A more sustainable future begins with collaboration and data

Stacy Kish

- The US Environmental Protection Agency's (EPA) Office of Chemical Safety and Pollution Prevention administers the Safer Choice program to help consumers, businesses, and purchasers find cleaning and other products that perform and contain ingredients that are safer for human health and the environment.

- The EPA's Office of Research and Development offers a series of linked databases and online applications where users can access chemical information on thousands of chemicals including the structure and properties of chemicals previously studied by the agency, as well as methods available to analyze chemicals.

- ExxonMobil* has leveraged over three decades of in-house testing to develop a model to help their customers identify potential alternate chemicals for cleaning product formulations.

- With the use of EPA databases and industry assistance, many companies create new, safer product formulations and submit their product information to the EPA to obtain Safer Choice-certification to advertise that their products work well and contain only safer ingredients for people and the planet.

Developing environmentally safe products benefits from the collaboration of consumers, manufacturers, policy makers, and—it should be no surprise—computers. Multiple chemical databases have been operated by government agencies and private developers for decades. The tools have been used to store and access chemical information, thus far. However, in a world of chat bots—trained to produce human language—can we anticipate a system trained to predict chemical safety?

"We are trying to harvest, connect, and integrate data so we can bring it all into a much-needed network," says Antony Williams, cheminformatician at the United States Environmental Protection Agency (EPA), Center for Computational Toxicology and Exposure (CCTE).

Publicly available databases and models, like those described in this story, could provide artificial intelligence with the knowledge to help manufacturers create more consumer safe products. The wealth of data has the potential to provide a greener future for chemical manufacturing.

IDENTIFYING THE SAFE CHOICE

Beginning in the early 1990s, the EPA began working with companies to help them identify safer chemicals and processes. These partnerships lead to resources that would help manufacturers assess product safety and consider alternatives when needed. The work eventually resulted in the development of the Safer Choice Standard (https://www.epa.gov/saferchoice/standard).

The Safer Choice program reviews and analyzes chemicals based on their function as an ingredient within a formulation, such as surfactants, solvents, and chelating agents. For example, data for a novel surfactant compound can be compared to the EPA's Safer Choice Criteria for Surfactants to evaluate its biodegradability and toxicity. Then, formulators can evaluate if the substitution is safer before they make a swap. This approach allows formulators to use those ingredients with the lowest hazard in their functional class, while still formulating high-performing products.

Products meeting EPA's criteria are allowed to use the Safer Choice label. Many companies seek this voluntary label to highlight their company's sustainability commitment to purchasers and retailers. At the same time, customers can look for the label to give them reassurance that the product's formulation has been reviewed and the ingredients used are deemed safest in their class.

Qualified third parties compile hazard information for every chemical ingredient in a product submitted for Safer Choice certification. The EPA reviews the hazard information and makes the final determination of acceptability. This review is grounded in the cache of information housed in a chemical's structure which explains how it will behave in the environment and human body. Scientists can evaluate and compare these compounds to those with known toxicology profiles.

The Safer Choice program leverages the expertise and work of EPA scientists to compile a vast amount of data on chemicals of interest to the agency. Now, the EPA has taken the important step of making the information more easily accessible to the general public.

A LIBRARY OF DATA

Formatting such a diverse collection of data can be daunting. To help address this complicated task, the agency developed a series of new software tools that provide fit-for-purpose science applications that range from human hazard to risk characterization. The goal is to deliver empirical and predictive data to chemists to take chemical evaluations for new product formulation beyond today's capabilities.

