**Supporting information**

**Data engineering for tracking chemicals and releases at industrial end-of-life activities**

Jose D. Hernandez-Betancur1, Gerardo J. Ruiz-Mercado2\*[[1]](#footnote-1), John P. Abraham2, Mariano Martin3, Raymond L. Smith2, Wesley W. Ingwersen2

1 Oak Ridge Institute for Science and Education, hosted by Office of Research & Development, U.S. Environmental Protection Agency, Cincinnati, OH, 45268, USA

2 U.S. Environmental Protection Agency, Cincinnati, OH, 45268, USA

Office of Research & Development, U.S. Environmental Protection Agency, Cincinnati, OH, 45268, USA

3 Department of Chemical Engineering, University of Salamanca, Plz. Caido 1-5, Salamanca, 37008, Spain

**S1. Acronyms used along the manuscript**

**Table S1.** List of acronyms used along the manuscript to abbreviate environmental regulations, tools, technical reports, and organizations.

|  |  |  |  |
| --- | --- | --- | --- |
| Analog Identification Methodology | AIM | National Primary Drinking Water Regulations | NPDWRs |
| Chemical Screening Tool for Exposures and Environmental Releases | ChemSTEER | Recycling, Energy recovery, Treatment & Disposal Facility | RETDF |
| Clean Air Act | CAA | Resource Conservation and Recovery Act | RCRA |
| Clean Water Act | CWA | Resource Conservation and Recovery Act Information | RCRAInfo |
| Computational Toxicology | CompTox | RETDF Industry Sector | RETDFiS |
| Condition of Use | CoU | Safe Drinking Water Act | SDWA |
| Contaminant Candidate List | CCL | Simplified Molecular Input Line Entry System | SMILES |
| Emergency Planning and Community Right-to-Know Act | EPCRA | Substance Registry Services | SRS |
| End-of-Life | EoL | Toxic Release Inventory | TRI |
| Environmental Compartment | EC | Toxic Substances Control Act | TSCA |
| Facility Registry Service | FRS | TSCA Chemical Data Reporting | TSCA CDR |
| Generator Industry Sector | GiS | TSCA New Chemical Program | TSCA NCP |
| Hazardous Air Pollutants | HAPs | U.S. Environmental Protection Agency | USEPA |
| Methylene Chloride | MC | Waste Management Hierarchy | WMH |

**S2. Data engineering framework**

Table S2 together with Figure S1 describe the collected data entry (i.e., EoU database features or columns) names, sources, data type, and the entities in the EoU supply and management chain. Please note that data symbol entries (first column, Table S2) are used to show their entity sources in Figure S1.

