A DATA ENGINEERING APPROACH FOR TRACKING INDUSTRIAL CHEMICAL WASTES IN END-OF-LIFE SCENARIOS

Jose D. Hernandez-Betancur[1], Gerardo J. Ruiz-Mercado[2*], John P. Abraham[2]
[1] ORISE Research Fellow, Office of Research and Development, National Risk Management Research Laboratory, U.S. Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, Ohio 45268, USA
[2] Office of Research and Development, National Risk Management Research Laboratory, U.S. Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, Ohio 45268, USA

Abstract Overview

A chemical risk evaluation process is burdened by apparently time-consuming data requirements. These requirements include: the conditions of use; quantities of a chemical to which a receptor - either human or the environment - is exposed; the exposure pathway and route; the most susceptible receptors; and the hazards of a chemical. Such data requirements are even more difficult to determine at the end-of-life (EoL) of a chemical because the uncertainty in the disposition of the compound is greater. In this work, a data engineering approach is used to extract information from U.S. Environmental Protection Agency's (EPA) structured and non-structured data to build a dataset which will be useful for tracking a chemical into its probable EoL final waste stream(s). Different EoL management scenarios may result in potential significant releases of the chemical, and therefore relevant to a more complete Toxic Substance Control Act (TSCA) risk evaluation. This contribution aims the facilitation of chemical risk evaluation under TSCA needs.

Keywords

Data engineering, Data refinement, EPA databases, End-of-life management scenarios, Chemical risk evaluation

Introduction

The Toxic Substance Control Act (TSCA) contains a Chemical Substance Inventory (CSI) which lists the chemicals subject to TSCA regulations. This inventory is constantly growing due to the development of new chemicals for the U.S. market. Currently, the TSCA CSI lists over 85,000 chemicals (U.S Environmental Protection Agency, 2015). Because of the size of the CSI, TSCA came up with a process to prioritize those chemical substances for which further risk evaluation is warranted (U.S. Environmental Protection Agency, 2017). In 2016, EPA selected the first group of 10 high-priority chemicals for risk evaluation. However, the EPA problem formulation documents for those chemicals did not analyze all potential pathways associated with their EoL. (U.S. Environmental Protection Agency, 2017) because of some pathways have well-established assessment procedures under other (non-TSCA) EPA statutes such as the Clean Air Act (CAA), Safe Drinking Water Act (SDWA), and so on. Nonetheless, other pathways associated with recycling, reuse, recovery, and other treatments, and having high-epistemic-uncertainty risk evaluation procedures, were not further analyzed in the TSCA chemical risk evaluation, but the possibility of additional exposure was highlighted. Therefore, there is an opportunity to develop an approach for the rapid estimation of releases in EoL scenarios and determine if the risks from these TSCA chemicals of interest are significant at EoL and therefore relevant to a more complete TSCA risk evaluation. These chemicals are previously generated under

* Corresponding author: Ruiz-Mercado.Gerardo@epa.gov

industrial conditions of use and subsequently transferred for the above EoL management scenarios at an off-site facility.

EPA has valuable information on its existing databases, which contain historical information recorded by different EPA programs. However, the main challenge is that for obtaining such information, different databases should be merged, and selected the useful records based on data refinement criteria. Consequently, to deal with EPA's non-explicit, and both structured and non-structured databases, in this work is proposed a systematic approach based on data engineering and data refinement for tracking industrial chemicals in EoL scenarios.