"We have granted public access to our chemical data for a long time. We have provided computational products that deliver complex information in a way that addresses their needs," said Williams. "These tools can help people identify safer alternative chemicals when conducting environmental and human health risk assessments."
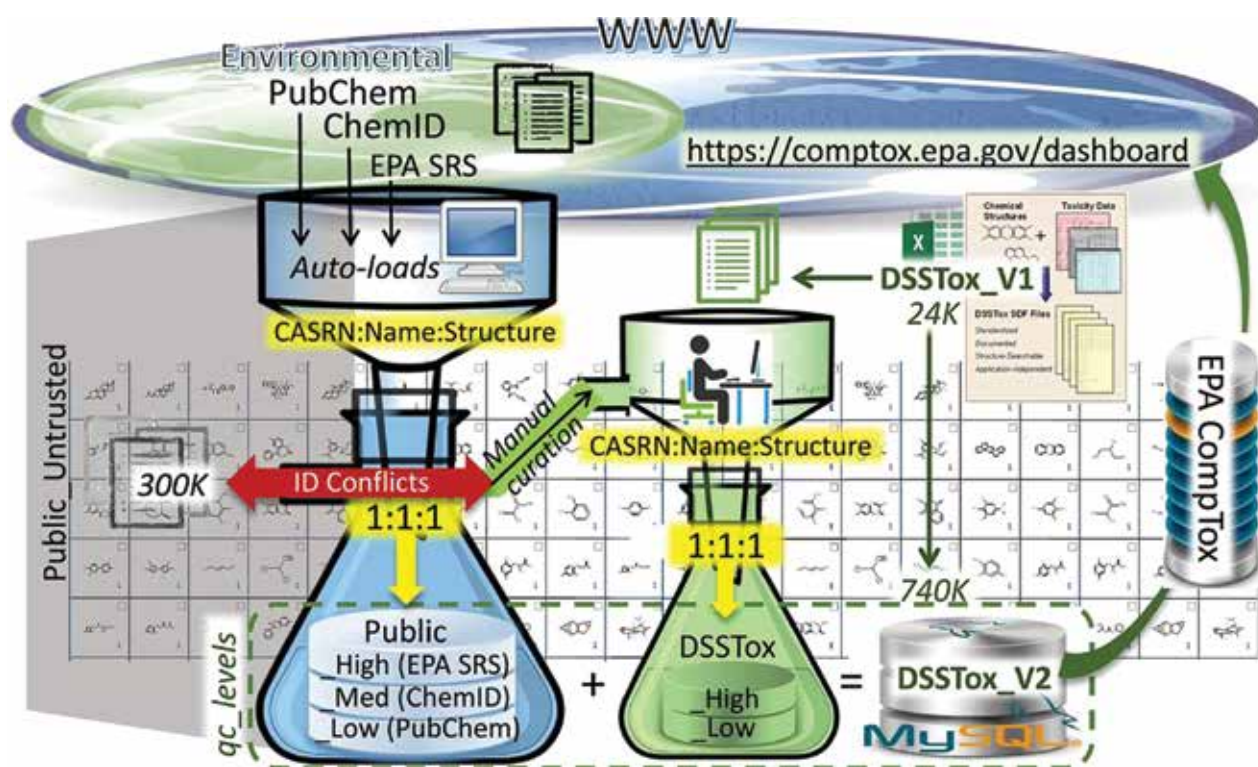
A visual representation of the development of the cheminformatics tools developed by the EPA and made public in the early 2000s. DSSTox serves as the core foundation of EPA's CompTox Chemicals Dashboard [https://comptox.epa.gov/dashboard], which provides public access to DSSTox content for use in modeling and research activities within EPA and, increasingly, across the field of computational toxicology. Source: Grulke, C.M., *Comp Tox, 12*, 100096, 2019.

The effort began with the Distributed Structure-Searchable Toxicity (DSSTox) database (https://doi.org/10.1016/j.comtox.2019.100096) which evolved to contain the chemical structures of more than 1.2 million substances. The DSSTox database provides the foundation for many of the searchable databases and 'dashboards' created by the CCTE.

The latest iteration is the CompTox Chemicals Dashboard (https://tinyurl.com/34h4z5n3) which provides access to information on the substances and multiple predictive models developed at the center and by third parties. Williams' team also created the Cheminformatics Modules (https://tinyurl.com/3czvr77y), *a set of proof-of-concept, web-based tools that provide novel ways to search and visualize chemistry-related data. Other databases under development help gather more than 3,000 analytical methods* for measuring specific chemical structures. The collection is focused on liquid chromatography–mass spectrometry and gas chromatography–mass spectrometry methods, but also includes nuclear magnetic resonance and infrared methods.

The beauty of these searchable databases is that they are linked using unique substance identifiers (called DTXSIDs) to allow a user to move seamlessly between chemical details, from information on their properties to their toxicity data. As more databases come online, like the methods for analysis database, the identifier will allow users to scan them too.

The chemicals can be organized into lists by chemical classes or by regulatory assignment, like the Integrated Risk Information System, the Toxics Release Inventory, and various iterations of the Toxic Substances Control Act. The database does not just manage distinct chemical structures but also substances of unknown or variable composition (UVCBs), a grouping that commonly includes surfactants.

