A Web-Based Application for Viewing Exposure Data

Internal Documentation

# Overview

Under the amended Toxics Substances Control Act (TSCA), the U.S. Environmental Protection Agency (EPA) is tasked with evaluating both new and existing chemicals in commerce to determine if the pose an unreasonable risk to human health or the environment. TSCA also requires that EPA consider “highly exposed or susceptible subpopulations” as part of its risk evaluations. Often, the lack of easily accessible data relevant for exposure creates challenges for efforts to evaluate a chemical’s risk.

Recently, the EPA’s Office of Research and Development (ORD), has developed databases, models, and other computational tools that provide exposure-relevant data for many different chemicals. These tools typically use only publicly available data, and thus allow transparency in where and how the data are obtained. However, these tools require knowledge of multiple programming languages and/or database query languages. Learning and utilizing these tools is therefore time prohibitive to risk assessors who may need to rapidly assess many chemicals.

To assist assessors in more quickly surveying the exposure landscape around a specific chemical and quickly provide them with results from ORD’s computational tools, the Exposure Data Viewer application has been developed. This tool summarizes information from various EPA models and U.S. government data sources in a single place. Data are summarized visually with links to more details about the data and model predictions provided as links in the application.

A primary focus of this application is the ability to provide information relevant to workers, which are considers a “highly exposed” population. Specifically, this tool provides collated data from longitudinal indoor air concentration measurements from the 1980s onward. These data, which are collected and published by the U.S. Occupational Safety and Health Administration (OSHA), are presented for a searched chemical in different industrial sectors (provided by North American Industry Classification System, NAICS, codes). Further, results from Bayesian statistical models and equations from EPA’s ChemSTEER models are provided compared to general population exposure models.

Finally, when information for a given chemical is not sufficient, techniques such as read-across can be helpful for filling in information gaps. The Exposure Data Viewer also provides a means to determine chemical structure-based analogs. Exposure information is retrieved and summarized for users so that structurally similar chemicals may be used in determining risk for a data-poor chemical.

## Data Sets

The Exposure Data Viewer provides an overview of the chemical exposure landscape for a specific chemical by utilizing several different datasets that have been collected and curated by ORD from publicly available sources. Each of these datasets is described below:

### Chemical and Products Database

EPA’s Chemicals and Products Database (CPDat) contains information on how a chemical may be used in commerce as well as information about the composition of chemicals in products1,2. CPDat provides information on the functional use, product use, and sector of use for chemicals from documents published by manufacturers or other government or non-government organizations. Data from CPDat are obtained from ORD’s Computational Toxicology and Exposure (CTX) application programming interface (API). CPDat is a versioned database; the data provided by the Exposure Data Viewer are from CPDat v4.0.

### Chemical Exposure Health Data

OSHA’s Chemical Exposure Health Data (CEHD)3 provides industrial-hygiene samples of chemicals measured in a workplace by OSHA compliance officers from 1984 to 2022. While CEHD provides dermal wipes and air measurements, only the air concentrations are provided by the Exposure Data Viewer. These air concentration measurements are taken by sampling in either a worker’s personal breathing space or the shared working area.

### Chemical Data Reporting

EPA’s Chemical Data Reporting (CDR) consist of information required for manufacturer and importers to report to EPA under TSCA4. This resource contains information about the functional use of chemicals in commerce, how many workers are at sites using a chemical, and what NAICS sectors a chemical is used in, as well as if end-products that contain a chemical are intended to be used by children. CDR cycles are every four years. The data presented in the Exposure Data Viewer are from the 2020 data reporting cycle.

## Models

Models can be used to fill in gaps when data are lacking, especially from the sources provided in the Data Sets section. The following are details of models that have been developed by EPA whose predictions are provided in the Exposure Data Viewer.

### Quantitative Structure-Use Relationships

Quantitative Structure-Use Relationship (QSUR) models take a chemical’s structure as input and output a probability that the chemical could serve a given functional role in a product or process5. Currently, CTX provides access to QSUR predictions for a chemical. Training data for these models originates from the functional use data stored in CPDat.

