A perspective in data and software systems to streamline the transition towards a safer chemical end-of-life management

Jose D. Hernandez-Betancur[1,2*], Gerardo J. Ruiz-Mercado[3,4]

[1] Faculty of Mines, National University of Colombia, Medellin, 050041, Colombia
[2] Department of Chemical Engineering, University of Salamanca, Plz. Caidos 1-5, Salamanca, 37008, Spain
[3] Office of Research & Development, U.S. Environmental Protection Agency, Cincinnati, OH, 45268, USA
[4] Chemical Engineering Graduate Program, University of Atlántico, Puerto Colombia, 080007, Colombia

Abstract

Chemicals serve pivotal functions in many commercial and consumer products. To manage chemicals and their impact on the environment, chemical risk assessment (CRA) and material flow analysis (MFA) are employed. However, challenges arise in accessing data, particularly in the end-of-life (EoL) stage of products. This perspective manuscript explores how software and data systems can facilitate CRA and MFA at the EoL stage. This contribution reviews regulatory data sources like the Pollutant Release and Transfer Registers, information extraction from academic data via natural language processing, and real-time data to improve understanding of the EoL supply and management chain. Additionally, the manuscript discusses the application of graph neural networks and transfer learning techniques to improve the representation and performance of EoL supply chain models.

[*] Corresponding author:

Email address: jodhernandezbe@unal.edu.co (J.D. Hernandez-Betancur)

1. Introduction

Chemicals are essential components of commercial products like batteries and industrial lubricants. However, certain uses of hazardous chemicals can pose an unreasonable risk to human health and the environment through their entire life cycle. Chemical risk assessment (CRA) is a method used to make well-informed decisions, choose compounds with safer characteristics, and develop and implement strategies to eliminate or decrease chemical risks[1].

Material flow analysis (MFA) quantifies and allocates material movement (e.g., chemicals) in production systems, releases, and ecosystems[2]. In CRA, MFA helps assess chemical risk and exposure. It assists in determining the receptors (e.g., workers) that may be exposed to a chemical in the work environment and the quantity of releases into the environment. MFA also identifies scenarios that may result in human and environmental exposure[3].

The data acquisition for MFA and CRA historically has been difficult due to data quality and accessibility[4, 5]. The growing amount of chemicals manufactured and brought into the worldwide market and the global integration of the chemical supply chain make this particularly concerning[6]. End-of-life (EoL) management is complex due to uncertainty about the quantity and pathway taken by a chemical and a dearth of data to assess chemical exposure which is why CRA sensitivity assessments are generally insufficient[7–9].

Artificial intelligence (AI) and information technology (IT) systems transformed the chemical industry, chemical engineering, sustainability, and life cycle assessment through digitalization[10–14]. Thus, this manuscript shows how digitalization may streamline of MFA and CRA at chemical EoL. The topics examined in this study encompass a data-centric approach, where the quality and accessibility of data improve over time (see Figure 1).