**Table S2**. Data entries collected from several sources (as shown in Figure S1) for reconciliation, harmonization, and built the EoL database. The 3rd and 6th columns describe the data type for each entry in the EoL database: ‘I’ for integer, ‘F’ for float type, and ‘C’ for character data (alphanumeric).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data symbol** | **Data name** | **Data Type** | **Data symbol** | **Data name** | **Data Type** |
|  | TRI reporting year1 | I |  | Reliability of the quantity transferred | F |
|  | Generator TRIF ID | C |  | EoL activity category based on TRI | C |
|  | Generator name | C |  | EoL activity category under TSCA previous reports [5] | C |
|  | Generator street | C |  | EoL activity category under waste management hierarchy [29] | C |
|  | Generator city | C |  | Receiver FRS ID2 | I |
|  | Generator county | C |  | Receiver TRIF ID2 | C |
|  | Generator state | C |  | Receiver RCRA ID | C |
|  | Generator zip | I |  | Receiver name | C |
|  | Generator primary NAICS code | I |  | Receiver street | C |
|  | Generator primary NAICS name | C |  | Receiver city | C |
|  | SRS chemical ID | I |  | Receiver county | C |
|  | TRI chemical ID | C |  | Receiver state | C |
|  | RCRA chemical ID | C |  | Receiver zip | I |
|  | CAS number | C |  | RETDF FRS ID2 | I |
|  | TRI chemical name | C |  | RETDF TRIF ID2 | C |
|  | SMILES | C |  | RETDF name | C |
|  | Chemical category 1 based on TSCA NCP3 | C |  | RETDF street | C |
|  | Chemical category 2 based on TSCA NCP3 | C |  | RETDF city | C |
|  | Chemical category 3 based on TSCA NCP3 | C |  | RETDF county | C |
|  | Is a HAP under the CAA?4 | C |  | RETDF state | C |
|  | Is it identified in biosolids by the CWA?5 | C |  | RETDF zip | I |
|  | Is a priority pollutant under the CWA?6 | C |  | RETDF primary NAICS code | I |
|  | Is it part of CCL under the SDWA?7 | C |  | RETDF primary NAICS name | C |
|  | Is it part of NPDWR under the SDWA?8 | C |  | Maximum amount of chemical present at RETDF | I |
|  | Is a chemical in the TSCA non-confidential inventory?9 | C |  | Total chemical generated as waste by RETDF | F |
|  | TRI classification10 | C |  | High-end scenario chemical flow | F |
|  | Metal indicator | C |  | Environmental compartments, EC-1-EC-4 | C |
|  | Generator condition of use based on TRI and matching TSCA CDR categories | C |  | RETDF chemical flow releases to the compartment *D57* 11 | F |
|  | Quantity transferred by generator | F |  | RETDF total chemical release11 | F |
|  | Unit of measurement | C |  |  |  |
| 1 is the year when the generator reported the off-site transfer in the TRI program. It is the starting point for the data engineering (see Figure S2).  2For traceability, if and , then the RETDF was the receiver, therefore, there is not a brokerage. On the other hand, when a transfer was to a waste broker and .  3For the starting point of hazard assessment, , , and are the categories in which a chemical may belong under TSCA NCP. However, when a chemical does not belong to any TSCA NCP category, then this is classified according to the TRI program, that means, .  4Link: <https://rb.gy/a97rbp> (access June 2, 2020)  5Link: <https://www.epa.gov/sites/production/files/2019-06/documents/2016-2017-biosolids-biennial-review.pdf> Appendix A, Table A-1(access June 2, 2020)  6Link: <https://rb.gy/d4zapo> (access June 2, 2020)  7Link: <https://rb.gy/xkoxx0> (access June 2, 2020)  8Link: <https://rb.gy/jc6u59> (access June 2, 2020)  9Link: <https://rb.gy/ej09ah> (access June 2, 2020)  10 TRI (general EPCRA Section 313 chemical), Persistent Bioaccumulative and Toxic, and Dioxin or Dioxin-like compound.  11  is the total emission to each environmental compartment that was reported by the RETDF to the TRI (see Section 3.3) instead of being the indirect release of a chemical that may occur after it is transferred to an off-site facility for EoL management. Moreover, is the sum of all releases to each compartment from the RETDF. | | | | | |

A screen shot of a computer

Description automatically generated

**Figure S1.** Schematic explanation about the relationship between the entities in the EoL supply and management chain and data entries and their source after refinement, harmonization, and building of the EoL database.

The data engineering process to transform the data into structures for the framework is depicted as sequential steps in Figure S2 and Figure S3. Figure S2 presents the data engineering process used to retrieve, clean, filter, and transform information from TRI considering waste brokering, while Figure S3 describes a similar approach for RCRAInfo. The following step-by-step instructions describe the proposed approach to build the EoL database and correspond to those processing steps shown in Figure S2 and Figure S3, respectively. Notice that Figure 4 is a simplified version of data engineering and refinement process after combining Figure S2 and Figure S3 for transforming the TRI database into structures for the EoL database and tracking chemical flows at EoL stages.

