

Background

The purpose of this information is to identify differences between the data maintained in the various data sets used to compile the inventory: Department of Revenue Tax Information, Environmental Health Database and Individual County Health Department lists. The notes contained herein are an attempt to explain irregularities in the information, how we overcame these obstacles and basic results.

Environmental Health Database (EHDB)

Alachua: (22,185 Address Records)

We were able to link 12,047 records using the tax identification numbers and addresses. The DOR data had spaces in the tax id where the EHDB contained dashes.

Baker: (2,595 Address Records)

We were able to link 1,744 records using the tax identification numbers and addresses. The DOR data did not have the dashes as the EHDB did.

Bay: (10,911 Address Records)

3,756 EHDB records are either blank or contain S/T/R only. Linking by tax identification numbers yielded 6,616 records. Utilizing addresses, another 2,486 records were successfully linked. The result is an 84% success rate.

Bradford: (3,071 Address Records)

The formatting of the tax id's in the EHDB are inconsistent. Many contain preceding S-T-R, often without leading zeros in one digit sections. The DOR data contains no reference to S-T-R. We were only able to link 158 records by tax id without a significant amount of manual review and entry. Utilizing addresses linked an additional 89 records. Lastly, we used a "Right" string concatenation to look at the EHDB tax id numbers absent S-T-R and were able to link 1,877 records.

Brevard: (19,569 Address Records)

The vast majority of EHDB records only contain S/T/R in the Property ID field. No records were matched using tax identification numbers. However, we were able to link 17,246 records using the addresses – a success rate of 88%.

Broward: (13,361 Address Records)

The EHDB does not have a single valid tax identification number in Broward County. However, using addresses alone, we were able to link nearly 11,000 records.

Calhoun: (1,635 Address Records)

Less than 20 records have tax id numbers that are something other than S-T-R or blank. Only 14 records linked using tax identification numbers. 663 records do not have a street number associated with the address. Therefore, only 153 records were linked by address.

Citrus: (25,598 Address Records)

There are no valid tax identification numbers in the EHDB. Those that have data are only the Section, Township and Range – which does not define a parcel. The addresses in the tax roll had leading zeros in the street number to make them a consistent length of 5 characters. We parsed the leading zeros and linked to the EHDB. Following the updates, we were able to match over 19,200 records or just over 75%.

Clay: (15,125 Address Records)

There are over 6,500 EHDB records without tax identification numbers or they are simply the S/T/R. We were able to link slightly more than 5,700 records by tax identification numbers. An additional 6,600 records were linked by address giving an 82% success rate.

Collier: (14,929 Address Records)

The tax identification numbers in the EHDB are very inconsistent. They are believed to be a combination of the S/T/R with the true tax identification number. However, many are missing the S/T/R portion making varying length id's. There are also hundreds of records that have leading dashes. Using addresses alone, we were only able to successfully match half of the records.

Columbia: (10,811 Address Records)

There are about 500 EHDB records without tax identification numbers. We were able to link over 7,000 using the tax identification numbers. A negligible number of additional records were linked using the addresses.

DeSoto: (3,720 Address Records)

About 450 EHDB records do not have tax identification numbers. Using tax id, we were able to join over 2,600 records. Linking by address provided 500 additional joins. The DOR data has the "PreDirection" after the suffix. It was altered for the purposes of the join.

Dixie: (3,377 Address Records)

Almost EHDB 600 records have null tax id or only the S-T-R. The EHDB has dashes in the tax identification number while the DOR data does not. We formatted the parcels table with the dashes as shown in the tax ID's within the EHDB. After the modifications, we were able to link 1,831 records by tax ID. There are 1,901 records without Street Numbers for the addresses which make linking by address not possible for those records. Another 129 records were linked using the addresses alone.

Duval: (31,243 Address Records)

There are approximately 19,000 records with blank or invalid tax identification numbers in the EHDB. Of the records with apparently valid id's 6,991 were linked using this method; while another 14,900 were linked using the addresses.

Escambia: (27,804 Address Records)

There are 3,405 records in the EHDB without Tax Identification numbers. The EHDB has dashes in the tax identification number while the DOR data does not. We formatted the parcels table with the dashes as shown in the tax ID's within the EHDB. After the modifications, we were able to link 21,681 records by tax ID. There are almost 500 records without street numbers. Another 4,143 were linked using addresses for a linked success rate of 93%.

