
DR	DRIVE	STREET	ST
DRV	DRIVE	STREEET	ST
DRIVES	DRIVE	STN	STATION
ESTE	EAST	STATN	STATION
EXPRESSWAY	EXPY	STA	STATION
EXPW	EXPY	SUITE	STE
EXP	EXPY	SUMIT	SUMMIT
EXPR	EXPY	SUMITT	SUMMIT
EXPRESS	EXPY	SMT	SUMMIT
EXTENSION	EXT	SOUTHWEST	SW
EXTN	EXT	SUROESTE	SW
EXTNSN	EXT	TER	TERRACE
EXTENSIONS	EXT	TERR	TERRACE
FRRY	FERRY	TRWY	THROUGHWAY
FRY	FERRY	TRKS	TRACK
FLD	FIELD	TRK	TRACK
FLDS	FIELD	TRAK	TRACK
FIELDS	FIELD	TRFY	TRAFFICWAY
FALLS	FLS	TRLS	TRAIL
FLAT	FLT	TR	TRAIL
FLATS	FLT	TRL	TRAIL
FORESTS	FOREST	TUNNL	TUNNEL
FORD	FRD	TUNEL	TUNNEL
FORDS	FRD	TUNL	TUNNEL
FORGE	FRG	TUNLS	TUNNEL
FORGES	FRG	TNPK	TURNPIKE
FORK	FRK	TPKE	TURNPIKE
FORKS	FRK	TURNPK	TURNPIKE
FORT	FT	TRPK	TURNPIKE
FREEWAY	FWY	TRNPK	TURNPIKE

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
FREEWY	FWY	TPK	TURNPIKE
FRWAY	FWY	UPAS	UNDERPASS
FRWY	FWY	UN	UNION
GARDEN	GDN	URB	URBANIZATION
GARDN	GDN	URBANIZACION	URBANIZATION
GRDEN	GDN	VLYS	VALLEY
GRDN	GDN	VLY	VALLEY
GARDENS	GDN	VLLY	VALLEY
GDNS	GDN	VALLY	VALLEY
GRDNS	GDN	VIADCT	VIADUCT
GLN	GLEN	VIA	VIADUCT
GLENS	GLEN	VDCT	VIADUCT
GROV	GROVE	VWS	VIEW
GROVES	GROVE	VW	VIEW
GATEWAY	GTWY	VLGS	VILLAGE
GATEWY	GTWY	VLG	VILLAGE
GATWAY	GTWY	VILLIAGE	VILLAGE
HVN	HAVEN	VILLG	VILLAGE
HAVN	HAVEN	VILLAG	VILLAGE
HARBOR	HBR	VILL	VILLAGE
HARBR	HBR	VL	VILLE
HARB	HBR	VSTA	VISTA
HARBORS	HBR	VST	VISTA
HRBOR	HBR	VIST	VISTA
HEIGHTS	HTS	VIS	VISTA
HEIGHT	HTS	OESTE	WEST
HGTS	HTS	WAY	WY
HIGHWAY	HWY		

Appendix D: FRS Reference Table for Standardized City and County Data Values

Below is a list of city and county name standardization values used during the city and county standardization procedure. The FRS standardization routine replaces the value entered in the 'Original Value' columns with the value in the 'Standardized Value' columns below. See **Section 4.3** for a description of this process.

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
AFB	*(null)*	MOUNT	MT
AIR FORCE BASE	*(null)*	MOUNTAIN	MT
BCH	BEACH	MTN	MT
BORO	*(null)*	NE	NEBRASKA
BYU	BAYOU	NEAR	*(null)*
C	*(null)*	NEARBY	*(null)*
CA	CALIFORNIA	NJ	NEWJERSEY
CENTRE	CENTER	NO	N
CIT	*(null)*	NORTH	N
CITY	*(null)*	NORTHEAST	NE
CT	CONNECTICUT	NORTHWEST	NW
CNTR	CENTER	NV	NEVADA
CNTY	*(null)*	NY	NEWYORK
COLO	CO	NYC	NEWYORK
COLORADO	CO	OH	OHIO
COMPANY	CO	PARK	PK
CORNERS	CORNER	PEAK	PK
COUNTY	*(null)*	PHX	PHOENIX
CT	CONNECTICUT	POINT	PT
CTR	CENTER	POINTE	PT
CTY	*(null)*	PORT	PT
CY	*(null)*	SAINT	ST
DC	DISTRICTOFCOLUMBIA	SHORES	SHORE
DIV	*(null)*	SO	S
EAST	E	SOUTH	S
EASTWEST	EW	SOUTHEAST	SE
FORT	FT	SOUTHWEST	SW

ORIGINAL VALUE	STANDARDIZED VALUE	ORIGINAL VALUE	STANDARDIZED VALUE
HEIGHTS	HEIGHT	SPG	SPRING
HGT	HEIGHT	SPGS	SPRING
HGTS	HEIGHT	SPRINGS	SPRING
HTS	HEIGHT	STA	ST
IL	ILLINOIS	STATION	ST
IN	INDIANA	STE	ST
IS	ISLAND	STREET	ST
ISL	ISLAND	THE	*(null)*
ISLE	ISLAND	TN	TENNESSEE
ISLND	ISLAND	TOWNSHIP	*(null)*
JCT	JUNCTION	TWNSP	*(null)*
JUNC	JUNCTION	TWP	*(null)*
KC	KANSAS	TX	TEXAS
KS	KANSAS	UNICORP	*(null)*
KY	KENTUCKY	UNINCORP	*(null)*
LK	LAKE	UNINCORPORATED	*(null)*
MD	MARYLAND	UNIV	UNIVERSITY
ME	MAINE	UNIVER	UNIVERSITY
MH	MARSHALLISLANDS	VA	VIRGINIA
MI	MICHIGAN	VLY	VALLEY