Accessing information for Per- & Polyfluoroalkyl Substances (PFAS) using the US EPA CompTox Chemicals Dashboard

Antony J. Williams*, Christopher M. Grulke, Grace Patlewicz and Ann M. Richard
U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology (NCCT), Research Triangle Park, NC

1181/P206
Society of Toxicology Meeting
Baltimore, MD
March 10-14, 2019

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

United States Environmental Protection Agency

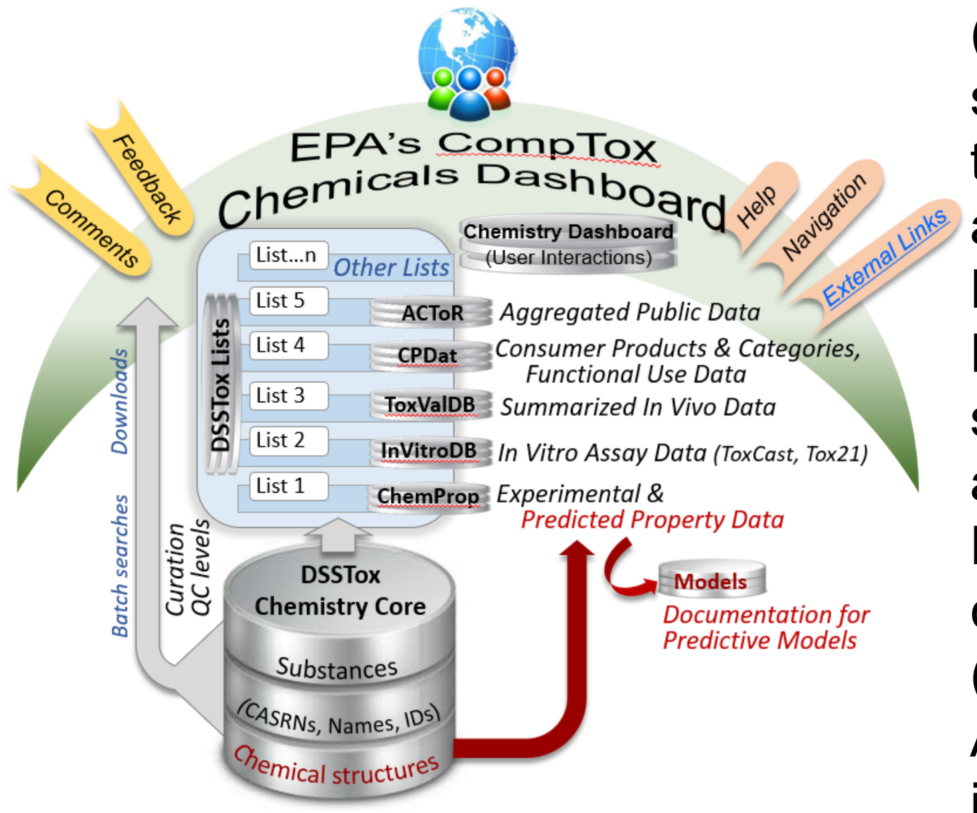
Abstract

EPA's National Center for Computational Toxicology is developing automated workflows for curating large databases within the DSSTox project and for providing accurate linkages of data to chemical structures, exposure and hazard information. The data are made available via the EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard), a publicly accessible website providing access to data for ~875,000 chemical substances, the majority of these represented with chemical structures. The web application delivers a wide array of computed and measured physicochemical properties, *in vitro* high-throughput screening data and *in vivo* toxicity data, and linkages to a growing list of literature, toxicology, and analytical chemistry websites. The application provides access to lists of chemicals of specific interest to EPA and stakeholders. One area of high current interest is that for Per- & Polyfluoroalkyl Substances (PFAS). Several PFAS lists are sourced from EPA, the European Union and the OECD, with the total exceeding 5000 substances. A procured testing library of hundreds of PFAS chemicals, with a portion of the list annotated into structure-based categories, has been integrated into the Dashboard. Resulting benefits include a searchable database of chemical properties, hazard and exposure predictions and links to the open literature. This poster provides an overview of the Dashboard, the extensive library of PFAS chemicals and associated categories, and efforts underway to develop new physicochemical property and environmental fate and transport QSAR prediction models for these chemicals.

Problem Definition and Goals

Problem: High quality chemical structure information in association with chemical lists of interest to the environmental chemistry community is required for data aggregation, structure-based indexing, read-across, and development of QSAR prediction models. However, structural information is often unavailable for classes of chemicals of current and emerging interest. A prime example is that of PFAS substances. EPA undertook an effort to procure several hundred PFAS chemicals to support development of analytical methods, environmental monitoring and toxicity testing and to register and structurally annotate these and several public lists of PFAS chemicals in the DSSTox database. A further need within this chemical domain was to create usable structure-based category definitions. The goal was to employ manual curation and cheminformatics approaches to make these lists, the categories, and associated data (e.g. hazard, physicochemical, fate and transport) publicly available and searchable within the Dashboard to serve as a consolidated resource for the toxicology community.