***Data engineering and refinement process of TRI, Figure S2:***

1. A TRI reporting year () is selected. For this work, TRI reporting year 2017 is used, which is the most recent data year available online at the time of study. The documents that described the content of the data reported to the TRI Program can be found in <https://rb.gy/tcao4t> (last access May 15, 2020). Note that the database that generates these files is updated 4-6 times per year. Thus, analysts might not be able to reproduce the case study with the same data source.
2. Generator TRIF ID, location, and primary NAICS code (), TRI chemical ID (), TRI chemical name (), TRI classification (), and generator condition of use () are retrieved from TRI File 1a and TRI File 1b (*Release and Other Waste Management* and *Chemical Activities and Uses*, respectively) . In contrast, quantity transferred off-site (), EoL activity category based on TRI (), receiver name and location (), and in some cases receiver RCRA ID (, when it has the correct structure of the ID assigned by USEPA to the physical location of a hazardous RETDF), are gathered from TRI File 3a (*Off-site Transfers*).
3. If belonging to TSCA CDR industrial CoU, the receiver is located in the U.S., and is an EoL activity of interest, the records are selected to form a partial dataset called TRI dataset 1.
4. As shown by the red diamond 1, or are used to access to FRS. When has the correct reporting structure, this is used to retrieve receiver FRS ID () and subsequently receiver TRIF ID (). Otherwise, are compared to find whether or not these are similar to any facility location recorded in the FRS, and after gathering , , and .
5. To be able to continue the tracking, records are selected when the receivers have reported at least to TRI or RCRAInfo ( and are not null) to form TRI dataset 2.
6. As shown by the red diamond 2, is used to connect SRS and then obtain SRS chemical ID () and CAS number (). After that, as indicated by the red diamond 3, is used to confirm using whether or not an TRI chemical belongs to one of the lists of substances of concern named in Section 3.1 (e.g., RCRA listed hazardous waste). Thus, TRI dataset 3 is obtained.
7. According to the red diamond 4, is used to retrieve the SMILES () of each TRI chemical from CompTox. After that, as presented by the red diamond 5, SMILES of each chemical is the input so that AIM tool, predicts whether a chemical may belong to any TSCA NCP categories (). Thus, TRI dataset 4 is built.
8. As shown by the red diamond 6, , , , , and are used to merge TRI database 4 with RCRA database 3 (resulting as explained below in *Data engineering and refinement process of RCRAInfo*). Thus, obtaining RETDF location and primary NAICS code (), maximum amount of chemical at RETDF (), total chemical generated as waste at by RETDF (), and chemical flow released to a compartment from RETDF and sum of releases to all compartments (), which are data entry group 3 in Figure S1. Moreover, the next sub-steps shall be followed:
   1. Brokering case: If and  are non-null, these are used to check whether the TRI chemical may be considered a listed hazardous waste and the receiver has reported to RCRAInfo. Then, , and are used to access to RCRA dataset 3. If several data records satisfy the search for , , and , then the one closest to the year 2017 is selected, as far as it satisfies the consistency analysis explained in Section 3.3.
   2. No brokering case: If the previous search is not successfully completed and the record has a non-null (i.e., receiver has reported to TRI), it goes through a recursive search from TRI reporting year 2001 to 2017. The same applies when a receiver has not reported to RCRAInfo (null ), the TRI chemical is not listed hazardous waste (null ), or both. For this recursive search, , , and (for this case ) are used. If various records are found having the same , , and (or ), then the record closest to the year 2017 is selected.
9. Finally, the maximum possible flow of chemical transferred (), is computed as described in Section 3.3. As was explained before, is the maximum amount of chemical transferred that could end up as emissions at an off-site facility during its EoL. As explained in Section 3.3, it is given by the minimum flow found by tracking a chemical using the procedures described in Figure S2 and Figure S3.

A screenshot of a cell phone

Description automatically generated

***Figure S2.*** *Data engineering and refinement process for transforming the TRI database into structures for the EoL database and tracking chemical flows at EoL stages. RCRAInfo data engineering process is presented in Figure S3.*