Flagler: (2,769 Address Records)

The DOR data did not contain the same formatting as the EHDB tax identification numbers. DOR was updated to similar formatting. We were able to link 2,142 records using the modified tax numbers. A nominal number of additional records were linked using addresses. Many of the unmatched do not include tax identification numbers in the EHDB.

Franklin: (3,885 Address Records)

The Parcel Tax ID field within the EHDB is sparsely populated – 1,362 are not populated. Those that are populate have varying formats, lengths and completeness. Therefore, we were only able to get 33 matches. When using the addresses, we are able to match slightly greater than 800 records.

Gadsden:

DOR did not have “-” in the tax identification number. EHDB is missing preceding “0” in majority of tax identification numbers with single-digit section numbers. EHDB does not have preceding digit for jurisdiction (1-4). Utilized modified Tax ID numbers and physical addresses to link nearly 60% of addresses.

Gilchrist: (3,172 Address Records)

There are 895 records in the EHDB that do not have tax identification numbers or only have S-T-R values. We were able to link more than 1,700 records by tax id and another 200 by address. There are also 107 “Legacy” records with no information.

Glades: (2,131 Address Records)

Majority of tax id's in EHDB did not have dashes. Removed the dashes from the ones that did in order to create a join with the DOR data – also removed dashes. Using the reformatted tax identification numbers yielded over 1,100 links. Joining based on addresses yielded an additional 230 links.

Gulf: (3,467 Address Records)

There are 471 records without tax identification numbers in the EHDB. Using those that are populated and formatted correctly, we are able to match 1,240 to corresponding parcels. We were also able to match over 900 using addresses.

Hamilton: (1,766 Address Records)

390 of the records in the EHDB do not have Tax ID numbers. Many more had S-T-R preceding the correct tax identification number. Therefore, only 118 records linked automatically. Using addresses, we were able to link 23 records automatically. Using a “right” string concatenation on tax id – removing S-T-R – we were able to link another 704 Records. There are also 318 records without street numbers which makes linking by address difficult.

Hardee: (2,046 Address Records)

1,563 records were joined utilizing the tax identification numbers. Another 284 records were linked using addresses. This provided a success rate in excess of 90%.

Hendry: (4,209 Address Records)

The tax identification numbers in the EHDB are highly irregular and inconsistent. Therefore only 67 matches were found. However, using addresses, almost 2,000 records were successfully linked. After determining that there were an excessive number of addresses in the DOR data with multiple spaces between content and repairing, another 700+ records were successfully linked.

Hernando: (21,611 Address Records)

There are 14,509 records in the EHDB missing tax identification numbers. Of the ones that are not null, the vast majority contain only the Section/Township/Range and are therefore useless. Using the addresses, we were able to link 16,500 – over 75%.

Highlands: (12,276 Address Records)

The DOR data did not contain the preceding digit to the Township or Range as it should. The EHDB also used periods where there should have been dashes. Both the parcels table and EHDB records were modified to be consistent. Following the updates, 5,386 records were successfully linked. Utilizing addresses, we were able to link an additional 3,141 records.

Hillsborough: (26,789 Address Records)

The tax identification numbers in the EHDB are actually the alternate key stored by the county. This is actually better since the tax identification number is much longer and more likely to be entered improperly. The DOR tax roll does not have the period after the 6th digit, but the database was updated to reflect it. Using the Alternate Key and PropertyID, we were able to link over 18,000 records successfully. Linking by addresses yielded another 4,000 linked records – providing a link rate greater than 82%.

Holmes: (3,294 Address Records)

Most of the Tax ID's in the EHDB are invalid. They simply contain the Section, Township and Range for the property – and include “SEC”, “TWN” and “RNG” in the field. Therefore, we were only able to link 40 records based on tax identification number. An additional 320 were linked using the addresses. The physical address in the tax roll for Holmes County is sparsely populated.

Indian River: (17,234 Address Records)

There are 7,369 EHDB records without tax identification numbers or only have S-T-R values. There were also inconsistencies between the DOR tax ID's and the EHDB property id's. After iterative cleaning, we were able to link 8,258 records. Using the addresses, we were able to link 2,574 records. The numbered address for street names in the EHDB do not contain suffixes (24 versus 24th) as the DOR address does. This causes address not to match.