### Occupational Air Concentration Model

The Occupational Air Concentration Model uses data provided by OSHA’s CEHD to determine 1) if a chemical is likely to be detected in air in a given workplace and 2) if detected, what is the likely air concentration of that chemical6. These concentrations can then be passed to models, like EPA’s Chemical Screening Tool for Exposures and Environmental Releases (ChemSTEER)7, to obtain a workplace exposure for a chemical.

### Command-line Occupational Exposure Tool

The Command-line Occupational Exposure Tool (cloet) provides access to the suite of models available in ChemSTEER8,9. Using cloet, both measured and predicted occupational air concentrations can be converted to exposure concentrations.

### Systematic Empirical Evaluation of Models for Exposure, version 3

The Systematic Empirical Evaluation of Models (SEEM) was originally developed to help quantify uncertainty in exposure estimates based on inferred exposures from human biomonitoring data10,11. The third iteration of SEEM, SEEM3, uses a consensus modeling approach, combined with Bayesian methods to first predict possible exposure pathways for a chemical and then predict the exposure of the chemical12. Predictions from many different models (both developed by EPA and by outside entities) are used to form a consensus exposure prediction. SEEM3 exposure predictions are obtained via CTX.

# Instructions for Using the Exposure Data Viewer

The Exposure Data Viewer is built using Streamlit13, which is a Python programming language package that is used to quickly build interactive, web-based applications for data science purposes. In order to open the Exposure Data Viewer, user must first have a Python installation (version 3.10 or higher). For users unfamiliar with Python, an installation of Python will typically come with the pip command, which is used for installing packages that the Exposure Data Viewer depends on.

|  |
| --- |
| **Note** |
| The Exposure Data Viewer heavily depends on obtaining data via CTX. In order to do this, each user must first obtain an API key for CTX. Please request an API key by emailing ccte\_api@epa.gov prior to running the Exposure Data Viewer. |

## Installation

1. Ensure that the CTX API key is properly saved into your system’s .env file. More information for how to do this can be found in the ctx-python package’s documentation14.
2. Either use git to clone the repository or download the repository from GitHub and save it in a relevant location on your hard drive. The git command for clone this repository is:

**git clone https://github.com/USEPA/Exposure\_Read-Across.git**

1. If you downloaded the repository, unzip the file.
2. Navigate to the Exposure\_Read-Across folder on your hard drive by navigating to the folder’s location in Windows Explorer and then right clicking and choosing “Open in Terminal”.
3. In the Terminal application, install the Exposure Data Viewer by typing

**pip install .**

1. Once the package has successfully installed, you can launch the Exposure Data Viewer by typing the following command into the Terminal Application:

**streamlit run Home\_Page.py**

1. Upon running this command, a browser tab should automatically open in your default web browser. If it does not, then output from the above command should provide a link to click to access the application.

## Navigation

Once launched, the Exposure Data Viewer has three different webpages listed in the left-hand navigation bar:

* Home: provides general information about using the application
* Chemical Exposure Information Search: allows users to search for reported and predicted information available for a searched chemical
* Chemical Exposure Read-Across: allows users to search for chemical analogs for a specified chemical, once the analogs are identified, then relevant exposure information for those analogs are provided.

Both of the Chemical Exposure pages allow users to search for a chemical by either drawing a structure of interest or by providing a Distributed Structure-Searchable Toxicity (DSSTox) Substance Identifier (DTXSID) or by providing a Chemical Abstract Services Registration Number (CAS-RN).