***Data engineering and refinement process of RCRAInfo, Figure S3:***

1. A reporting year of the RCRA report is selected as a starting point. However, the procedure in Figure S3 is done for each RCRA biennial report from 2001 to 2017 (odd years). The idea is to try to find an off-site facility (broker(s), receiver, and RETDF) in the information stored to date, such that it is possible to track a chemical flow.
2. Using the amount of hazardous waste received by a facility and the reported management method code (at the red diamond 1 stage), the Nonbrokers dataset 1 and Brokers dataset 1 are built. Therefore, If the record has an amount received other than 0, this goes to Nonbrokers dataset 1, unless it has a management method code equal to H141, which means that a facility receives the waste only for storage and then transfer, therefore, it would go to Brokers dataset 1.
3. With the information collected at the red diamond 2 stage, the Brokers dataset 1 passes through the grey block composed of 4 stages, obtaining a cleaner data set called Brokers dataset 2 by:
   1. Searching the quantity of waste containing the chemical of concern shipped by a broker if source code is G61, that means, the waste was received from off-site for storage and then transfer. Here, information about “to where” and “for what” the waste was shipped is identified as well.
   2. Knowing the management activity for which a hazardous waste was shipped (e.g., biological treatment) and where it was transferred, the off-site facility which received the hazardous waste is searched in the information stored in RCRAInfo to date.
   3. If the searching is successful, a flow consistency analysis is done to ensure the flow shipped by a broker is less than the flow received by the off-site facility
   4. If several records are found having the same off-site facility and management activity, it is selected the record closest to the reporting year of the RCRA report, which was selected in step 1.
   5. Step 3 is done until the RETDF is found.
4. Broker dataset 2 and Nonbrokers dataset 1 are combined.
5. After that, RCRA IDs of facilities, at the red diamond 3, are used to find their FRS IDs. Subsequently, as shown the red diamond 4, these are used to check if a facility has reported to the TRI program at any time. Thus, if a facility satisfies the above, a record is kept in RCRA dataset 1.
6. As shown at the red diamond 5, RCRA IDs of hazardous wastes are used to retrieve the SRS IDs. They are then used to verify if hazardous waste is a TRI chemical, which means, if it has a TRIF ID. Thus, RCRA dataset 2 is generated at this step.
7. Finally, the TRIF ID of RETDFs and hazardous wastes are used to connect with TRI reports (the red diamond 7). Then, information about data entry group 9 in Figure S1 tried to be found in the existing and available TRI reports to date. If the search is successful, the record is kept; otherwise, it is discarded. Subsequently, RCRA dataset 3 is obtained and can be used by step 8.1 of the *Data engineering and refinement process of TRI.*

Note that both processes shown in Figure S2 and Figure S3 can be done in parallel, at least until the process presented in Figure S2, requires RCRA dataset 3. It is essential to clarify that the recursive search shown in Figure S2 for all existing and available TRI reports to date, is independent of the data engineering process presented in Figure S3 for RCRAInfo, making it possible to address both procedures in parallel.

A screenshot of a cell phone

Description automatically generated

**Figure S3.** *Data engineering and refinement process for transforming the RCRAInfo database into structures for the EoL database and tracking chemicals in industrial wastes. Management method code, H141, “describe the type of hazardous waste management system used to treat, recover, or dispose of a hazardous waste”. The source code, G61, “describe the type of process or activity (i.e. source) where a hazardous waste was generated”.*

**S3. Development of the Equation 1**

The framework assumes that a RETDF receives a chemical flow for (i) energy recovery, recycling, treatment (physical/chemical/thermal/biological), or disposal, and (ii) the recycled chemical is used as a product or incorporated in a valuable product. Additionally, as shown in Figure S4(A), the RETDF may also generate wastes containing a chemical. The total waste having the chemical can be recycled, treated, transferred, released, disposed of, used for recovering energy. Thus, Equation S1 presents the annual mass balance for the chemical. is the annual change of the amount of chemical present at the RETDF, which can be calculated as using the amount of chemical present at the start of the year () and at the end of the year ().

|  |  |
| --- | --- |
|  | (S1) |

As depicted in Figure S4(B), the amount of chemical present or accumulated at the RETDF can change over the reporting year. Thus, if is the range code representing the maximum amount of chemical present at RETDF at any moment over the reporting year and is a value that belongs to , and would be and and even they can be . Thus, must be estimated to calculate the emission factor for each record . Also, two boundary-value cases are:

1. Figure S5(A) shows a first case where and with . If the highest value for is taken from , . This case represents the lowest possible value that can take.
2. Figure S5(B) exposes a second case where and . Similarly, taking the highest value for , . This value is the highest possible value for .

Due to the above boundary-value cases, has a constraint such . However, this is the first of three constraints for , as shown in Figure S5(C). The other two constrains for are obtained as:

**A close up of a logo

Description automatically generated**

**Figure S4.** (A) Summary of the elements included in the annual mass balance for a chemical at a RETDF based on the assumptions presented in Section 3.3. (B) Hypothetical curve representing the amount of a chemical present at a RETDF over the course of a reporting year as time function.