Jackson: (9,471 Address Records)

The vast majority of the tax identification numbers in the EHDB contain only the S/T/R or are blank. Therefore, only 92 records were linked utilizing this method. Using the addresses, we were able to link an additional 5,100 records. Many of the unlinked records have incomplete addresses – either missing street numbers or suffixes.

Jefferson: (3,499 Address Records)

The tax identification numbers in the EHDB are fairly consistent. Therefore, we were able to link over 2,500 records using this method. Additionally, another 200 records were linked by address. We were therefore able to obtain an 80% success rate. Many of the unlinked records have incomplete addresses – either missing street numbers or suffixes – and are lacking the property identification number.

Lafayette: (1,101 Address Records)

The DOR data did not contain the same formatting as the EHDB tax identification numbers. DOR was updated to similar formatting. We were able to link 669 records using the modified tax numbers. However the DOR data has no addresses entered. Therefore we were not able to link up additional records.

Lake: (25,357 Address Records)

There are nearly 1,700 records with missing tax identification numbers in the EHDB. The property ID in the EHDB for nearly all of the populated fields is incorrect. The values entered are a combination of the S/T/R and alternate key. First, we linked 13,400 using the addresses alone. Using a subset of the Property ID and linking to the Tax Roll Alternate Key, we were able to link an additional 9,100 records – giving us an 89% success rate.

Lee: (63,099 Address Records)

There are 8,172 records in the EHDB without Tax Identification numbers. We were able to link 46,016 records using tax identification numbers. Another 14,147 records were linked using addresses providing a 95% success rate.

Leon: (13,445 Address Records)

There are 1,731 records in the EHDB without Tax Identification numbers. We were able to link 10,231 records using tax identification numbers. Another 1,772 records were linked using addresses providing an 89% success rate.

Levy: (9,285 Address Records)

There are 1,384 records in the EHDB without Tax Identification numbers or only S-T-R. We were able to link 3,931 records using tax identification numbers. Another 1,134 records were linked using addresses.

Liberty: (1,204 Address Records)

The EHDB contains no valid tax identification numbers. Most are blank, while the remaining are STR only. 770 of the records contain no street number for the permitted septic system. We were able to link only about 50 of the remaining 400 records.

Madison: (4,053 Address Records)

2,622 records have no street number – only street name. A large number of EHDB records contained “/” instead of “-” in tax identification numbers. The formatting was updated. Joining using Tax ID’s linked nearly 1,900 records while using Addresses another 200 linked.

Manatee: (7,124 Address Records)

Many of the tax identification numbers stored in the EHDB contained the “/” or “.” formatting (there are also 1,758 with null values). Those characters were removed for linking to the DOR tax roll which did not include the formatting. Linking by tax ID yielded over 2,600 records. Using the address, another 3,000 records were successfully linked – many of the unlinked are numbered roads without suffixes (i.e. 17 Street versus 17th Street).

Marion: (44,873 Address Records)

There are 2,666 records in the EHDB without tax identification numbers. We were able to successfully link 40,752 records using tax id – a very substantial percentage relative to other counties. Another 2,232 records were linked by addressing yielding a 96% success rate.

Martin: (10,323 Address Records)

The EHDB is lacking more than 7,500 tax identification numbers for the properties or may only contain the S-T-R. The DOR data did not have dashes as the EHDB had. They were updated to coincide with the formatting. Also, numerous records in the EHDB have incorrect length and are missing characters. We were only able to link about 100 records by tax id. After only being able to link 1,200 records by address, we determined that there were thousands of address records in the DOR data with multiple spaces between address parts. After scrubbing the data, we were able to link another 2,700 records.

Monroe: (8,263 Address Records)

The DOR data contains the preceding S-T-R information, while the EHDB does not. There are 1,423 EHDB records that do not have tax id’s. Utilized string concatenation queries to join portion of DOR tax roll data to the EHDB and was able to achieve on 6,091 records. Additionally, we were able to link 1,136 records by address giving us an 87% success rate.

Nassau: (10,417 Address Records)

Majority of EHDB records have valid tax identification numbers. Over 8,000 addresses joined based on tax ID alone. 948 records do not have a street number and do not join based upon Tax ID numbers. Using addresses, we were able to link an additional 440 parcels – providing an 82% success rate.