# Examples

## Chemical Exposure Information Search

### Chemical Structure Search

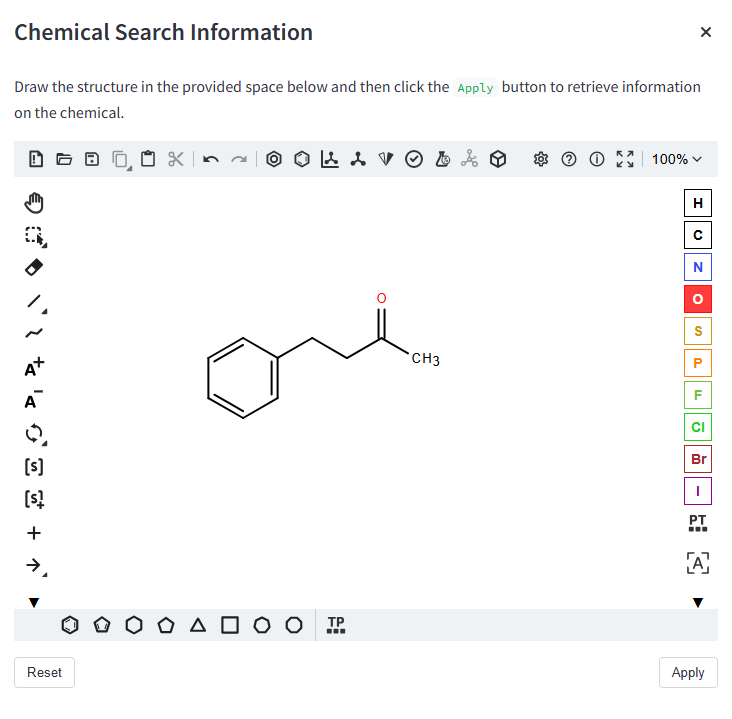
1. Launch the Exposure Data Viewer and navigate to the Chemical Exposure Information Search page.
2. From the Chemical Search Method drop-down menu select “Structure”
3. Using the Ketcher modal interface, draw the chemical structure of interest and click “Apply”

Figure 1: Ketcher modal interface for inputing chemical structure to be searched.

1. The Exposure Data Viewer will take a few seconds to query, retrieve, and then display the information relevant to the chemical, but once all data are displayed the Ketcher interface will close automatically, and the user can examine the displayed information.

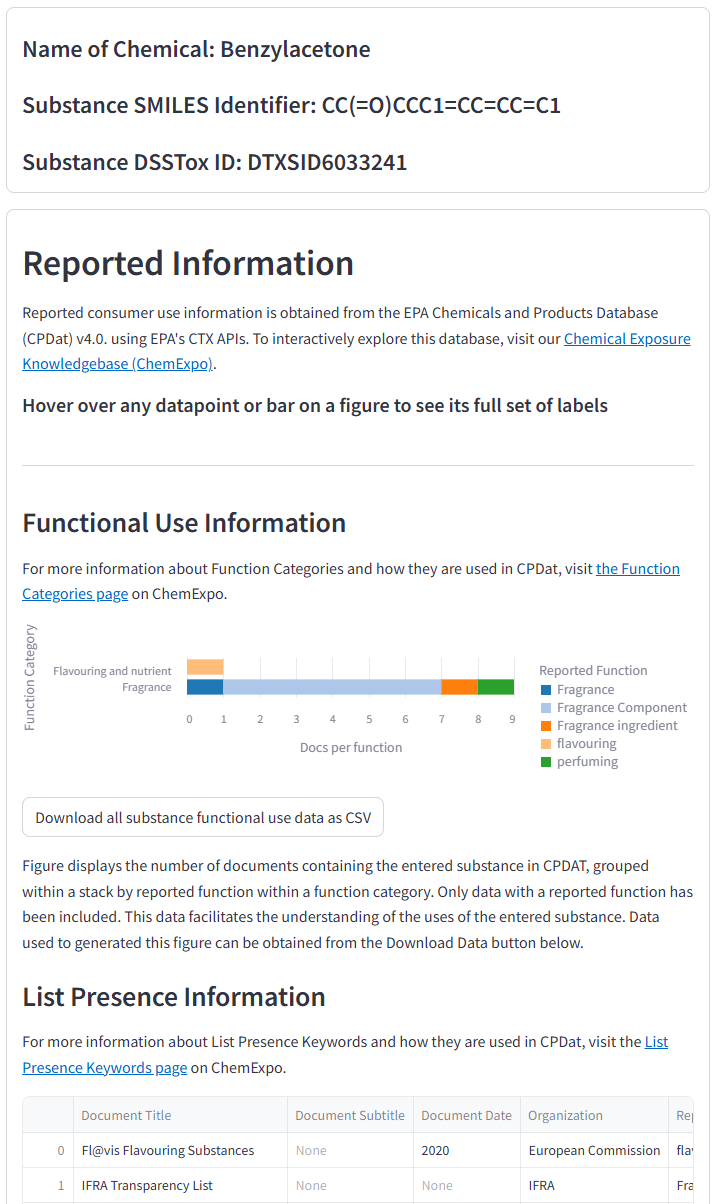


Figure 2: A snapshot of information returned by the Exposure Data Viewer.