1. The term cannot be negative and the denominator of the emission factor must not be to avoid indetermination. Thus, using the Equation S1, a second constraint for is obtained, . i.e., , which is depicted in Figure S5(C)
2. Additionally, the sum of the emission factors must be a value between 0 to 1. As Figure 5 shows, the framework considers 4 compartments: indoor air (EC-1), on-site soil (EC-2), on-site surface water (EC-3), and outdoor air (EC-4). Due to this criterion in the sum of the emission factors, a third constraint can be obtained for . Thus, developing the Equation S2 results in the constraint . However, considering the second constraint, the third constraint for can be expressed as , i.e. , as presented in Figure S5(C).

|  |  |
| --- | --- |
|  | (S2) |
|  |
|  |
|  |
|  |

Figure S5(C) shows the intersection of the three intervals which represent the three constraints for . Thus, from the discontinued vertical purple line on the right, is the upper value for the intersection. However, as indicated by the green curly bracket, the lower value for the intersection depends on the largest value between and , i.e., Thus, the domain for is . Thus, for a record , has several values in that interval so that may be represented by a random number.

A screenshot of a video game

Description automatically generated

**Figure S5.** (A) Hypothetical case with the highest value for . (B) Hypothetical case with the lowest value for . (C) Intersection of the intervals obtained by applying the constraints.

**S4. Tables with labels for the case study: Tracking of MC flows at the EoL stage at off-site locations (transfers)**

**Table S3.** NAICSclassification for the generator industry sector (GiS) used in the chemical flow analysis shown in Figure 6.

|  |  |
| --- | --- |
| **Name** | **Label** |
| Adhesive manufacturing | GiS-1 |
| All other basic organic chemical manufacturing | GiS-2 |
| All other miscellaneous chemical product and preparation manufacturing | GiS-3 |
| Ammunition (except small arms) manufacturing | GiS-4 |
| Artificial and synthetic fibers and filaments manufacturing | GiS-5 |
| Cement manufacturing | GiS-6 |
| Copper rolling, drawing, extruding, and alloying | GiS-7 |
| Ethyl alcohol manufacturing | GiS-8 |
| Guided missile and space vehicle propulsion unit and propulsion unit parts manufacturing | GiS-9 |
| Hazardous waste treatment and disposal | GiS-10 |
| In-vitro diagnostic substance manufacturing | GiS-11 |
| Industrial gas manufacturing | GiS-12 |
| Irradiation apparatus manufacturing | GiS-13 |
| Materials recovery facilities | GiS-14 |
| Medicinal and botanical manufacturing | GiS-15 |
| Metal coating, engraving (except jewelry and silverware), and allied services to manufacturers | GiS-16 |
| Metal crown, closure, and other metal stamping (except automotive) | GiS-17 |
| National security | GiS-18 |
| Other basic inorganic chemical manufacturing | GiS-19 |
| Other chemical and allied products merchant wholesalers | GiS-20 |
| Paint and coating manufacturing | GiS-21 |
| Pesticide and other agricultural chemical manufacturing | GiS-22 |
| Petrochemical manufacturing | GiS-23 |
| Petroleum lubricating oil and grease manufacturing | GiS-24 |
| Pharmaceutical preparation manufacturing | GiS-25 |
| Photographic and photocopying equipment manufacturing | GiS-26 |
| Plastics material and resin manufacturing | GiS-27 |
| Precision turned product manufacturing | GiS-28 |
| Special die and tool, die set, jig, and fixture manufacturing | GiS-29 |
| Sporting and athletic goods manufacturing | GiS-30 |
| Surgical and medical instrument manufacturing | GiS-31 |

**Table S4.** Chemical condition of use (CoU) categories (at generator facility) employed in the chemical flow analysis shown in Figure 6. Also, Table S5 describes the corresponding TRI activities and uses of chemicals (TRIU) combinations.