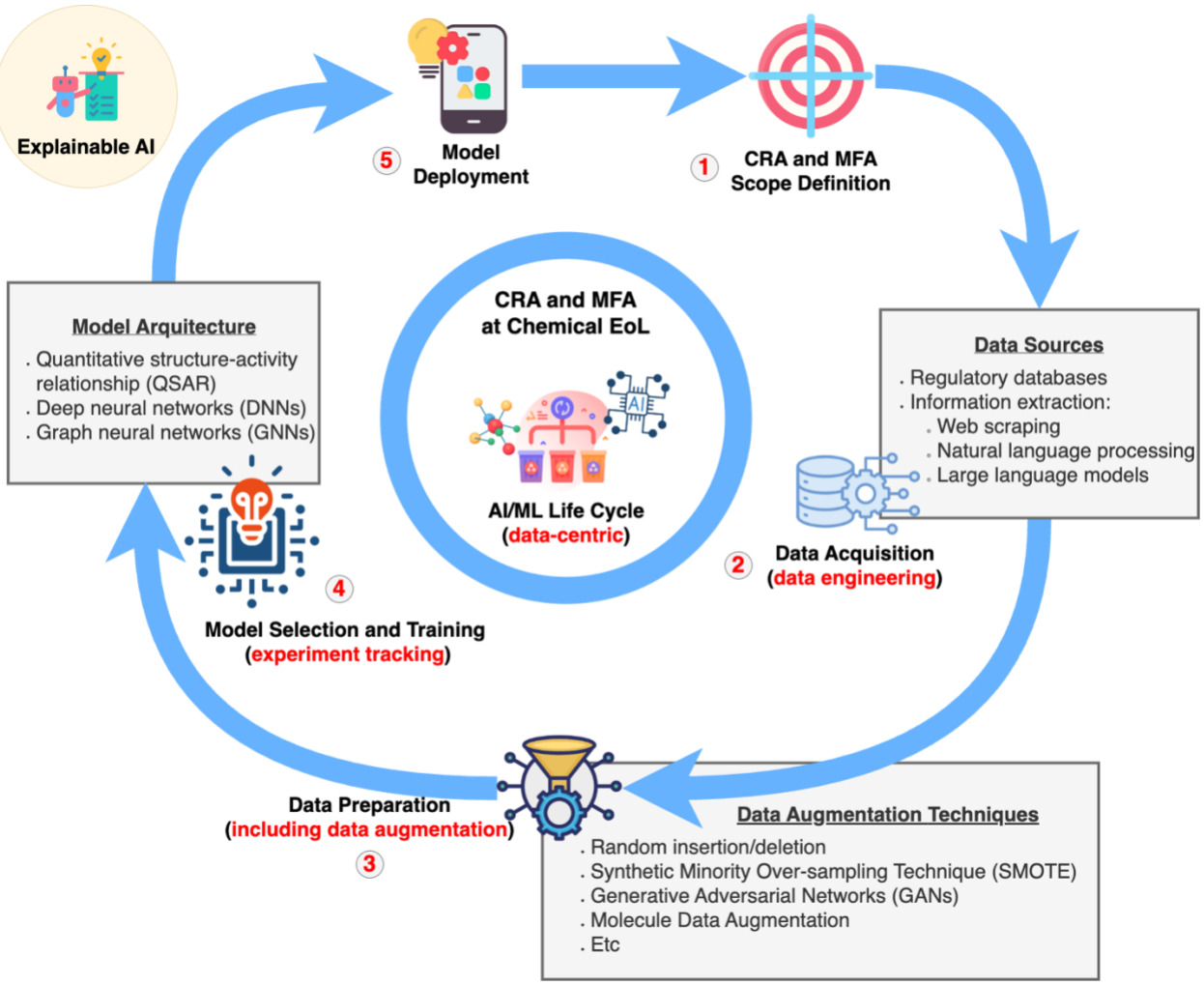

Figure 1. This figure presents the machine learning life cycle applied to Chemical risk assessment (CRA)and material flow analysis (MFA)at the end-of-life (EoL) stage. It integrates data sources, including regulatory databases and information extraction tools, feeding into advanced artificial intelligence (AI) models such as Deep Neural Networks (DNNs), Graph Neural Networks (GNNs), and Generative Adversarial Networks (GANs). These models support analysis and predictions related to exposure pathways and regulatory compliance. The framework highlights how these tools collaboratively enable a data-driven approach to effective EoL chemical management.

2. End-of-life chemical data availability

2.1. Regulatory data sources

High-performance AI models require domain-specific data to extract "knowledge" about the EoL supply and management chain's behavior through data patterns. Regulatory database systems can help collect facility-level EoL and chemical release data. The Pollutant Release and Transfer Registers (PRTRs), an international publicly accessible database system created by the Organization of Economic Co-

operation and Development (OECD), provides data on the releases of toxic chemicals into the air, water, and land by industrial facilities and transferred off-site for treatment[15]. Also, the OECD uses the PRTR to evaluate progress towards the United Nation's twelve sustainable development goal, which promotes sustainable consumption and production practices[16, 17].

The PRTRs from Australia, Canada, and the USA provide individual chemical transfer data[18–20]. These publicly access databases depict EoL management scenarios with greater quantitative and qualitative detail. The off-site material transfer report is done by chemical for all EoL scenarios, which support AI models performance because the risk of bias is lower than with other PRTRs reporting aggregated EoL material transfer amounts by whether the transferred material is a hazardous waste instead of informing the constituent chemicals[21].

The USA PRTR, also known as the Toxics Release Inventory (TRI), is highly granular, allowing data engineering incorporation into environmental impact assessment applications. US government entities use TRI in developing environmental input-output life cycle assessment models[22]. It helps track chemicals through the EoL supply chain and identify waste brokering and intermediaries[23]. It also provides statistics on on-site EoL management operations, potential pollution abatement technologies (e.g., steam stripping), and abatement efficiencies[24]. Also, TRI integration with other data sources supports the design and evaluation of potential chemical circular economy scenarios[25].

CRA sensitivity analysis can use regulatory data sources to reflect worst-case environmental release scenarios despite reporting quantity thresholds, EoL material industry sector generator, and chemical species[23]. As part of data and AI modeling, these data sources can be integrated into a data-centric framework, where the systematic procedure remains unchanged while the sample data size (e.g., chemicals, threshold values, reporting facilities, etc.) and quality increase over time[26]. Thus, data engineering pipelines can merge data silos to measure the influence of environmental regulatory

stringency and economic feasibility on the EoL supply chain across countries and time[21], expanding the AI model domain[26].

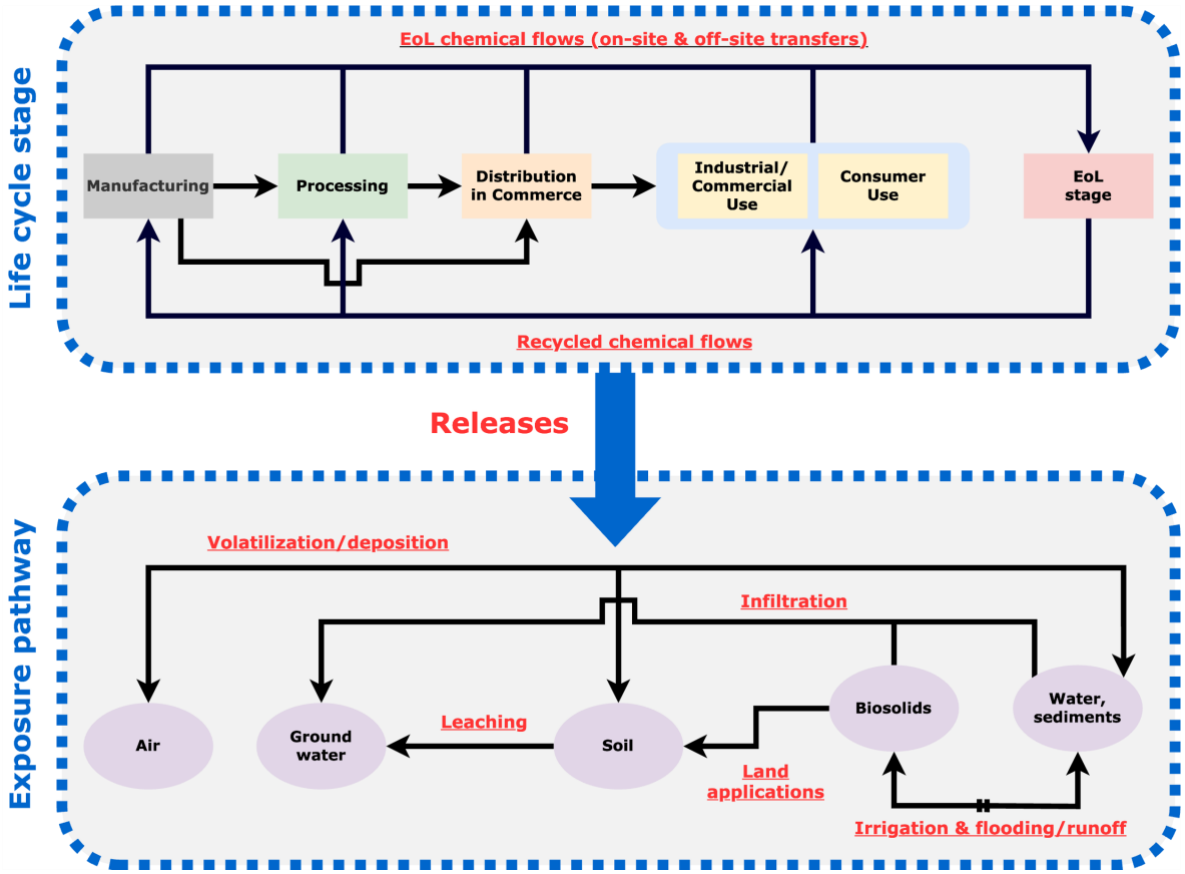

Figure 2. An overview of the chemical life cycle stages and environmental exposure pathways, showing recycled and end-of-life (EoL) material flows (top), and transport between environmental release compartments (bottom) via mass transfer processes like runoff, leaching, and volatilization.