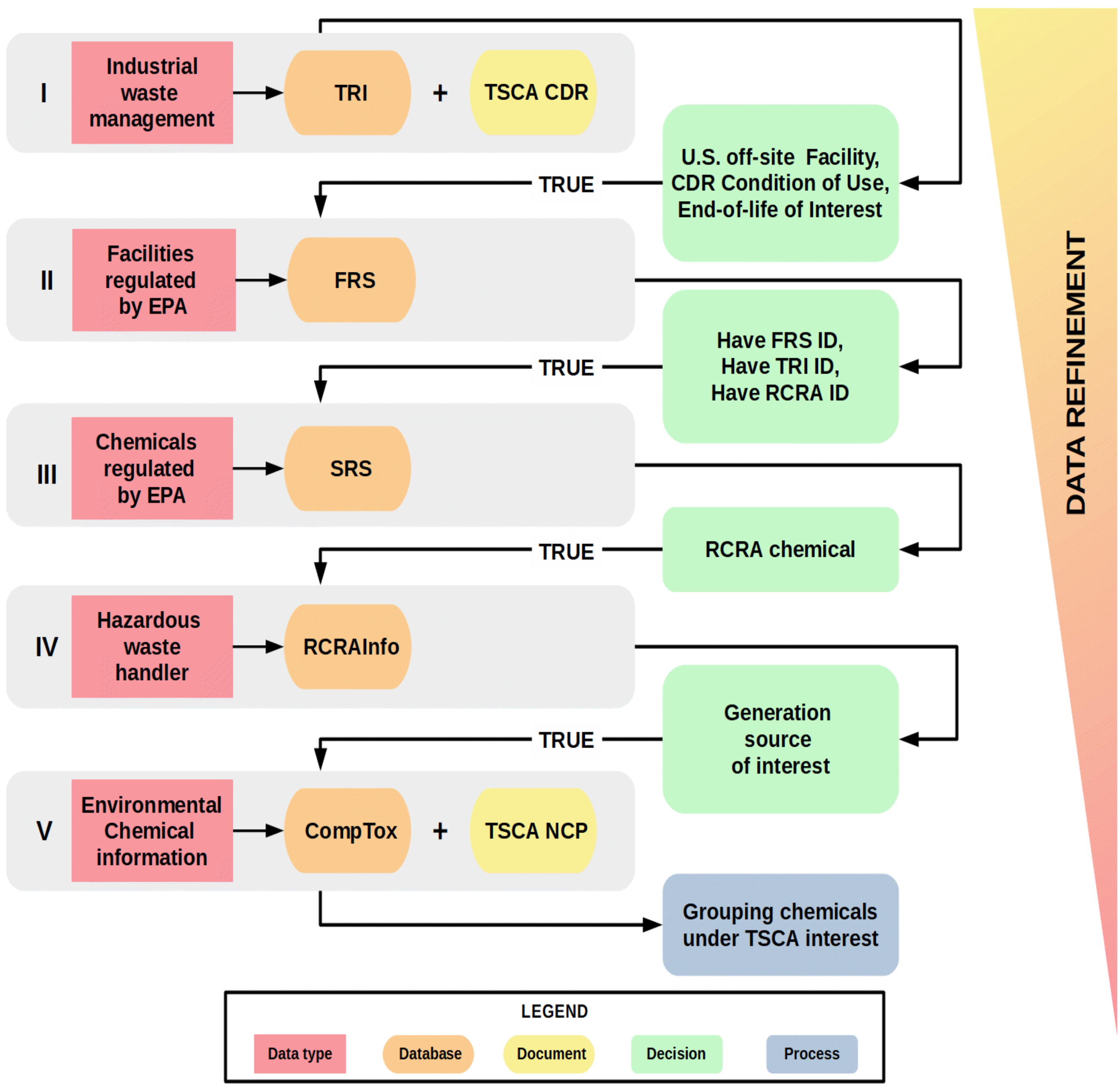

Figure 1. Data engineering and data refinement flowsheet for end-of-life scenario

Methodology

In this work, many computational codes are developed for retrieving information from EPA's Toxic Release Inventory (TRI), the Resource Conservation and Recovery Act Information (RCRAInfo), the Facility Registry System (FRS), the Substance Registry System (SRS), and the CompTox Chemistry Dashboard. This is complemented by using the industrial Condition of Use from the TSCA Chemical Data Reporting (CDR), which categorizes and groups industrial applications of chemicals. In addition, to obtain more valuable information for TSCA risk evaluation, the chemicals are classified according to the TSCA chemical categories established in the TSCA New Chemicals Program (TSCA NCP).

A general outline of the data refinement is shown in Figure 1. Five data wrangling stages support the data acquisition, analysis, and engineering procedures, where at each stage, some records from EPA's databases go through a consolidation filter. At each stage, some databases and supporting documents are used in conjunction with several decision-making criteria for EoL evaluation. For instance, at the first stage, the TRI database and the TSCA CDR are used to retrieve information regarding industrial waste management by considering aspects about the Industrial Condition of Use which generates the chemical waste, and the subsequent EoL scenario at a U.S. off-site facility. Therefore, with the general outline in Figure 1, the most useful records would be organized and consolidated in ways for constituting a TSCA-EoL scenario database.

Conclusions

The constructed database will be useful for a subsequent use in Data Analytics to perform descriptive statistics and to extract meaningful insights for EoL chemical risk, as well as in Machine Learning Engineering to come up with algorithms to forecast and do a learning-from-data generalization for existing and new chemicals at U.S. market not being reported in EPA databases. This contribution aims the facilitation of chemical risk evaluation under TSCA needs.

Disclaimer

The views expressed in this extended abstract are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Acknowledgement

This research was supported in part by an appointment of Jose D. Hernandez-Betancur to the Postmasters Research Program at the National Risk Management Research Laboratory, Office of Research and Development, U.S. EPA, administered by the Oak Ridge Institute for Science and Education through Interagency Agreement No. DW-89-92433001 between the U.S. Department of Energy and the U.S. EPA.

References

U.S Environmental Protection Agency. (15 de September de 2015). *About the TSCA Chemical Substance Inventory*. Obtenido de https://www.epa.gov/tsca-inventory/about-tsca-chemical-substance-inventory

U.S. Environmental Protection Agency . (30 de June de 2017). *Assessing and Managing Chemicals under TSCA*. Obtenido de Risk Evaluations for Existing Chemicals under TSCA: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/risk-evaluations-existing-chemicals-under-tsca

U.S. Environmental Protection Agency. (19 de December de 2016). *Designation of Ten Chemical Substances for Initial Risk Evaluations Under the Toxic Substances Control Act*. Obtenido de https://www.federalregister.gov/documents/2

U.S. Environmental Protection Agency. (20 de June de 2017). *Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act.* Obtenido de https://www.regulations.gov/document?D=EPA-HQ-OPPT-2016-0654-0108