|  |  |
| --- | --- |
| **Label1** | **Combination** |
| CoU-1 | TRIU-1 |
| CoU-2 | TRIU-1 + TRIU-2 |
| CoU-3 | TRIU-1 + TRIU-3 + TRIU-4 |
| CoU-4 | TRIU-1 + TRIU-3 + TRIU-2 |
| CoU-5 | TRIU-1 + TRIU-3 + TRIU-2 + TRIU-4 |
| CoU-6 | TRIU-4 |
| CoU-7 | TRIU-5 |
| CoU-8 | TRIU-5 + TRIU-3 |
| CoU-9 | TRIU-5 + TRIU-6 + TRIU-1 + TRIU-7 + TRIU-2 + TRIU-8 + TRIU-4 |
| CoU-10 | TRIU-3 + TRIU-2 + TRIU-4 |
| CoU-11 | TRIU-2 |
| CoU-12 | TRIU-2 + TRIU-8 |
| CoU-13 | TRIU-2 + TRIU-8 + TRIU-4 |
| CoU-14 | TRIU-6 + TRIU-1 + TRIU-3 + TRIU-2 |
| CoU-15 | TRIU-6 + TRIU-2 + TRIU-4 |
| CoU-16 | TRIU-1 + TRIU-4 |
| CoU-17 | TRIU-5 + TRIU-3 + TRIU-7 |
| CoU-18 | TRIU-6 + TRIU-2 |
| CoU-19 | TRIU-8 |
| CoU-20 | TRIU-3 + TRIU-7 + TRIU-4 |
| CoU-21 | TRIU-3 |
| CoU-22 | TRIU-7 |
| CoU-23 | TRIU-7 + TRIU-4 |
| CoU-24 | TRIU-5 + TRIU-1 |
| CoU-25 | TRIU-5 + TRIU-9 + TRIU-3 + TRIU-7 + TRIU-4 |
| CoU-26 | TRIU-5 + TRIU-3 + TRIU-4 |
| CoU-27 | TRIU-3 + TRIU-4 |
| CoU-28 | TRIU-9 + TRIU-1 + TRIU-4 |
| CoU-29 | TRIU-5 + TRIU-2 + TRIU-4 |
| CoU-30 | TRIU-2 + TRIU-4 |
| CoU-31 | TRIU-6 + TRIU-1 |
| CoU-32 | TRIU-1 + TRIU-3 |
| CoU-33 | TRIU-5 + TRIU-6 + TRIU-1 |
| CoU-34 | TRIU-6 |
| CoU-35 | TRIU-6 + TRIU-7 + TRIU-2 + TRIU-8 + TRIU-4 |
| CoU-36 | TRIU-5 + TRIU-9 + TRIU-3 + TRIU-2 + TRIU-8 + TRIU-4 |
| CoU-37 | TRIU-1 + TRIU-7 + TRIU-2 |
| CoU-38 | TRIU-5 + TRIU-1 + TRIU-2 + TRIU-4 |
| **1** In the TRI program, facilities do not report quantities of chemical generated at each activity, instead of that, they submit a form with all the aggregated quantities. | |

**Table S5.** TRI activities and uses of chemicals (TRIU) for defining the CoU categories described in Table S4.

|  |  |
| --- | --- |
| **Name** | **Label** |
| Added as a formulation component | TRIU-1 |
| Used as a chemical processing aid | TRIU-2 |
| Repackaging | TRIU-3 |
| Ancillary or other use | TRIU-4 |
| Produce the chemical | TRIU-5 |
| Used as a reactant | TRIU-6 |
| As a process impurity | TRIU-7 |
| Used as a manufacturing aid | TRIU-8 |
| Import the chemical | TRIU-9 |
| Used as an article component | TRIU-10 |

**Table S6.** NAICSClassification for the RETDF industry sector (RETDFiS) and used in the chemical flow analysis shown in Figure 6.

|  |  |
| --- | --- |
| **Name** | **Label** |
| All other miscellaneous chemical product and preparation manufacturing | RETDFiS-1 |
| Hazardous waste treatment and disposal | RETDFiS -2 |
| Cement manufacturing | RETDFiS -3 |
| Materials recovery facilities | RETDFiS -4 |
| Solid waste combustors and incinerators | RETDFiS -5 |
| Petroleum lubricating oil and grease manufacturing | RETDFiS -6 |
| All other basic organic chemical manufacturing | RETDFiS -7 |
| Photographic film, paper, plate, and chemical manufacturing | RETDFiS -8 |
| Pesticide and other agricultural chemical manufacturing | RETDFiS -9 |