However, PRTR systems, including TRI, encompass no more than 770 chemicals. This is a difficult scenario given the continually rising prevalence of toxic and hazardous chemicals. The Toxic Substance Control Act (TSCA) inventory in the US lists more than 86,000 chemicals. Approximately 2,000 new chemicals are introduced annually in the US. Also, around 2,500 chemicals in the TSCA inventory are classified as high production volume (HPV), with nearly 45 % of these HPV chemicals lack sufficient toxicological studies to assess their health impacts on humans and wildlife[27].

CRA requires the collection of potential exposure scenarios across various life cycle stages, including EoL. It also involves tracking chemical movements between these stages. This process must include detailed data on chemical releases into environmental compartments and transport between them. Additionally, environmental fate factors must be considered, such as biodegradation, bioaccumulation, chemical transformation, and environmental persistence (see Figure 2). It is crucial to develop strategies for the automatic data collection and management. This will expand the applicability of datasets used in training data-driven models.

2.2. Information extraction

Regulatory data sources can provide the EoL supply chain elements like generator/waste handler industry sectors, inter-industry sector transfer amount, and EoL activities. But, these data sources have limitations such as being designated as confidential business information, a limited range of regulated substances and industrial sectors/activities, data granularity, and annual report cycle requirements[21]. Information extraction (IE) may be used to increase dataset size for AI modeling. Computer programs that scrape webpages for data identification and collection are one example. Web scrapers have been used in epidemiology research and public health planning[28], chemical hazards attributes and physical properties[24, 25, 29], textile data extraction for forensic science[30], and the pharmaceutical industry medicament requirements analysis based on prevalent diseases[31]. However, web scraping may be illegal in some jurisdictions and prohibited by website owners[28].

Moreover, AI models can go beyond predicting EoL activities and supply chain constituents. Natural language processing (NLP) can be used in data engineering pipelines to catalog EoL-related academic papers and extract information from portable document format (PDF) files[32]. IE systems have used NLP models to automatically label a corpus including superalloy names and property values for materials research[33]. In toxicology, NLP-based EI systems have help construct biological response pathways from literature. This advances non-animal toxicology research[34]. An NLP-based IE system helped create a

framework that automatically examines and extracts incidents reports. The framework generates risk matrices and analyzes failure modes and effects to address wildfire damage[35].

Large language models (LLMs), a subset of NLP models, have gained attention for their ability to understand and generate text across diverse conditional tasks. LLMs are particularly effective in IE through prompt-based instructions, including from scientific texts[36, 37]. They are valuable for IE tasks in data engineering pipelines for MFA and CRA during the EoL stage. LLMs have proven useful in extracting materials science knowledge from peer-reviewed articles, including phase-property relationships in aluminum alloys and aiding alloy design[35, 38]. They are also valuable in chemical fields for tasks like compound entity recognition, reaction role labeling, and building databases of thermally activated fluorescent molecules [39].

2.3. Data augmentation

In cases of EoL data scarcity, which can cause overfitting or imbalance data for classification learning tasks, data augmentation can improve AI model performance. Data augmentation creates more training data using inherent patterns in existing data[40]. Previously, synthetic minority over-sampling technique (SMOTE) and multilabel SMOTE (or MLSMOTE) were used for classification learning in the context of MFA during the EoL stage[26].

Moreover, data augmentation was used in chemical reaction prediction to improve the synthesis planning of reaction templates and reaction-based molecule optimization. The reaction data was supplemented with template applicability information[41]. Other uses of data augmentation include altering functional groups inside molecules to generate synthetic data and improve chemical reaction predictions[42]. Also, data augmentation is used in chemical process design to digitize chemical process flowsheets by randomly changing branches[43].

Advanced deep learning techniques can find patterns in data and generate artificial sample data for CFA and CRA at EoL. For example, AI models have been improved to predict protein sequence solubility

using deep learning techniques inspired in generative adversarial networks (GANs)[44]. GANs also predict supercritical water gasification in hydrogen production[45]. Conditional GANs augment data of corrosion generated in industrial process pipelines[46]. In biology and other fields, advanced models like the generative pre-trained transformer 4 (GPT-4) have improved predictive modeling[47].

3. Evolution of data-driven models' architecture

In cheminformatics and CRA, data-driven toxicity prediction models inspired in quantitative structure-activity relationship (QSAR) models are common. QSAR models quantitatively correlate molecular structure descriptors with response variables like the water-ethanol partition coefficient[48]. Combining traditional AI tree-based models with QSAR models to understand the EoL supply and management chain can yield high-performance models powered by regulatory datasets[26]. When developing a QSAR model, tree-based algorithms can explain how each model contributes to the response variable prediction.

CRA and MFA can use state-of-the-art AI algorithms to capture more complex data patterns, but model explainability is important. Researchers have used QSAR and deep neural networks (DNNs) in CRA and drug development[49]. AI models that predict chemical toxicity in rats and mice have also used this modeling synergy to reduce the need for in-vitro animal trials for hazard assessment[50]. Using previously learned knowledge in related tasks, transfer learning improves DNNs performance. Transfer learning has been evaluated for chemical process design using reinforcement learning in process system engineering[51]. Transfer learning has also been applied to predict chemical properties using deep graph neural networks (GNNs)[52].

