standardizedinventories-master folder

1. The setup.py was modified to include BeautifulSoup, argparse, selenium, and regex in the installation requirements.
2. A new file config.yaml was added. This file is useful to handle changes in a web site localization and/or the code source of a website is modified. In this moment only, there is information about RCRA, TRI, FRS, and SRS.
3. A new file common.py is used to handle config.yaml.

stewi folder

1. TRI.py was modified
2. The following packages were included:
   1. argparse: to write the inputs directly from Windows CMD
   2. beautifulsoup: to retrieve information directly from non-dynamic html
   3. regex: to handle regular expressions
   4. io: to deal with various type of input/output
3. The files TRI\_File\_1\_columns.txt and TRI\_File\_3a\_columns.txt were added in stewi/data folder. These include all the names of the columns for TRI Data Plus File 1 and File 3a. Maybe, for data since 2018 a new column will be needed to add (recycling a new condition of use included since 2018). I used TRI Fila 3a because it has information about off-site transfer. I could note that StEWI did not include information about the basis of estimate for off-site land treatment and other off-site land disposals.
4. TRI\_required\_fields.txt was modified to add new information
5. TRI\_keys.txt was added in stewi/data folder to include new TRI keys easier.
6. Files called TRI\_chem\_release\_**Year**.csv were added to include new validation source from <https://iaspub.epa.gov/triexplorer/tri_release.chemical>
7. To use TRI.py you can navigate to standardizedinventories-master (or stewi) in Windows CMD and write:

python “stewi/”TRI.py Option Year -F File1 File2 … FileN

Where Options are E, O, N: E is for extracting files from TRI Data Plus web site. N for organizing TRI National Totals files from TRI\_chem\_release\_**Year**.csv (this is expected to be download before and to be organized as it is described in TRI.py). O for organizing TRI as required by StEWI

E,g,. You want to use File 1 and File 3a of TRI Data Plus (as it is our case), retrieve information for TRI 2018. Therefore, you write in Windows CMD:

python “stewi/”TRI.py E 2018 -F 1 3a

After, if you want to create TRI\_2017\_NationalTotals.csv for validation:

python “stewi/”TRI.py O 2018

The flag -F and the files are not needed, but you need to have TRI\_chem\_release\_**2018**.csv in data folder.

Finally, you want to organize this for StEWI:

python “stewi/”TRI.py O 2018 -F 1 3a

Note: As you know it is possible that TRI include new columns. Therefore, you only need to include them in TRI\_File\_1\_columns.txt and TRI\_File\_3a\_columns.txt (or other files you want).

1. RCRAInfo.py was modified
2. Selenium package was used to handle dynamic html due to requests and urllib3 do not handle this.
3. To use RCRAInfo.py you can navigate to standardizedinventories-master (or stewi) in Windows CMD and write:

python “stewi/”RCRAInfo.py Option Year -T Table1 Table2 … TableN

Where Options are E, O, C: E is for extracting files from RCRAInfo web site. O for organizing Biennial Report for each year due to the current flat file has information of all year totally mixed. C for creating the files StEWI needs.

E,g,. You want to retrieve table BR\_REPORTING. Therefore, you write in Windows CMD:

python “stewi/” RCRAInfo.py E -T BR\_REPORTING

After, if you organize for each year the table BR\_REPORTING (This not take data for existing RCRAInfo report)

python “stewi/” RCRAInfo.py O -T BR\_REPORTING

Finally, you want to organize Biennial Report 2017 for StEWI

python “stewi/” RCRAInfo.py C 2017

The flag -T and the files are not needed, but you need to have RCRAInfo\_**2017**\_NationalTotals.csv in data folder.

1. The RCRAInfo National Totals were obtained from <https://rcrapublic.epa.gov/rcrainfoweb/action/modules/br/trends/view>
2. RCRA\_FlatFile\_LineComponents\_2019.csv was added due to changes in the specification of flat files <https://rcrainfo.epa.gov/rcrainfo-help/application/publicHelp/index.htm>
3. ValidationSets\_Sources.csv file was modified to include the source for TRI National Totals and RCRAInfo National Total. For TRI for year between 2001 and 2017 and RCRAInfo from 2001 to 2009 and 2017.
4. chromedriver.exe was added as the driver for selenium

chemicalmatcher folder

1. Some modifications in globals.py, programsynonymlookupbyCAS.py, and writeStEWIchemicalmatchesbyinventory.py were made.

facilitymatcher folder

1. Some modifications in globals.py, WriteFacilityMatchesforStEWI.py, and WriteFRSNAICSforStEWI.py were made.