Okaloosa: (10,384 Address Records)

820 EHDB records have NULL tax identification numbers. While another 1,366 records are either spaces or only contain STR. Over 6,100 records linked utilizing tax identification numbers. Using the address as a link, another 2,600 records were successfully joined – providing an 84% success rate.

Okeechobee: (4,628 Address Records)

153 Records have no tax Id and the address is “Legacy”. Utilizing addresses, we were able to link 60%. Tax ID’s in the EHDB are inconsistent with formatting and do not match the length of the DOR. We were unable to determine proper formatting between the databases.

Orange: (33,938 Address Records)

17,210 Records in the EHDB do not have valid tax identification number – either null or not enough characters. Utilizing tax identification numbers, we were able to link slightly over 400 records. However, using addresses, we matched over 24,200.

Osceola: (9,434 Address Records)

The EHDB tax ID numbers contain formatting which the DOR tax roll did not have. The database was updated with formatting. Following the update, nearly 6,700 records were linked using the formatted tax identification numbers. After finding that linking by address yielded only a negligible amount of additional records, it was determined that the DOR tax roll contained multiple spaces between the street number and street name portions of the address. When corrected, an additional 2,000 records were linked giving us a 92% success rate.

Palm Beach: (22,688 Address Records)

Only 1,224 records in the EHDB do not have a tax identification number. Using the tax id’s to link records, we obtained 17,167 links. Another 3,045 records were linked using addresses which gave us an 89% success rate.

Pasco: (24,601 Address Records)

It was determined that the tax identification numbers in the DOR tax roll data were in the format of Range/Township/Section; while, the EHDB had them formatted as Section/Township/Range. We updated the DOR information to follow the EHDB convention and were able to successfully link 18,400 records. The physical addresses contained in the tax roll also contained numerous extraneous spaces. These were parsed before linking. Linking by address yielded an additional 4,000 records and giving us a 91% combined success rate.

Pinellas: (3,815 Address Records)

The formatting for the Tax ID in the EHDB, when populated, is very inconsistent along with the use of Township and Range Directions (not included in DOR information). Using tax id number only yielded 37 matches. Utilizing the addresses, resulted in an additional 2,700 matches.

Polk: (35,454 Address Records)

There are 4,119 records in the EHDB without Tax Identification numbers. Property ID formatting in the EHDB was inconsistent. We removed the dashes to provide for a link to the parcels. After reformatting, 10,133 records were successfully linked. Utilizing addresses, we were able to link an additional 13,831 records.

Putnam: (8,888 Address Records)

There are 1,132 records in the EHDB without Tax Identification numbers. The EHDB has dashes in the tax identification number while the DOR data does not. We formatted the parcels table with the dashes as shown in the tax ID's within the EHDB. After the modifications, we were able to link 7,094 records by tax ID. Another 1,022 were linked using addresses for a linked success rate of 91%.

Saint Johns: (20,948 Address Records)

Almost a quarter of the records do not have tax identification numbers. The tax id in the EHDB also contains a dash. We updated the parcels table to include similar formatting. Of those that have tax ids, we were able to link nearly all of them, 14,224 records. Another 3,993 records were linked using addresses for a success rate of 87%.

Saint Lucie: (10,119 Address Records)

There are over 9,100 records in the EHDB without tax identification number or contain incomplete information. Therefore, we were only able to link 490 records using this method. Another 2,891 records were linked utilizing the addresses. After reviewing the results, we determined that the DOR data contained multiple spaces between address parts for many records. We cleaned the data and were able to link an additional 1,957 records.

Santa Rosa: (24,953 Address Records)

There are 11,844 tax identification numbers in the EHDB that are invalid – blank, null or too short (S/T/R only). Therefore, only 7,800 records were linked using the tax identification number. We were able to join an additional 8,400 using addresses.

Sarasota: (8,014 Address Records)

There are about 2,600 tax identification numbers in the EHDB that are blank or invalid – STR only or incomplete. We were able to link 3,245 records using tax ids. While reviewing the DOR data, we identified that the county placed a dash (-) and an extra space between the street number and name. After cleaning, we were able to link another 3,373 records.

