Public access to environmental chemistry data via the EPA CompTox Chemistry Dashboard

Antony Williams[1], Chris Grulke[1], Andrew D. McEachran[2], Rebecca Jolley[1], Jeremy Dunne[1], Elizabeth Edmiston[1] & Jeff Edwards[1]

[1]U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Research Triangle Park, NC

[2]Oak Ridge Institute of Science and Education (ORISE) Research Participant, Research Triangle Park, NC

ORCID: 0000-0002-2668-4821

Antony Williams I williams.antony@epa.gov I 919-541-1033

ACS Fall 2017
Washington, DC
August 20-24, 2017

Problem Definition and Goals

Problem: There is limited access online to freely available and highly curated data to support environmental science research.

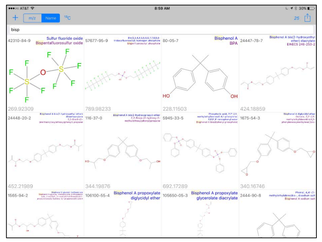
Goals: Provide access to a rich collection of data for ~760,000 chemical substances. Deliver the data via a simple to use web-based interface supporting diverse types of data associated with environmental chemistry, and specifically computational toxicology. Provide support for substances that can be represented as chemical structures as well as for Unknown or Variable Composition, Complex Reaction Products and Biological Materials (UVCB substances). Make the data available as downloadable Open data for repurposing and reuse in other applications.

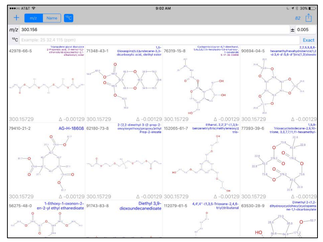
Abstract

The National Center for Computational Toxicology (NCCT) has assembled and delivered an enormous quantity and diversity of data for the environmental sciences through the CompTox Chemistry Dashboard. These data include high-throughput *in vitro* screening data, *in vivo* and functional use data, exposure models and chemical databases with associated properties.

The web application provides access to data associated with ~760,000 chemical substances and the data can be used to inform fate, exposure, hazard and risk assessment. Much of the data are downloadable and fully open for reuse, therefore allowing for integration into third party applications and databases.

Recently a mobile application was developed for the iOS platform (i.e. iPhone and iPad) that contains hundreds of thousands of chemicals for searching [1].

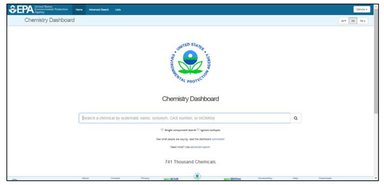




The Open data associated with the dashboard was reformatted for use in both iPhone and iPad applications.

This poster will provide an overview of the dashboard and its capabilities for delivering data to the environmental chemistry community.

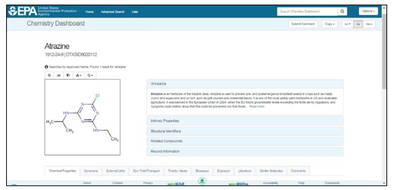
The CompTox Chemistry Dashboard



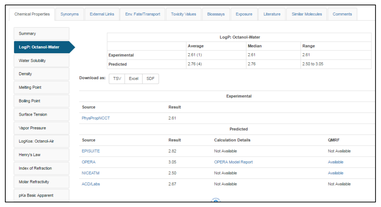
Dashboard Entry Page

The landing page of the dashboard is a simple text entry box allowing a type-ahead search for systematic, trade and trivial names, CAS Registry Numbers and InChIs.

Where possible, links are provided to related Wikipedia articles. An associated mol file is available for download to the desktop, and a summary report containing record data can be provided as a PDF file.



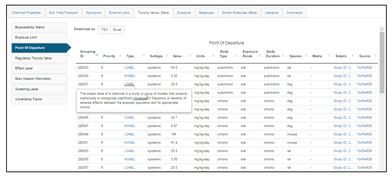
Chemical Record Page: Atrazine



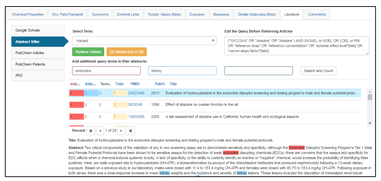
Chemical Properties Panel

For records with chemical structure representations, various inherent properties (e.g. formula and mass) and predicted physicochemical properties (logP, water solubility etc.) are provided.

The Toxicity Values tab provides access to data assembled from a series of public resources including EPA data (i.e. IRIS and PPRTV reports, ToxRef DB). Data can be downloaded as TSV and Excel files.



Toxicity Values Panel

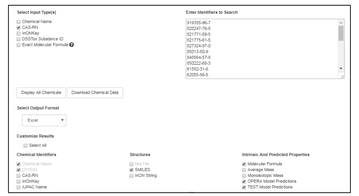


Literature: Pubmed Abstract Sifter

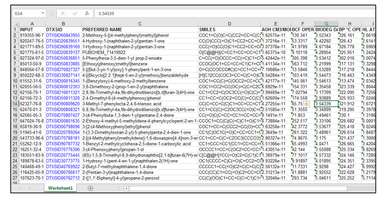
Literature searching using integration to a series of online resources can be performed using the CASRN and chemical name. This includes Google Scholar, PubMed and PubChem patents.

Batch Searching

Users of the dashboard may be interested in searching for collections of thousands of chemicals. The batch search allows inputs of CASRNs, chemical names, InChIKeys, DTXSIDs and exact molecular formulae. Data that can be downloaded includes SMILES, Molecular Formula and Mass (to support mass spectrometry analyses) and predicted OPEn saR Application (OPERA) [2] and Toxicity Estimation Software Tool (TEST) [3] predictions. The chemicals and associated data can be downloaded as tab-separated values, Excel or SDF files.



The Batch Search input page. Inputs include CASRN and Names. Results can be downloaded as Excel and SDF.



Downloaded Search Results in Excel format. DTXSIDs are hyperlinked to the substance page for easy viewing.

Future Work

- Continue to expand the data in terms of chemicals, toxicity data, additional experimental data.
- Provide the ability to perform structure, substructure and similarity searching of the database.
- Provide access to OPERA and TEST models as real-time online prediction tools.
- Add additional functionality supporting the display of bioassay data.

References

1. Mobile CompTox Dashboard App for both iPad and iPhone available at: https://itunes.apple.com/us/app/comptox-mobile/id1179517689
2. Mansouri *et al.* An automated curation procedure for addressing chemical errors & inconsistencies in public datasets used in QSAR modelling, *SAR QSAR Environ Res. 2016 Nov;27(11):939-965.*
3. EPA Toxicity Estimation Software Tool (T.E.S.T.) software http://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test

Future Work

The authors would like to acknowledge colleagues for their contributions to the development of the dashboard: Nancy Baker, Richard Judson, John Wambaugh, Jon Sobus, Kathie Dionisio and Katharine Phillips.