Williams and his team perform both semi-automated and manual curation that is both iterative and ongoing to *streamline searches for chemicals with similar structures, properties, and analytical methods in the databases. The process also* underpins *predictive modeling.*

MODELING A EPA SAFER CHOICE CERTIFIED SURFACTANT

Finding easier ways to formulate products that can meet consumer expectations and help reduce harm to the environment is not an activity restricted to the federal government. ExxonMobil developed models for predicting surfactant toxicity.

The company compiled detailed environmental profiles of ethoxylated derivatives of their Exxal™ isomeric branched, primary alcohols. The profiles included alcohols with both even- and odd-numbered hydrocarbon chains, ranging from C8 to C13. These compounds are commonly used by ExxonMobil customers to synthesize new surfactant products. Recently, their customers expressed an interest in pursuing a Safer Choice designation on their new product formulations. To help support their customers, ExxonMobil used computational models to characterize the environmental profiles of new compounds that could potentially be used in a wide-range of applications.
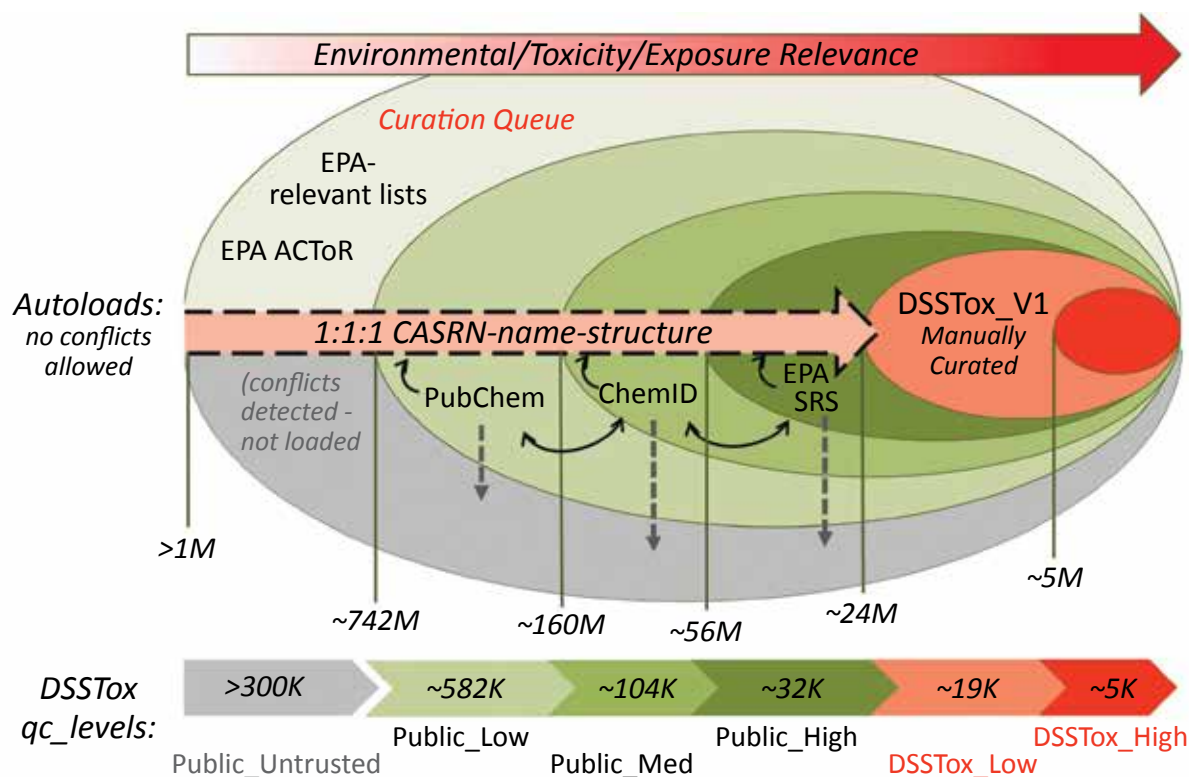
Content from three public databases (EPA's Substance Registry Services – SRS, NLM's ChemID, and PubChem) are quality filtered before being loaded into the DSSTox_Core portion of the DSSTox_V2 data model or rejected and placed in the Public_Untrusted bin, requiring further curation review along with other queued EPA lists. Source: Grulke, C.M., *Comp Tox, 12*, 100096, 2019.