Seminole: (13,038 Address Records)

There are 3,532 tax identification numbers in the EHDB that are invalid – blank, null or too short (S/T/R only). Therefore, only 4,900 records were linked using the tax identification number. The DOR tax roll did not contain similar formatting as the EHDB tax id's. Therefore, they were updated before the link was able to occur. Additionally, 5,400 records were linked using addresses.

Suwannee: (6,651 Address Records)

There are 1,676 records in the EHDB with blank or invalid tax identification numbers – STR only or some unidentifiable format. Using tax identification numbers, we were able to link 1,292 records. The DOR information does not have site addresses and we could not locate an readily available alternative source. It should be noted that the EHDB does not contain the leading zero in the township portion of the tax id, while the county DOR data does. We updated the parcels table to coincide. There are other inconsistencies in the EHDB tax id's as well. It would take a significant amount of work to normalize or clean.

Taylor: (10,931 Address Records)

6,327 Records have no tax Id and the address is "Legacy". DOR did not contain formatting as shown in EHDB (/ and -). Copied DOR Tax ID to formatted field and added formatting to tax_id field for join. Removed the "/" in the EHDB parcels because it was inconsistently used versus the "-". The county also separates the street number from the street name in the DOR physical address fields. They were combined for the join.

Union: (1,878 Address Records)

Nearly one-quarter of the tax identification numbers in the EHDB are incomplete or blank. Of the ones that are populated, we were able to link 1,212 records. Only 8 additional records were linked using addresses. Many of the remaining unlinked records have incomplete tax ids and missing street numbers in the address.

Volusia: (34,863 Address Records)

The EHDB contained tax identification numbers with and without formatting (dashes in ID). Therefore, the dashes were removed for consistency and for creating the link to the tax roll. There are 2,481 EHDB records without Tax ID numbers. However, after inspection the tax identification numbers in the EHDB and tax roll are not similar. We could not determine how they were modified or recreate them from other information. Using the addresses, we were able to link over 19,000 records.

Wakulla: (5,284 Address Records)

389 Records have no tax Id and the address is "Legacy". Properly formatted and consistent tax identification numbers in the EHDB are lacking – many do not include jurisdictions or include punctuation. A large number of the unmatched addresses do not include street numbers. We were able to successfully link 56% of addresses. Another 17% were linked utilizing the tax identification numbers.

Appendix - State of the Data

Walton: (9,314 Address Records)

There are 2,839 EHDB records with missing or incomplete tax identification numbers. The DOR data did not have the dashes (formatting) that the EHDB records do. We updated the parcels table to reflect similar formatting. We were able to link 2,954 records using tax numbers. Another 2,033 records were linked using addresses.

Washington: (3,962 Address Records)

Most records do not have valid tax id's. They are just STR or blank. Created physical address field and utilized to join to the DOR data – had 50% success rate.

County Health Departments - Legacy (CHD)

Alachua:

23,403 Records were supplied in an Excel Spreadsheet. Of these records; 4,715 only have Permit # and no other information. The county also provided a list of all known sewer parcels, which we identified as being serviced by facility id “FLALACHUA.”

Bay:

Data was submitted but files do not open properly. There were numerous “doc” files which are not word documents.

Bradford:

The proprietary OSDS Database was provided. There were 841 permits none of which contained tax identification numbers. Many of the addresses provided are incomplete or not valid. Therefore only 102 records were joined.

Brevard:

Provided Excel Spreadsheet containing 28,485 Permits, almost all have Tax ID numbers – with multiple spaces. We parsed all of the spaces out of the tax id numbers and the extra spaces between the address parts. Following the database clean up, we linked 27,310 records using the tax id and addresses provided.

Calhoun:

The county provided an Excel Spreadsheet of 490 permits. The address field of the permits only includes road name, no numbers or other unique information.

Charlotte:

Provided Excel Spreadsheet of Septic Tank Permits granted each year. Also provided a list of approximately 55,000+ Short Legal Numbers for septic parcels, there are no other attributes.

Appendix - State of the Data

Clay:

Proprietary Database provided. We attempted to extract information; however we were unable to get anything from the database.

Dixie:

The county supplied an Excel Spreadsheet with 2,280 permits all having permit #'s. There are no tax identification numbers and the site addresses provided are locations, not specific addresses. No addresses were successfully linked.

