Assembly and curation of a list of per- and polyfluoroalkyl substances (PFAS) to support environmental science research

Antony J. Williams[1*], Linda Gaines[2], Charles Lowe[1], Inthirany Thillainadarajah[3], Brian Meyer[3], Grace Patlewicz[1] and Ann M. Richard[1]

1. Center for Computational Toxicology & Exposure, Office of Research & Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA
2. Office of Land and Emergency Management, US Environmental Protection Agency, Washington, DC, USA
3. Senior Environmental Employment Program, US Environmental Protection Agency, Research Triangle Park, USA

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

Society of Toxicology
San Diego, CA
March 27-31, 2022

United States Environmental Protection Agency

Problem Definition and Goals

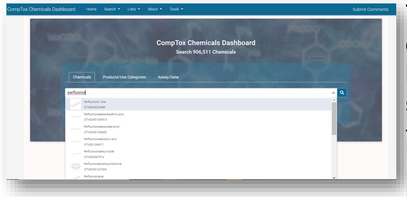
Problem: There are many sources of PFAS data online to support computational toxicology. However, curated datasets for the thousands of known PFAS chemicals are not available in structured formats,

Goals: Deliver online access to hundreds of thousands of chemicals of interest to environmental science and computational toxicology. Provide lists of PFAS substances via a simple to use web-based interface. Deliver application to support diverse types of data including experimental and predicted physicochemical properties, *in vivo* hazard data and *in vitro* toxicity and toxicokinetic data. Make the data available as downloadable data for reuse and repurposing in other databases.

Abstract

PFAS are a group of fluorinated substances that have generated increased public attention due to their potential health hazard and widespread presence in the environment. An attempt to define PFAS and establish standard categories was established in Buck et al[1]. Since then, various groups, including the OECD and EPA, have defined PFAS differently to serve their respective needs. The EPA's CompTox Chemicals Dashboard is a public resource providing access to hazard, bioactivity, property, and exposure data for ~900k substances, the majority with defined structures. We have dedicated efforts to the assembly and curation of thousands of PFAS substances. A set of more than 35 PFAS lists spanning EPA and international regulatory interests, as well as lists pertaining to research activities and analytical methods, are currently delivered via the Dashboard. Depending on the definition of what is deemed to be a PFAS, approximately 6,000 to upwards of 38,000 compounds currently in the Dashboard could be considered PFAS. We have defined an inventory of over 10700 PFAS structures using a set of clearly defined structural filters that encompass current EPA programmatic interests. This list is serving as a benchmark for the community to apply cheminformatics approaches and better understand and survey the scope of PFAS chemistries of potential concern. Such resources can help support the PFAs community to better frame the PFAS problem and aid in the development of categorization approaches for PFAS, as well as support the reporting of relevant data for the agency. *Abstract does not reflect EPA policy.*

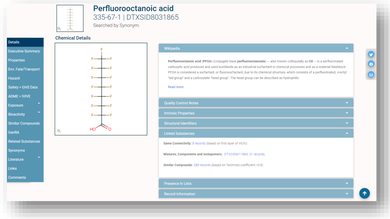
The CompTox Chemicals Dashboard



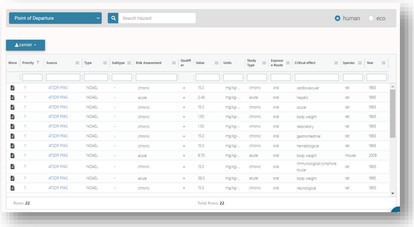
Dashboard Entry Page

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 InChI chemical identifiers.

Where possible, links are provided to related Wikipedia articles. Structure file formats are available for download to the desktop (SMILES and molfile) and an executive summary report regarding chemical toxicity is provided.



Chemical Record Page: PFOA



Toxicity Values Panel

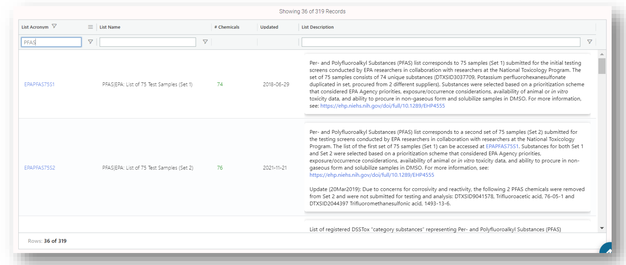
The Hazard tab provides access to data assembled from a series of public resources including EPA data (i.e., IRIS and PPRTV reports, ToxRef DB). Data can be downloaded as TSV and Excel files.

ToxCast Bioactivity data measured over the past decade are under the Bioassay Tab. Data can be downloaded as Excel files. New *in vitro* data are being generated on a library of ~150 PFAS in collaboration with NTP.

Bioactivities: e.g., ToxCast Data

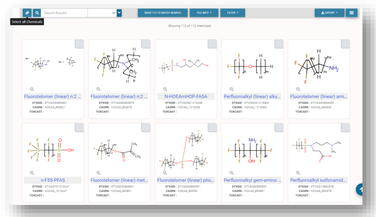
Accessing PFAS Chemical Lists

The dashboard provides access to ~35 individual PFAS chemical lists. These include a list based on structural filters, UVCB chemicals, PFAS categories represented as Markush structures, and regulatory lists.



List of chemical lists of PFAS chemicals: 36 lists and growing

A growing list of Markush structure representations is available. These structures can be enumerated and mapped to distinct chemical structures as members of the categories.



Future Work

- *In vitro* toxicity and toxicokinetic measurements have been measured for ~150 PFAS. These will be released in the future on the dashboard.
- Experimental property data are being harvested from literature and online resources to include into the dashboard and OPERA models.
- Chemical categorization efforts continue in order to be encompassing of more of the PFAS library.

References

1. Perfluoroalkyl and polyfluoroalkyl substances in the environment: Terminology, classification, and origins, Buck et al. https://doi.org/10.1002/ieam.258
2. Assembly and curation of lists of per- and polyfluoroalkyl substances (PFAS) to support environmental science research, Williams et al. https://www.frontiersin.org/articles/10.3389/fenvs.2022.850019
3. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, Williams et al. https://doi.org/10.1186/s13321-017-0247-6

Acknowledgements

The authors thank the chemical curation team for their rigorous work and the software development team for the development of the dashboard.