DNNs can be used in conjunction with explainable AI (XAI) to evaluate AI models. XAI helps stakeholders who are not computer scientists understand and rely on AI model outputs[53]. DNNs and QSAR have been used in estimating fish bioconcentration factor research. By adding XAI, researchers were able to score molecules' moieties that most affect bioconcentration factor prediction[54]. Also, XAI has been

used in supply chain management and analysis to improve explainability by combining DNNs and logic-based reasoning[55]. Thus, DNNs and QSAR could model the EoL supply and management chain for MFA and CRA. In summary, XAI could help stakeholders and decisionmakers to prioritize modeling variables (e.g., industry sectors, physical properties), while transfer learning could improve QSAR-inspired model predictability.

GNNs have become more popular in chemoinformatic research, including QSAR modeling. This AI model architecture can manage graph-structured data and learn complex topological relationships. QSAR-inspired GNN algorithms predict synthetic compounds toxicity, environmental behavior, and physiochemistry[56]. GNN is useful for EoL supply chain analysis due to its edge-, node-, or full graph-level prediction. Edge-level GNN predicts hidden linkages and tracks goods and information from suppliers to consumers[57]. Node-level GNN has also been used to classify companies by industry sectors[58]. GNN can introduce edges and links attributes that connect with regulatory and economical constrains, e.g., if an off-site transfers could be considered for legitimate recycling or waste-to-energy under local environmental regulations.

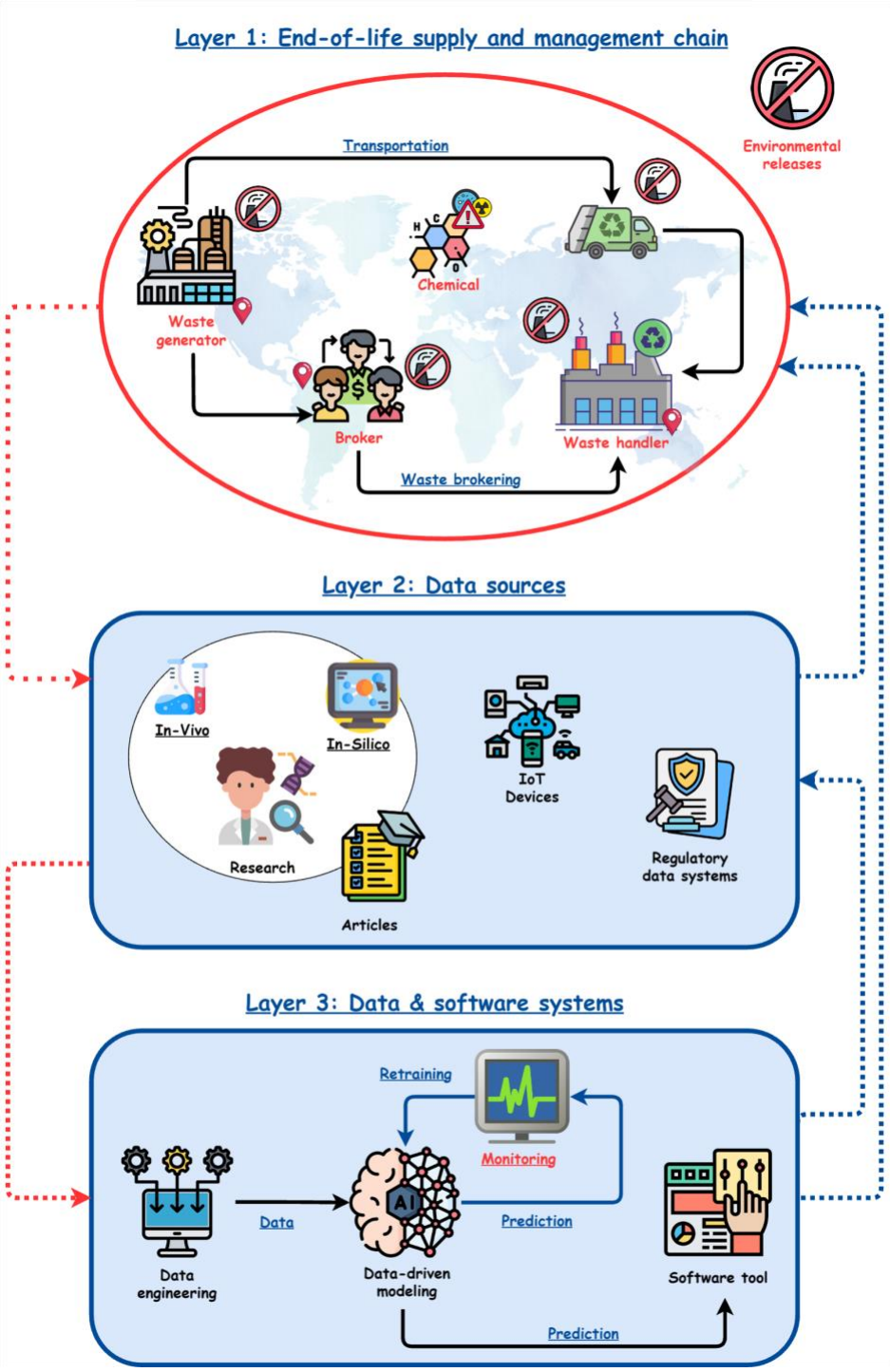

Figure 3. An examination of integrating information systems to depict the end-of-life (EoL) supply and management chain, as well as using real-world systems to gather data (e.g., internet of thing (IoT) technology) for the development of data-driven systems. These systems aid in comprehending and tracking the progression and dynamics of the supply chain, as well as forecasting potential risks to both humans and the environment. The system has the potential to undergo automatic re-optimization and re-training in response to changes in relationships and EoL supply chain behavior, hence mitigating the decline in predictive performance.

4. Software and data systems infrastructure

Figure 3 presents a tiered structure that encompasses the EoL supply and management chain, diverse data sources, and cutting-edge technologies. To understand material flow and classification, the first layer covers the EoL supply and management chain agents like waste handler. The second layer includes various data sources, including publicly available regulatory application programming interfaces (APIs) to analyze the EoL chain constituents. The third layer uses AI models and internet of thing (IoT) technology to analyze and use data sources to improve EoL supply chain decision-making, efficiency, and sustainability.