### Chemical Identifier Search

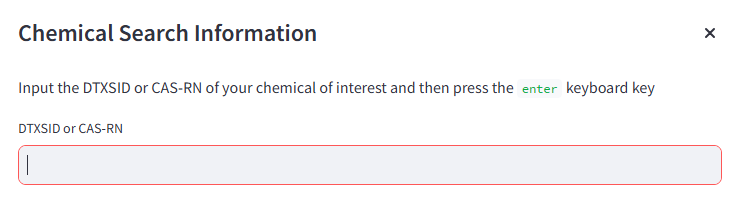
1. Launch the Exposure Data Viewer and navigate to the Chemical Exposure Information Search page.
2. From the Chemical Search Method drop-down menu select “DTXSID or CAS-RN”
3. Using the chemical identifier modal, draw the chemical structure of interest and click “Apply”
4. The Exposure Data Viewer will take a few seconds to query, retrieve, and then display the information relevant to the chemical, but once all data are displayed the textbox interface will close automatically, and the user can examine the displayed information.

Figure 3: Chemical identifier modal used to input DTXSID or CAS-RN for chemical search.

## Chemical Exposure Read-Across

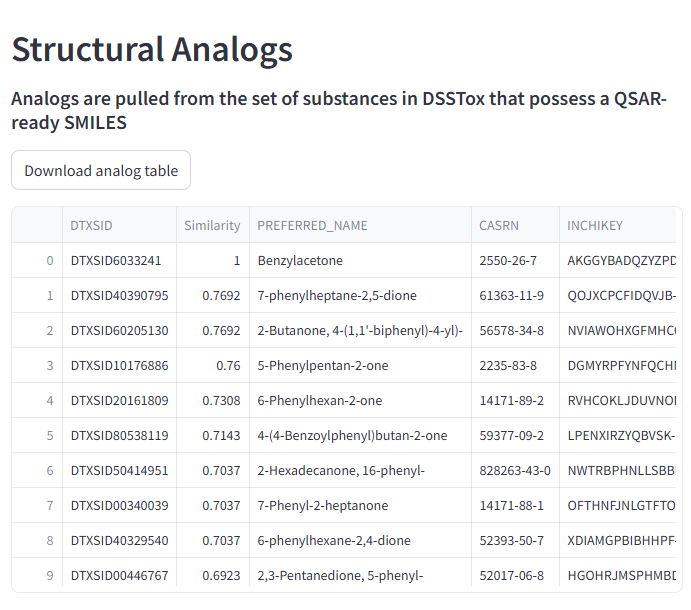
****The search functionality for the Chemical Exposure Read-Across page operates inthe exact same way as the Chemical Exposure Information Search Page. Once the chemical has been submitted information about structural analogs will be displayed as well as exposure-relevant information for those analogs. Depending on the technical specifications of the user’s computer, calculation of structural analogs may cause the user’s computer to briefly freeze.

Figure 4: Analog information for searched chemical.

