Accessing integrated chemistry and toxicology data via the US-EPA CompTox Chemicals Dashboard

Antony Williams, Chris Grulke, Richard Judson, Ann Richard, Katie Paul-Friedman, Jeremy Dunne and Jeff Edwards

U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, NC

SETAC 2019
Toronto, CAN
November 3-7, 2019

ORCID: 0000-0002-2668-4821
Antony Williams I williams.antony@epa.gov I 919-541-1033

Problem Definition and Goals

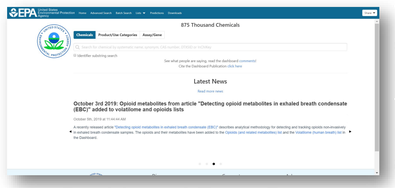
Problem: There is limited access online to freely available and highly curated data to support environmental science research.

Goals: Provide access to a rich collection of data for ~875,000 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 US-EPA Center for Computational Toxicity and Exposure (CCTE) has been generating data and building software applications and web-based chemistry databases for over a decade. During this period the center has analyzed thousands of chemicals in hundreds of bioassays, has researched high-throughput physicochemical property measurements and investigated approaches for high throughput toxicokinetics. Over the past few years some of the data have been delivered through prototype web-based “dashboards” for public consumption. The latest of these web applications, the CompTox Chemicals Dashboard, is an integrated access point to obtain information associated with 875,000 chemical substances and providing experimental and predicted data of various types. This includes physicochemical and fate and transport data, bioactivity data, exposure data and integrated literature searches. Real-time predictions and *generalized read-across* are possible and advanced search capabilities are available to support EPA-related projects including mass spectrometry non-targeted analysis. The dashboard is also an integration hub for ~70 public resources using cross website linking. This helps the user navigate to other data and information for a particular chemical on other websites. The application also provides integrated literature searches for chemical related information using Google Scholar and PubMed. A batch search also allows users to search using inputs of thousands of chemical names or CAS Registry Numbers and download details regarding the availability of bioassay, exposure and toxicity data. This poster provides an overview of the CompTox Chemicals Dashboard and the efforts to provide a seamless experience for both chemists and toxicologists. *This abstract does not necessarily represent the views or policies of the U.S. Environmental Protection Agency.*

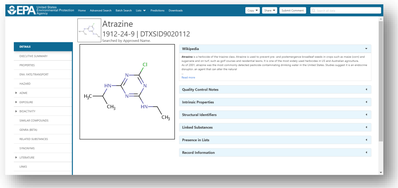
The CompTox Chemicals Dashboard



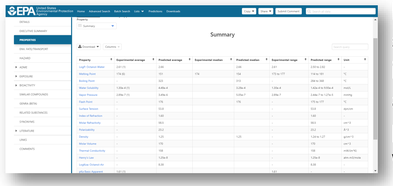
Dashboard Entry Page

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.



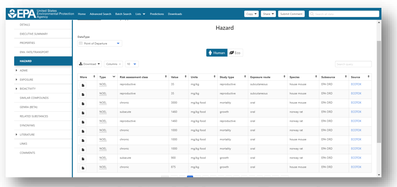
Chemical Record Page: Atrazine



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.

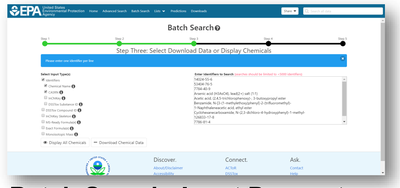


Toxicity Values Panel

Bioassays: ToxCast Data

The ToxCast Bioassay data measured over the past decade is accessible via the dashboard under the Bioassays Tab. PubChem bioassays data are also available. Data can be downloaded as Excel files.

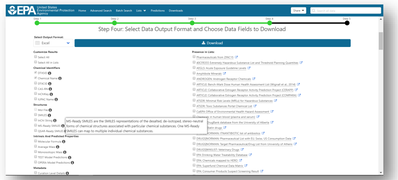
Batch Search



Batch searching has been delivered to allow a user to search for thousands of chemicals, using names, CAS Registry Numbers, InChI Keys, molecular formulae or masses.

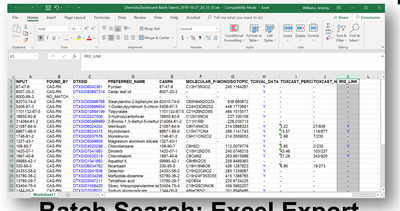
Batch Search: Input Parameters

The user can choose to export in multiple formats including Excel, TSV, SDF format. Multiple data types can be included – property data, hazard data (human and eco), bioassay data, identifiers etc.



Batch Search: Export Parameters

Excel files can be populated with various forms of data, from the multiple databases underpinning the dashboard, with a couple of button clicks. The data include hyperlinked data elements in the dashboard (highlighted in blue)



Batch Search: Excel Export

Future Work

- Continue to add chemicals and maintain updates to all integrated databases with every release. Next release is March 2020.
- Structure, substructure and similarity searching is already in testing
- Real-time predictions are already available but OPERA predictions will be added in the future

References

1. Williams *et al.* The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, J. Cheminformatics, 9, Article number: **61** (2017), DOI: 10.1186/s13321-017-0247-6
2. Grulke *et al.* EPA’s DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research, Computational Toxicology, **12**, (2019), 100096. DOI: 10.1016/j.comtox.2019.100096

Acknowledgements

The authors would like to acknowledge specific colleagues within our center for their contributions to the development of the dashboard: Todd Martin, Nancy Baker, John Wambaugh, Kristin Isaacs and Kathie Dionisio.