Recent years have seen exponential growth in data systems and software infrastructure. Open-source initiatives make modern technologies more accessible and created an ecosystem for quickly developing and testing new ideas. The OECD's QSAR Toolbox allows CRA practitioners to share research papers, datasets, and models for future projects[59, 60]. Other open-source QSAR modeling projects include QSAR-Co-X for multitarget QSAR modelling[61], and MRA Toolbox for mixture risk assessment[62]. These tools can speed up AI model training for EoL supply chain understanding as shown in Figure 3. DNNs can also provide data for new model training or transfer learning using QSAR-inspired models.

New technologies make it easier to supply data assets, especially with the rise of AI applications and the data transparency and accessibility. The U.S. Census Bureau provides APIs on the country's industry economy[63], the U.S. Center for Disease Control Prevention provides environmental public health APIs[64], the U.S. Occupational Safety and Health Administration delivers workplace injuries APIs[65], and the U.S. Environmental Protection Agency's CompTox provides computational toxicology APIs[66]. Data engineering pipeline can combine siloed API systems as shown in Figure 3, to extract useful features for modeling the EoL supply chain while considering various factors.

Also, the rise of IoT has great potential for EoL supply chain integration. The IoT uses sensors, software, data processing, and other technologies to simplify internet and communication network connections. IoT

may revolutionize EoL material management by increasing efficiency, reducing environmental impact, and promoting material circularity[67]. IoT may be a valuable source of real-time data that can be integrated into DNNs for real-time biological and non-biological EoL material categorization and sorting via computer vision[68]. IoT and blockchain technology have been used in the chemical supply chain to ensure traceability and transparency[69]. IoT and blockchain could provide real-time data for CRA and MFA. They can track chemicals in the EoL supply chain and use real-time data for environmental decision-making and AI modeling. Figure 3 shows how data can classify EoL material generators, brokers, and handlers and retrain AI models to maintain performance.

Conclusion

This contribution shows how advanced software and data systems can enhance CRA and MFA at the EoL stage. As shown in earlier publications [25, 26], using n-Hexane as an example, data engineering pipelines play a crucial role by integrating information from regulatory databases like PRTRs and TRI and extracting additional insights from academic and industrial documents using NLP. This combined approach allows the identification of specific facilities responsible for n-Hexane releases, quantities transferred off-site, industrial processes involved, and missing data to complete regulatory records. Such integration helps identify key exposure pathways, track waste flows, and analyze abatement technologies.

ML models, including DNNs and GNNs, leverage these enriched datasets to predict exposure scenarios and assess risks associated with environmental parameters like bioaccumulation, volatility, and persistence. By uncovering patterns in chemical releases and transport, these models help estimate risks across the entire lifecycle of n-Hexane. XAI techniques further enhance the interpretability of model predictions, providing stakeholders with a clear understanding of the factors that most influence the model's risk assessments. The approach promotes safer and more efficient chemical management practices and supports regulatory compliance.

Disclaimer

The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. EPA. Any mention of trade names, products, or services does not imply an endorsement by the U.S. Government or the U.S. EPA. The U.S. EPA does not endorse any commercial products, service, or enterprises.

References and recommended readings

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest.

[1] World Health Organization. *WHO Human Health Risk Assessment Toolkit: Chemical Hazards*, 2nd ed.; Geneva, 2021; Vol. 8.

[2] Lombardi, M.; Amicarelli, V.; Bux, C.; Varese, E. Sustainable Development and Waste Management. In *Reference Module in Earth Systems and Environmental Sciences*; Elsevier, 2023. https://doi.org/10.1016/B978-0-323-93940-9.00013-X.

[3] Meyer, D. E.; Mittal, V. K.; Ingwersen, W. W.; Ruiz-Mercado, G. J.; Barrett, W. M.; Gonzalez, M. A.; Abraham, J. P.; Smith, R. L. Purpose-Driven Reconciliation of Approaches to Estimate Chemical Releases. *ACS Sustain Chem Eng*, **2019**, *7* (1), 1260–1270. https://doi.org/10.1021/acssuschemeng.8b04923.

[4] Kullmann, F.; Markewitz, P.; Stolten, D.; Robinius, M. Combining the Worlds of Energy Systems and Material Flow Analysis: A Review. *Energy Sustain Soc*, **2021**, *11* (1), 13. https://doi.org/10.1186/s13705-021-00289-2.

[5] Woodruff, T. J.; Rayasam, S. D. G.; Axelrad, D. A.; Koman, P. D.; Chartres, N.; Bennett, D. H.; Birnbaum, L. S.; Brown, P.; Carignan, C. C.; Cooper, C.; et al. A Science-Based Agenda for Health-Protective Chemical Assessments and Decisions: Overview and Consensus Statement. *Environmental Health*, **2023**, *21* (S1), 132. https://doi.org/10.1186/s12940-022-00930-3.

[6] Chen, X.; Xu, L.; Ren, Z.; Jia, F.; Yu, Y. Sustainable Supply Chain Management in the Leather Industry: A Systematic Literature Review. *International Journal of Logistics Research and Applications*, **2023**, *26* (12), 1663–1703. https://doi.org/10.1080/13675567.2022.2104233.

[7] Hernandez-Betancur, J. D.; Ruiz-Mercado, G. J. Sustainability Indicators for End-of-Life Chemical Releases and Potential Exposure. *Curr Opin Chem Eng*, **2019**, *26*, 157–163. https://doi.org/10.1016/j.coche.2019.09.004.

[8] Voulvoulis, N.; Skolout, J. W. F.; Oates, C. J.; Plant, J. A. From Chemical Risk Assessment to Environmental Resources Management: The Challenge for Mining. *Environmental Science and Pollution Research*, **2013**, *20* (11), 7815–7826. https://doi.org/10.1007/s11356-013-1785-8.

[9] McPartland, J.; Shaffer, R. M.; Fox, M. A.; Nachman, K. E.; Burke, T. A.; Denison, R. A. Charting a Path Forward: Assessing the Science of Chemical Risk Evaluations under the Toxic Substances Control Act in the Context of Recent National Academies Recommendations. *Environ Health Perspect*, **2022**, *130* (2). https://doi.org/10.1289/EHP9649.

[10] Chiang, L. H.; Braun, B.; Wang, Z.; Castillo, I. Towards Artificial Intelligence at Scale in the Chemical Industry. *AIChE Journal*, **2022**, *68* (6). https://doi.org/10.1002/aic.17644.