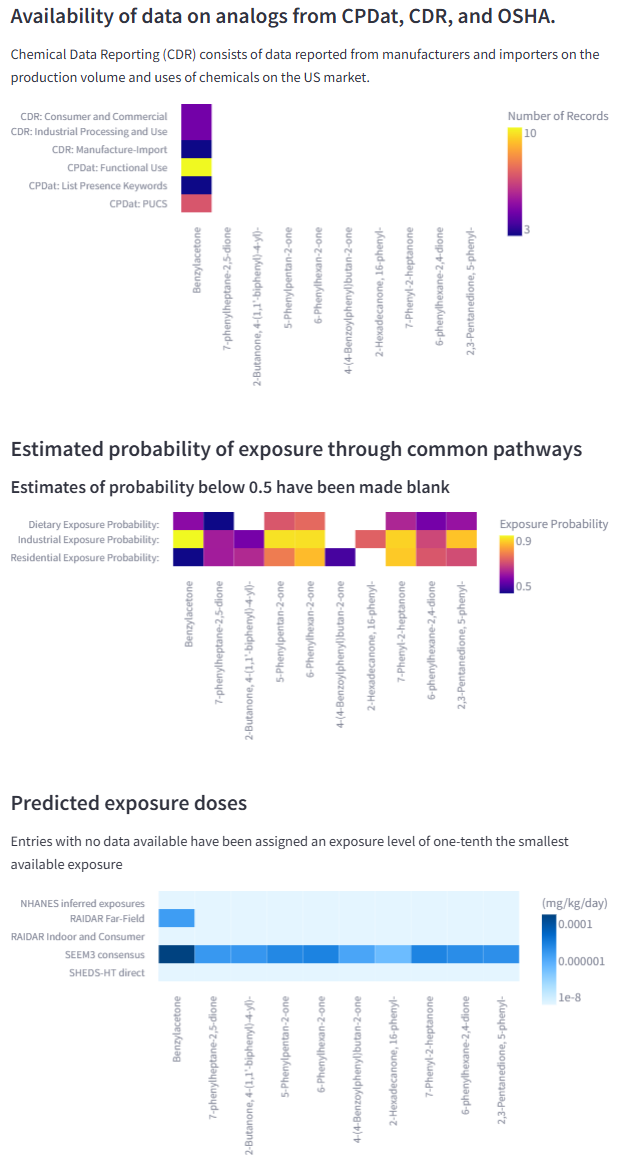
****

Figure 5: Exposure-relevant information displayed for analog chemicals.

# References

1. Dionisio, K. L. et al. The chemical and products database, a resource for exposure-relevant data on chemicals in consumer products. Scientific Data 5, 1–9 (2018).
2. The chemical and products database (CPDat) MySQL data file, version 3. https://doi.org/10.23645/epacomptox.5352997.v3 doi:10.23645/epacomptox.5352997.v3.
3. U.S. Occupational Safety and Health Administration, Chemical Exposure Health Data, <https://www.osha.gov/opengov/health-samples>, accessed 2019.
4. U.S. Environmental Protection Agency, Chemical Data Reporting, <https://www.epa.gov/chemical-data-reporting/access-chemical-data-reporting-data#2020>, access 2024.
5. Phillips, K. A., Wambaugh, J. F., Grulke, C. M., Dionisio, K. L. & Isaacs, K. K. High-throughput screening of chemicals as functional substitutes using structure-based classification models. Green Chemistry 19, 1063–1074 (2017).
6. Minucci, J. M., Purucker, T. S., Isaacs, K. K., Wambaugh, J. F., Phillips, K. A. A Data-Driven Approach to Estimating Occupational Inhalation Exposure Using Workplace Compliance Data. Environmental Science and Technology, 57, 5947—5956 (2023).
7. U.S. Environmental Protection Agency, Chemical Screening Tool for Exposures and Environmental Releases (ChemSTEER), <https://www.epa.gov/tsca-screening-tools/chemsteer-chemical-screening-tool-exposures-and-environmental-releases>, accessed 2019.
8. U.S. Environmental Protection Agency, Command-Line Occupational Exposure Tool (CLOET), <https://github.com/kaphillips/cloet>, accessed 2024.
9. Lowe, K., et al. Incorporating human exposure information in a weight of evidence approach to inform design of repeated dose animal studies, Regulatory Toxicology and Pharmacology, 127 (2021).
10. Wambaugh, J. F. et al. High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environmental science & technology 47, 8479–8488 (2013).
11. Wambaugh, J. F. et al. High throughput heuristics for prioritizing human exposure to environmental chemicals. Environmental science & technology 48, 12760–12767 (2014).
12. Ring, C. L. et al. Consensus modeling of median chemical intake for the U.S. Population based on predictions of exposure pathways. Environmental Science & Technology 53, 719–732 (2019).
13. Streamlit [Python package] (version 1.43.2). Retrieved from https://pypi.org/project/streamlit/.
14. ctx-python [Python package] (version 0.0.1a7). Retrieved from https://pypi.org/project/ctx-python/.