Escambia:

The county provided Excel Spreadsheet of 11,610 Permits, almost all have Tax ID Numbers. Using the tax id's we were able to link over 10,000 records. Using the addresses created an additional 800 linked records.

Flagler:

The county provided an Excel Spreadsheet containing 5,593 permits; almost all have Tax ID numbers. Using the tax identification numbers to create the link, we were able to achieve 2,663 linked records. We were not able to link any records by address. Many of the records are void of addresses.

Gilchrist:

The county provided an Excel Spreadsheet consisting of 4,146 permits with permit #'s, 35% do not have addresses – none have tax identification numbers. The majority of those with address information populated are not valid addresses. They contain descriptions, locations and other non-identifying information. Therefore, we were only able to link 78 records.

Hamilton:

The county provided an Excel Spreadsheet with 1,554 permits; most of the permits have a location and not a valid address. There are no tax identification numbers for permits. There were no valid joins to the parcel table using address.

Indian River:

The county provided a proprietary database program which had numerous DBF files – amounting to 10,167 septic records. We utilized those files to get information. There were no tax id numbers and the addresses were street names. Where there are numbers, they are in parentheses at the end of the field. We could not make joins without very rigorous cleanup and database work.

Lafayette:

Supplied an Excel Spreadsheet consisting of 216 permits, all having permit numbers. Incomplete Site Address where people's names are there and no actual addresses.

Appendix - State of the Data

Lee:

The county provided an Excel Spreadsheet of 39,742 permits where most have either a Tax ID or Address. We were able to link about 9,000 records using the tax identification numbers; while another 11,700 were linked using addresses.

Leon:

A CSV file consisting of 26,777 permits was supplied by Leon County. Permits in the mid-1980's start having Tax ID numbers, however there are permits with no Tax ID and tend not to have Street Number with Location. Using Tax ID, slightly more than 17,000 records were successfully linked. The addresses only yielded 44 additional links, likely due to the absence of street numbers for most addresses.

Levy:

An Excel Spreadsheet containing 15,626 permits was provided by Levy County. The lot addresses are sparse; those that are populated are PO boxes or Rural Route Boxes. There are no tax identification numbers either. In all, we were able to link 1 record by address.

Marion:

An Access Database was supplied and would not open using various versions. Also provided an Excel spreadsheet with 21,232 Records; Tax ID Numbers are sparse and the address provided is mailing address – Location and Subdivision are not easily locatable. We were only able to link 794 records by tax identification number.

Martin:

The county supplied an Access Database with 263 permit numbers with addresses/general locations. There were no tax identification numbers provided. We were only able to link 31 records after cleaning up punctuation and other inconsistencies in the database provided.

Monroe:

The county provided 2 Excel Spreadsheets with 2,306 permits; included permit#, subdivision and owner and real estate number (one of their tax identification numbers). There are 1,826 records without tax id numbers. Therefore, we were only able to link 404 records.

Orange:

Supplied a DBASE1.exe file (86 kb) – too small to be a program or self-extracting executable. No other information provided

Osceola:

Supplied a 297 page PDF file of CENTRAX database reports – duplicate of EHD. No archived data provided.

Palm Beach:

The county provided a Microsoft Access database of 8,186 records. We computed the tax identification numbers from multiple columns. The numbered streets in the addresses do not contain suffixes (i.e. 24 versus 24th) as the DOR data does. We were able to link 1,333 records.

Sarasota:

Supplied two databases PERMSEPN & PERMSEPR; which contains 60,784 Permit #'s along with owner, tank size and installer. There are no addresses or tax identification numbers. We cannot link records simply by using owner name or other information provided.

St. Johns:

Provided 3 Excel Spreadsheets with the same information just sorted differently. There are a total of 10,959 permits. Files only contained date, permit number, owner, lot/block and subdivision. No method for automating location.

St. Lucie:

The county provided 2 Excel Spreadsheets for the 1980's and 1990's totaling 15,055 permits. No Tax ID's provided, but most have Addresses. We were able to link slightly more than 4,400 records.

Union:

We received a spreadsheet with 1,360 records. All of the records have tax identification numbers. We were able to link 1,224 records by tax identification number.

Walton:

We received a delimited text file from the county of permits. There were no tax identification numbers and the site addresses were mainly locations – absent street numbers and usually lists a road, intersection or subdivision. Therefore, we only were able to link about 150 records.