[11] Schweidtmann, A. M. Generative Artificial Intelligence in Chemical Engineering. *Nature Chemical Engineering*, **2024**, *1* (3), 193–193. https://doi.org/10.1038/s44286-024-00041-5.

[12] Schoormann, T.; Strobel, G.; Möller, F.; Petrik, D.; Zschech, P. Artificial Intelligence for Sustainability—A Systematic Review of Information Systems Literature. *Communications of the Association for Information Systems*, **2023**, *52*, 199–237. https://doi.org/10.17705/1CAIS.05209.

[13] Ghoroghi, A.; Rezgui, Y.; Petri, I.; Beach, T. Advances in Application of Machine Learning to Life Cycle Assessment: A Literature Review. *Int J Life Cycle Assess*, **2022**, *27* (3), 433–456. https://doi.org/10.1007/s11367-022-02030-3.

[14] U.S. Environmental Protection Agency. *Artificial Intelligence Tools and Open Data Practices for EPA Chemical Hazard Assessments*; Beebe, J., Wassel, R., Beins, K., Guyton, K. Z., Eds.; National Academies Press: Washington, D.C., 2022. https://doi.org/10.17226/26540.

[15] OECD Environment Directorate. Uses of PRTR Data and Tools for Their Presentation. In *OECD Series on Pollutant Release and Transfer Registers*; OECD Publishing: Paris, 2023; Vol. 27.

[16] OECD Environment Directorate. Harmonised List of Pollutants for Global Pollutant Release and Transfer Registers (PRTRs). In *OECD Series on Pollutant Release and Transfer Registers*; OECD Publishing: Paris, 2022.

[17] OECD Environment Directorate. Using PRTR Information to Evaluate Progress Towards the Sustainable Development Goal 12. In *OECD Series on Pollutant Release and Transfer Registers*; OECD Publishing: Paris, 2021.

[18] Department of Agriculture Water and the Environment. *DAWE 2022, Review of the National Pollutant Inventory 2021*; Canberra, 2022.

[19] Environment and Climate Change Canada. *Guide for Reporting to the National Pollutant Release Inventory*; Gatineau, 2021.

[20] U.S. Environmental Protection Agency. *2021 TRI National Analysis*; 2023.

[21] ** Hernandez-Betancur, J. D.; Ruiz-Mercado, G. J.; Martin, M. Tracking End-of-Life Stage of Chemicals: A Scalable Data-Centric and Chemical-Centric Approach. *Resour Conserv Recycl*, **2023**, *196*, 107031. https://doi.org/10.1016/j.resconrec.2023.107031.

This manuscript presents the integration a data engineering strategy to integrate PRTR data across years and territories, generating a dataset to be used for ML applications.

[22] Young, B.; Ingwersen, W. W.; Bergmann, M.; Hernandez-Betancur, J. D.; Ghosh, T.; Bell, E.; Cashman, S. A System for Standardizing and Combining U.S. Environmental Protection Agency Emissions and Waste Inventory Data. *Applied Sciences*, **2022**, *12* (7), 3447. https://doi.org/10.3390/app12073447.

[23] Hernandez-Betancur, J. D.; Ruiz-Mercado, G. J.; Abraham, J. P.; Martin, M.; Ingwersen, W. W.; Smith, R. L. Data Engineering for Tracking Chemicals and Releases at Industrial End-of-Life Activities. *J Hazard Mater*, **2021**, *405*, 124270. https://doi.org/10.1016/j.jhazmat.2020.124270.

[24] Hernandez-Betancur, J. D.; Martin, M.; Ruiz-Mercado, G. J. A Data Engineering Framework for On-Site End-of-Life Industrial Operations. *J Clean Prod*, **2021**, *327*, 129514. https://doi.org/10.1016/j.jclepro.2021.129514.

[25] ** Hernandez-Betancur, J. D.; Martin, M.; Ruiz-Mercado, G. J. A Data Engineering Approach for Sustainable Chemical End-of-Life Management. *Resour Conserv Recycl*, **2022**, *178*, 106040. https://doi.org/10.1016/j.resconrec.2021.106040.

This manuscript presents the use of data engineering to integrate off-site transfer data, on-site pollution abatement technology data, and commercial, industrial, and consumer use data to create a markov random field to represent the EoL supply and management chain and recycling loops.

[26] ** Hernandez-Betancur, J. D.; Ruiz-Mercado, G. J.; Martin, M. Predicting Chemical End-of-Life Scenarios Using Structure-Based Classification Models. *ACS Sustain Chem Eng*, **2023**, *11* (9), 3594–3602. https://doi.org/10.1021/acssuschemeng.2c05662.

This manuscript shows the use of data obtained by QSAR and ML modeling to obtain predictive models to determine the probability of occurrence of EoL activities, using predictors like environmental regulatory stringency, added value by industry sectors, chemical flow transfer amount and chemical price.

[27] Muir, D. C. G.; Getzinger, G. J.; McBride, M.; Ferguson, P. L. How Many Chemicals in Commerce Have Been Analyzed in Environmental Media? A 50 Year Bibliometric Analysis. *Environ Sci Technol*, **2023**, *57* (25), 9119–9129. https://doi.org/10.1021/acs.est.2c09353.

[28] Rennie, S.; Buchbinder, M.; Juengst, E.; Brinkley-Rubinstein, L.; Blue, C.; Rosen, D. L. Scraping the Web for Public Health Gains: Ethical Considerations from a 'Big Data' Research Project on HIV and Incarceration. *Public Health Ethics*, **2020**, *13* (1), 111–121. https://doi.org/10.1093/phe/phaa006.

[29] Single, J. I.; Schmidt, J.; Denecke, J. Knowledge Acquisition from Chemical Accident Databases Using an Ontology-Based Method and Natural Language Processing. *Saf Sci*, **2020**, *129*, 104747. https://doi.org/10.1016/j.ssci.2020.104747.

[30] Muehlethaler, C.; Albert, R. Collecting Data on Textiles from the Internet Using Web Crawling and Web Scraping Tools. *Forensic Sci Int*, **2021**, *322*, 110753. https://doi.org/10.1016/j.forsciint.2021.110753.