"The models can help determine if substances are good candidates for the Safer Choice listing and if they are classed appropriately," said Jennifer Foreman, regulatory affairs advisor with ExxonMobil. "We are trying to provide a degree of certainty, so our customers can formulate products that can help achieve their environmental goals."

Ecotoxicologists in the company's biomedical sciences group took data from a vast historical dataset and conducted in-house testing consistent with Organization for Economic Co-operation and Development (OCED) guidelines. They then used it to help create predictive models for acute toxicity values of ethoxylated derivatives of Exxal™ alcohols and compared them against cutoff thresholds for the Safer Choice Surfactant Criteria. The model results are then validated against EPA models predictions to help increase confidence in both models.

New chemicals are incorporated into the model to predict its place along the continuum of known data points. From the known chemical structure information, it is possible to interpolate between the data to make assumptions on how a new compound will perform in individual tests. If the new compound's predicted value is close to a classification threshold for acute toxicity, the chemical is flagged for additional testing. If the value falls well outside the threshold, the compound is identified as a good candidate for consideration in the EPA's Safer Choice program.

ExxonMobil has historically shared testing data with customers so those seeking Safer Choice labeling could include

An example of the an EPA Safer Choice label. Products earn the label by meeting the EPA's strict requirements for human health, the environment, and performance.

the information in their submission package to the EPA. The company has recently shared the model output with the EPA to help validate ExxonMobil's model results and interpolation strategy. Based on the agency's response, this approach could help facilitate their customers being able to achieve the Safer Choice label without the expense and time needed for additional testing. This collaborative approach could create a win-win for both customers and the EPA.

"We are in the process of working with the EPA on pre-qualification work with the aim to make it more efficient for our customers to obtain the Safer Choice label and develop new formulations with confidence," Foreman said.

They initially used ethoxylated surfactants to establish their model with the EPA, because of the popularity of their alcohol compounds in the production of new cleaning products. ExxonMobil's alcohol ethoxylates contain more groups branching off the compound backbone. The branches have been shown to improve performance, which can make them a competitive option for new cleaning formulations. ExxonMobil's continued research and testing has disproven the perception that the additional branching hinders biodegradation.

"We have in-house data that supports that these compounds meet the more stringent 10-day biodegradation window," says Foreman, "and we continue to generate additional data on behalf of our customers."

To help increase the transparency of their testing, ExxonMobil published their testing data in peer-reviewed scientific journals for public comment and consideration. The company has also actively engaged the EPA, bringing their models, raw data, and model output to the agency for review and evaluation. Beyond reviewing ExxonMobil's ecotoxilogic models, ExxonMobil has approached the EPA to help determine what additional information—beyond ecotoxicology of the surfactant—an applicant would need to submit in order to obtain a Safer Choice label, like the thresholds for potential impurities.

"Everyone has to work together to bring all of the resources to bear," said Foreman. "There are multiple checks and balances in place to verify the right recommendations are being brought forward and evaluated."

ExxonMobil has now developed categories for derivatives of five of their branched alcohols, namely Exxal™ 8-11 and Exxal™ 13. Many of the ethoxylates in these categories are predicted to meet the more stringent EPA Safer Choice direct release criteria, which is designed for products that bypass sewage treatment and could end up directly in a waterway. Foreman is hopeful that some form of prequalification for these compounds will be available to customers within the next few years.

"Voluntary labels are a really interesting area, because it is a place where you can help support companies as they continuously improve and innovate," said Foreman. "Companies are trying to bring products with sustainability benefits to market for customers who want to make that choice. ExxonMobil is working across the value chain to help make these efforts possible."

In the past two decades, computational tools have reigned-in the scale and complexity of determining which chemicals can meet the EPA's Safer Choice labeling requirements. As scientists, like those at ExxonMobil and the EPA, continue to optimize data storage and platform connectivity the task will become even more manageable. Ultimately, artificial intelligence could be trained to analyze and interpret these data to provide cleaning formulations with both maximum performance and maximum benefit to society.

Stacy Kish is a science writer for INFORM and other media outlets. She can be contacted at earthspin.science@gmail.com.

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