**Table S7.** Classification of EoL activities used in the chemical flow analysis shown in Figure 6. The 1st column is the categorization used by TRI program for representing off-site transfers. Notice that EoL management by combustion is captured in both energy recovery and treatment via incineration. Also, this table shows the resulting TRI classification categories after applying the EoL data engineering framework for the MC case study. Other chemical substances might have other resulting categories from the full set of TRI classification categories <https://rb.gy/d2hwez> (access June 2, 2020).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TRI classification1** | **TSCA classification2** | **Label** | **USEPA WMH3** | **Label** |
| Solvents/organics recovery | Recycling4 | EoL-1 | Recycling4 | WMH-1 |
| Other reuse or recovery |
| Energy recovery | Energy recovery4 | EoL-2 | Energy recovery4 | WMH-2 |
| Incineration/insignificant fuel value5 | Incinerators | EoL-3 | Treatment | WMH-3 |
| Incineration/thermal treatment5 |
| Wastewater treatment (excluding POTWs) – nonmetals | Industrial wastewater treatment | EoL-4 |
| Solidification/stabilization – treatment - nonmetals | Other treatment | EoL-5 |
| Other waste treatment |
| RCRA Subtitle C landfills | Landfill (municipal, hazardous, or other land disposal) | EoL-6 | Disposal | WMH-4 |
| Underground injection class I wells | Underground injection | EoL-7 |
| 1 These correspond to  2 These correspond to .  3 These correspond to .  4 These classifications are part of , , and .  5 For a facility “to claim that a reported EPCRA Section 313 chemical sent off-site is used for the purpose of energy recovery and not for treatment for destruction, the EPCRA Section 313 chemical must have a significant heating value and must be combusted in an energy recovery unit such as an industrial boiler, furnace, or kiln”. Thus, “incineration/insignificant fuel value” category represents that the chemical goes into a “legitimate energy recovery unit”, but it does not “contribute to the heating value of the waste.” | | | | |

**Table S8.** Environmental compartment (EC) labels for the), the data engineering approach and shown in Figure 6.

|  |  |
| --- | --- |
| **Name** | **Label** |
| 1Indoor air | EC 1 |
| 1On-site soil | EC 2 |
| 1On-site surface water | EC 3 |
| 1Outdoor air | EC 4 |
| Net recycling | EC 5 |
| Net energy recovery | EC 6 |
| Net treatment | EC 7 |
| Net disposal | EC 8 |
| 1 These are the environmental compartments which are part of in the EoL database. | |

**S5. Data obtained from TRI explorer**

Based on <https://rb.gy/ot4mph> (access May 15, 2020), the TRI off-site transfers for further management of waste containing MC was gathered and summarized in Table S9 using percentages. That information was obtained by the TRI administrators by means of 257 facilities in all industries, for MC in the U.S., 2017”. In addition, from the EoL database, 259 pathways for MC were obtained after tracking using the data engineering approach, which were used to calculate the values in the 3rd column of the Table S9. Using chi-square hypothesis test, with 3-degree freedom and significance level 0.05, the critical value is 7.81 and the test statistic 1.9011, that means, it is not rejected the hypothesis that the obtained material distribution using the EoL database records are similar to those obtained using the information available online from the TRI Explorer tool since 1.9011 < 7.81.

**Table S9.** MC percentage distribution for each type of off-site waste management using the pathways found using the data engineering approach and the information from TRI administrators for.

|  |  |  |  |
| --- | --- | --- | --- |
| **Management type** | **%** | |  |
| **From TRI** | **From EoL** |
| Transfers to energy recovery | 19.016 | 18.215 | 0.0337 |
| Transfer to recycling | 36.720 | 42.779 | 0.9998 |
| Transfers to disposal | 0.348 | 0.017 | 0.3148 |
| Transfers to treatment | 43.916 | 38.989 | 0.5528 |
| Total Transfers off-site for further waste management | 100.000 | 100.000 | 1.9011 |

**S6. Python script written for the MC case study**

The Python code was uploaded to the following GitHub repository

Link: <https://github.com/jodhernandezbe/MC_Case_Study>

1. \* Corresponding author:

   E-mail: ruiz-mercado.gerardo@epa.gov

   Address: 26W Martin L. King Dr., Cincinnati, OH, 45268, USA [↑](#footnote-ref-1)