[31] Dahiya, R.; Nidhi; Kumari, K.; Kumari, S.; Agarwal, N. Usage of Web Scraping in the Pharmaceutical Sector. *EAI Endorsed Trans Pervasive Health Technol*, **2023**, *9*. https://doi.org/10.4108/eetpht.9.4312.

[32] Leonard, K. C.; Hasan, F.; Sneddon, H. F.; You, F. Can Artificial Intelligence and Machine Learning Be Used to Accelerate Sustainable Chemistry and Engineering? *ACS Sustain Chem Eng*, **2021**, *9* (18), 6126–6129. https://doi.org/10.1021/acssuschemeng.1c02741.

[33] Yan, R.; Jiang, X.; Wang, W.; Dang, D.; Su, Y. Materials Information Extraction via Automatically Generated Corpus. *Sci Data*, **2022**, *9* (1), 401. https://doi.org/10.1038/s41597-022-01492-2.

[34] Corradi, M. P. F.; de Haan, A. M.; Staumont, B.; Piersma, A. H.; Geris, L.; Pieters, R. H. H.; Krul, C. A. M.; Teunis, M. A. T. Natural Language Processing in Toxicology: Delineating Adverse Outcome Pathways and Guiding the Application of New Approach Methodologies. *Biomaterials and Biosystems*, **2022**, *7*, 100061. https://doi.org/10.1016/j.bbiosy.2022.100061.

[35] Andrade, S. R.; Walsh, H. S. Machine Learning Framework for Hazard Extraction and Analysis of Trends (HEAT) in Wildfire Response. *Saf Sci*, **2023**, *167*, 106252. https://doi.org/10.1016/j.ssci.2023.106252.

[36] Xu, D.; Chen, W.; Peng, W.; Zhang, C.; Xu, T.; Zhao, X.; Wu, X.; Zheng, Y.; Chen, E. Large Language Models for Generative Information Extraction: A Survey. **2023**.

[37] Dagdelen, J.; Dunn, A.; Lee, S.; Walker, N.; Rosen, A. S.; Ceder, G.; Persson, K. A.; Jain, A. Structured Information Extraction from Scientific Text with Large Language Models. *Nat Commun*, **2024**, *15* (1), 1418. https://doi.org/10.1038/s41467-024-45563-x.

[38] Polak, M. P.; Morgan, D. Extracting Accurate Materials Data from Research Papers with Conversational Language Models and Prompt Engineering. *Nat Commun*, **2024**, *15* (1), 1569. https://doi.org/10.1038/s41467-024-45914-8.

[39] Huang, D.; Cole, J. M. A Database of Thermally Activated Delayed Fluorescent Molecules Auto-Generated from Scientific Literature with ChemDataExtractor. *Sci Data*, **2024**, *11* (1), 80. https://doi.org/10.1038/s41597-023-02897-3.

[40] Mumuni, A.; Mumuni, F. Data Augmentation: A Comprehensive Survey of Modern Approaches. *Array*, **2022**, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258.

[41] Fortunato, M. E.; Coley, C. W.; Barnes, B. C.; Jensen, K. F. Data Augmentation and Pretraining for Template-Based Retrosynthetic Prediction in Computer-Aided Synthesis Planning. *J Chem Inf Model*, **2020**, *60* (7), 3398–3407. https://doi.org/10.1021/acs.jcim.0c00403.

[42] Wu, X.; Zhang, Y.; Yu, J.; Zhang, C.; Qiao, H.; Wu, Y.; Wang, X.; Wu, Z.; Duan, H. Virtual Data Augmentation Method for Reaction Prediction. *Sci Rep*, **2022**, *12* (1), 17098. https://doi.org/10.1038/s41598-022-21524-6.

[43] Balhorn, L. S.; Hirtreiter, E.; Luderer, L.; Schweidtmann, A. M. Data Augmentation for Machine Learning of Chemical Process Flowsheets; 2023; pp 2011–2016. https://doi.org/10.1016/B978-0-443-15274-0.50320-6.

[44] Han, X.; Zhang, L.; Zhou, K.; Wang, X. ProGAN: Protein Solubility Generative Adversarial Nets for Data Augmentation in DNN Framework. *Comput Chem Eng*, **2019**, *131*, 106533. https://doi.org/10.1016/j.compchemeng.2019.106533.

[45] Ma, Z.; Wang, J.; Feng, Y.; Wang, R.; Zhao, Z.; Chen, H. Hydrogen Yield Prediction for Supercritical Water Gasification Based on Generative Adversarial Network Data Augmentation. *Appl Energy*, **2023**, *336*, 120814. https://doi.org/10.1016/j.apenergy.2023.120814.

[46] Ma, H.; Geng, M.; Wang, F.; Zheng, W.; Ai, Y.; Zhang, W. Data Augmentation of a Corrosion Dataset for Defect Growth Prediction of Pipelines Using Conditional Tabular Generative Adversarial Networks. *Materials*, **2024**, *17* (5), 1142. https://doi.org/10.3390/ma17051142.

[47] Xiao, Z.; Li, W.; Moon, H.; Roell, G. W.; Chen, Y.; Tang, Y. J. Generative Artificial Intelligence GPT-4 Accelerates Knowledge Mining and Machine Learning for Synthetic Biology. *ACS Synth Biol*, **2023**, *12* (10), 2973–2982. https://doi.org/10.1021/acssynbio.3c00310.

[48] Shi, W.; Guo, J.; Bao, T. QSAR Tools for Toxicity Prediction in Risk Assessment—Comparative Analysis. In *QSAR in Safety Evaluation and Risk Assessment*; Elsevier, 2023; pp 203–218. https://doi.org/10.1016/B978-0-443-15339-6.00016-3.

[49] Xu, T.; Ngan, D. K.; Huang, R. Application of QSAR Models Based on Machine Learning Methods in Chemical Risk Assessment and Drug Discovery. In *QSAR in Safety Evaluation and Risk Assessment*; Elsevier, 2023; pp 245–258. https://doi.org/10.1016/B978-0-443-15339-6.00006-0.

[50] Bo, T.; Lin, Y.; Han, J.; Hao, Z.; Liu, J. Machine Learning-Assisted Data Filtering and QSAR Models for Prediction of Chemical Acute Toxicity on Rat and Mouse. *J Hazard Mater*, **2023**, *452*, 131344. https://doi.org/10.1016/j.jhazmat.2023.131344.

[51] Gao, Q.; Yang, H.; Shanbhag, S. M.; Schweidtmann, A. M. Transfer Learning for Process Design with Reinforcement Learning; 2023; pp 2005–2010. https://doi.org/10.1016/B978-0-443-15274-0.50319-X.

[52] Buterez, D.; Janet, J. P.; Kiddle, S. J.; Oglic, D.; Lió, P. Transfer Learning with Graph Neural Networks for Improved Molecular Property Prediction in the Multi-Fidelity Setting. *Nat Commun*, **2024**, *15* (1), 1517. https://doi.org/10.1038/s41467-024-45566-8.

[53] Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J. M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence. *Information Fusion*, **2023**, *99*, 101805. https://doi.org/10.1016/j.inffus.2023.101805.

[54] * Zhao, L.; Montanari, F.; Heberle, H.; Schmidt, S. Modeling Bioconcentration Factors in Fish with Explainable Deep Learning. *Artificial Intelligence in the Life Sciences*, **2022**, *2*, 100047. https://doi.org/10.1016/j.ailsci.2022.100047.

This manuscript shows the use of XAI to obtain valuable information about the predictor importance in the performance and results of ML models, which results important if a black box models like traditional DNNs are used to obtained predictive models for CRA at EoL.

[55] Kosasih, E. E.; Papadakis, E.; Baryannis, G.; Brintrup, A. A Review of Explainable Artificial Intelligence in Supply Chain Management Using Neurosymbolic Approaches. *Int J Prod Res*, **2024**, *62* (4), 1510–1540. https://doi.org/10.1080/00207543.2023.2281663.

[56] ** Wang, H.; Liu, W.; Chen, J. QSAR Modeling Based on Graph Neural Networks. In *QSAR in Safety Evaluation and Risk Assessment*; Elsevier, 2023; pp 139–151. https://doi.org/10.1016/B978-0-443-15339-6.00012-6.

This manuscript presents an interesting application of GNNs combined with QSAR modeling strategy, showing a promised performance of this model architecture in the CRA field.

[57] ** Kosasih, E. E.; Brintrup, A. A Machine Learning Approach for Predicting Hidden Links in Supply Chain with Graph Neural Networks. *Int J Prod Res*, **2022**, *60* (17), 5380–5393. https://doi.org/10.1080/00207543.2021.1956697.

This manuscript shows the use of GNN architecture to predict connections in the supply chain network by using link-level predictions. This is an interesting application that could be extrapolated to the EoL supply chain to identify EoL activities based on sectors and chemical conditions of use based on sectors and chemical descriptors.

[58] * Wu, D.; Wang, Q.; Olson, D. L. Industry Classification Based on Supply Chain Network Information Using Graph Neural Networks. *Appl Soft Comput*, **2023**, *132*, 109849. https://doi.org/10.1016/j.asoc.2022.109849.

This manuscript shows how GNNs could be used to obtain node-level predictions to provide important supply chain information like industry sectors involved in the value chain for products/services, which could be extended to identify sector-level information regarding chemical conditions of use for CRA.

[59] Kutsarova, S.; Schultz, T. W.; Chapkanov, A.; Cherkezova, D.; Mehmed, A.; Stoeva, S.; Kuseva, C.; Yordanova, D.; Georgiev, M.; Petkov, T.; et al. The QSAR Toolbox Automated Read-across Workflow for Predicting Acute Oral Toxicity: II. Verification and Validation. *Computational Toxicology*, **2021**, *20*, 100194. https://doi.org/10.1016/j.comtox.2021.100194.

[60] Kutsarova, S.; Mehmed, A.; Cherkezova, D.; Stoeva, S.; Georgiev, M.; Petkov, T.; Chapkanov, A.; Schultz, T. W.; Mekenyan, O. G. Automated Read-across Workflow for Predicting Acute Oral Toxicity: I. The Decision Scheme in the QSAR Toolbox. *Regulatory Toxicology and Pharmacology*, **2021**, *125*, 105015. https://doi.org/10.1016/j.yrtph.2021.105015.

[61] Halder, A. K.; Dias Soeiro Cordeiro, M. N. QSAR-Co-X: An Open Source Toolkit for Multitarget QSAR Modelling. *J Cheminform*, **2021**, *13* (1), 29. https://doi.org/10.1186/s13321-021-00508-0.

[62] Kim, J.; Seo, M.; Choi, J.; Na, M. MRA Toolbox v. 1.0: A Web-Based Toolbox for Predicting Mixture Toxicity of Chemical Substances in Chemical Products. *Sci Rep*, **2022**, *12* (1), 8880. https://doi.org/10.1038/s41598-022-13028-0.

[63] U.S. Census Bureau. Available APIs https://www.census.gov/data/developers/data-sets.html (accessed Mar 25, 2024).

[64] Centers for Disease Control and Prevention (CDC). Tracking Network Data Application Program Interface (API) https://ephtracking.cdc.gov/apihelp (accessed Mar 25, 2024).

[65] Occupational Safety and Health Administration (OSHA). Injury Tracking Application (ITA) https://www.osha.gov/injuryreporting/ (accessed Mar 25, 2024).

[66] U.S. Environmental Protection Agency. CompTox Data and APIs https://www.epa.gov/comptox-tools/comptox-data-and-apis (accessed Mar 25, 2024).

[67] Sharma, R. Leveraging AI and IoT for Sustainable Waste Management; 2023; pp 136–150. https://doi.org/10.1007/978-3-031-47055-4_12.

[68] Rahman, Md. W.; Islam, R.; Hasan, A.; Bithi, N. I.; Hasan, Md. M.; Rahman, M. M. Intelligent Waste Management System Using Deep Learning with IoT. *Journal of King Saud University - Computer and Information Sciences*, **2022**, *34* (5), 2072–2087. https://doi.org/10.1016/j.jksuci.2020.08.016.

[69] Bhattacharya, P.; Verma, A.; Sharma, G. Blockchain-Driven and IoT-Assisted Chemical Supply-Chain Management; 2022; pp 779–791. https://doi.org/10.1007/978-981-19-0284-0_57.