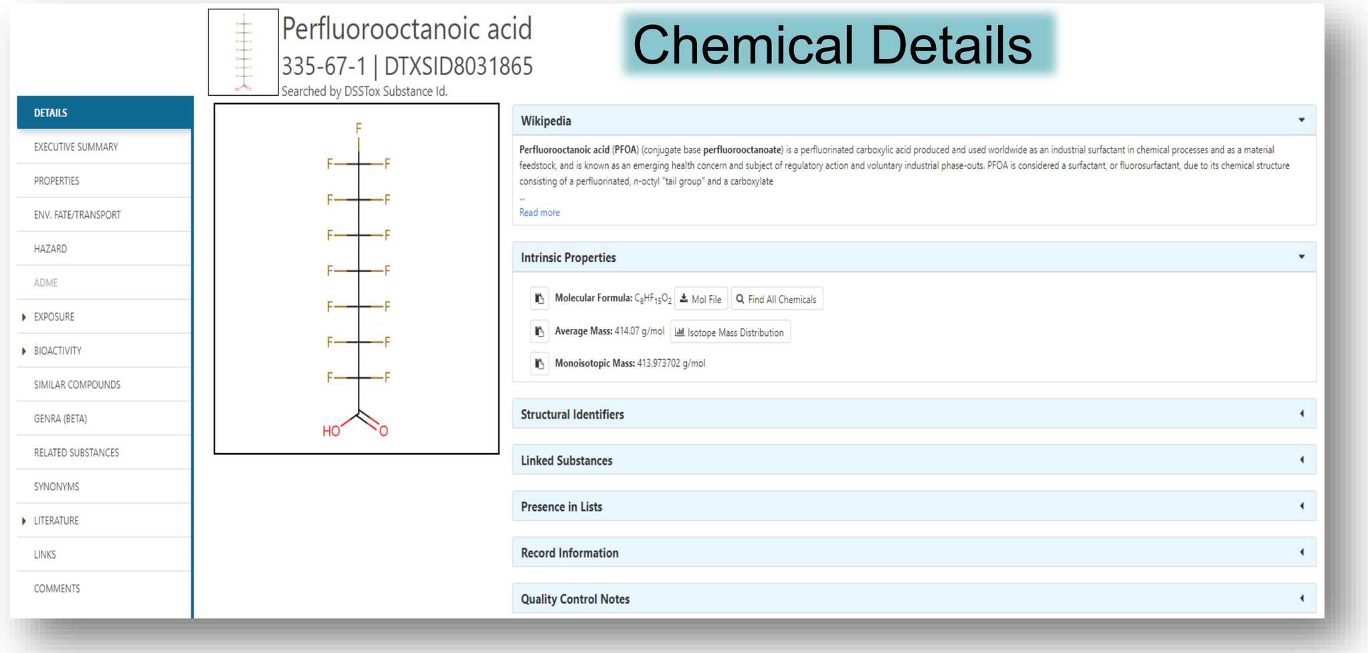
Source Data and Downloadable Forms

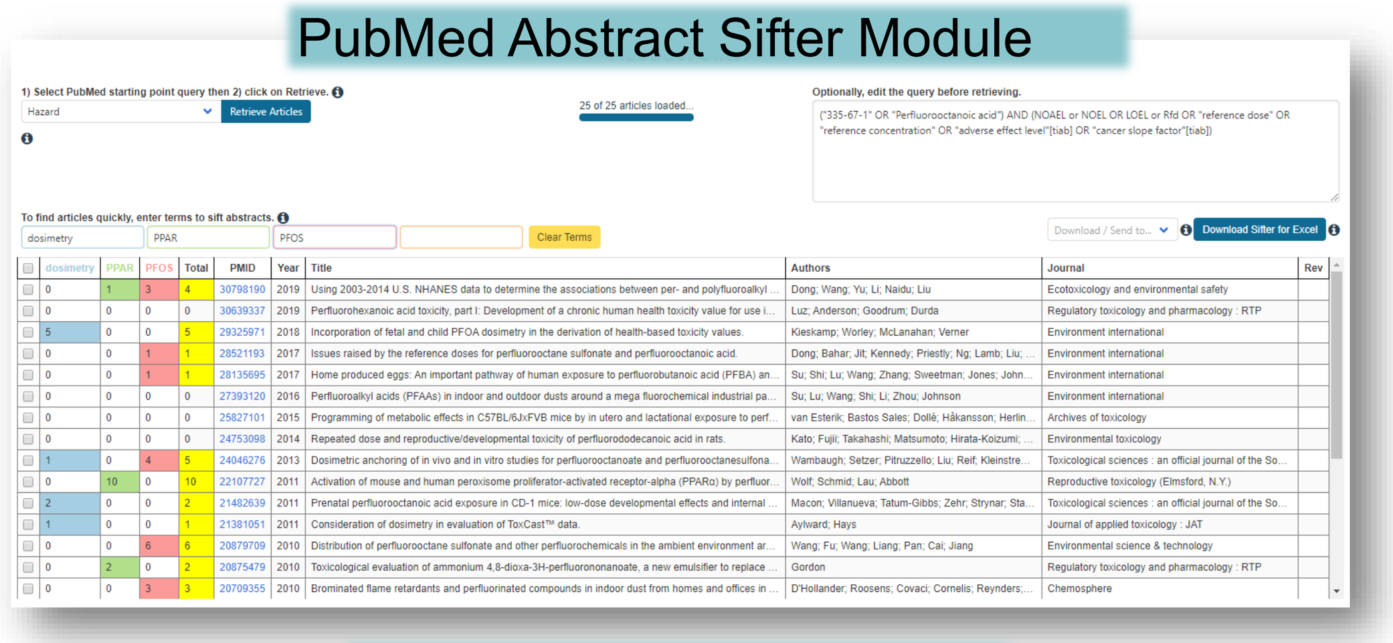


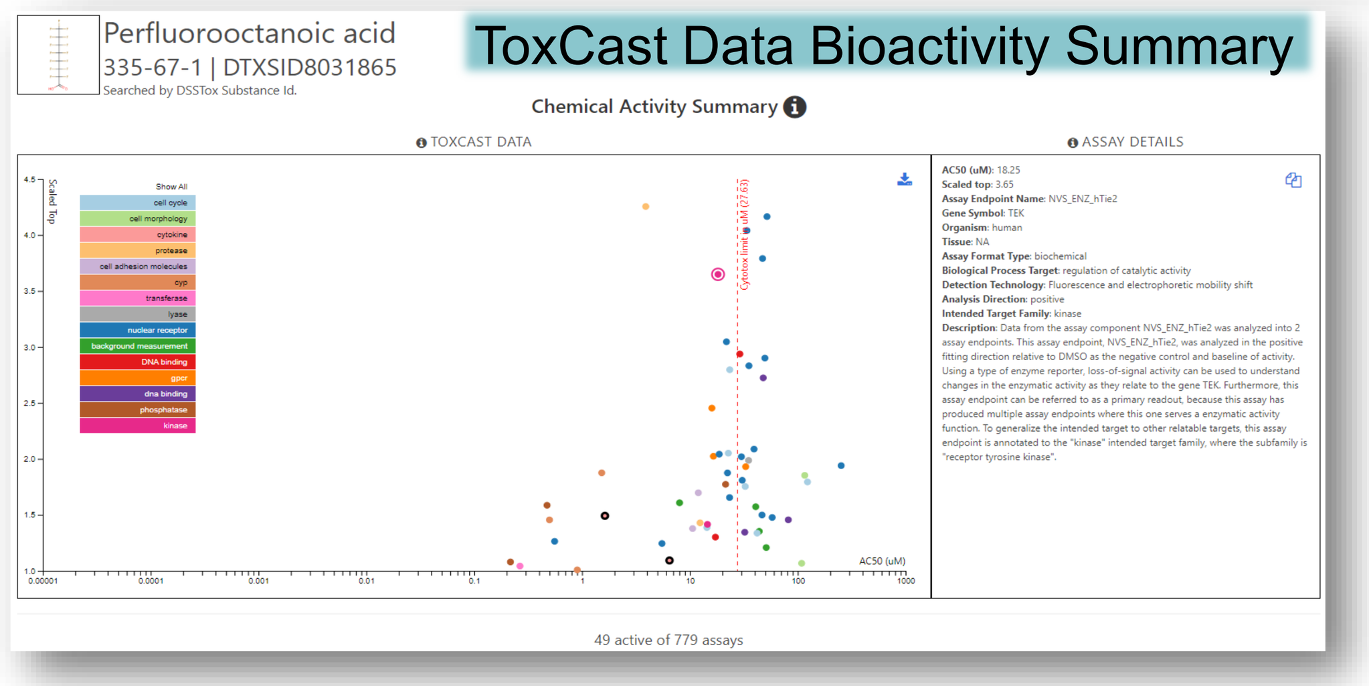
PFAS lists consisting of names and CAS Registry Numbers (CASRN) were assembled from various public and EPA sources, including: the Swedish Chemicals Agency report on the occurrence and use of highly fluorinated substances and alternatives[1], the OECD Comprehensive Global Database of PFAS chemicals[2], literature articles and public databases. PFAS lists were loaded into the DSSTox chemical registration system, and chemicals were curated over a period of months and assigned to one of five quality curation levels[3]. A master PFAS list page enables users to navigate to individual collections of PFAS chemicals (https://comptox.epa.gov/dashboard/chemical_lists/pfasmaster). All PFAS lists are downloadable in easy-to-consume formats, including Excel spreadsheets and SDF files containing structural layouts and associated data.

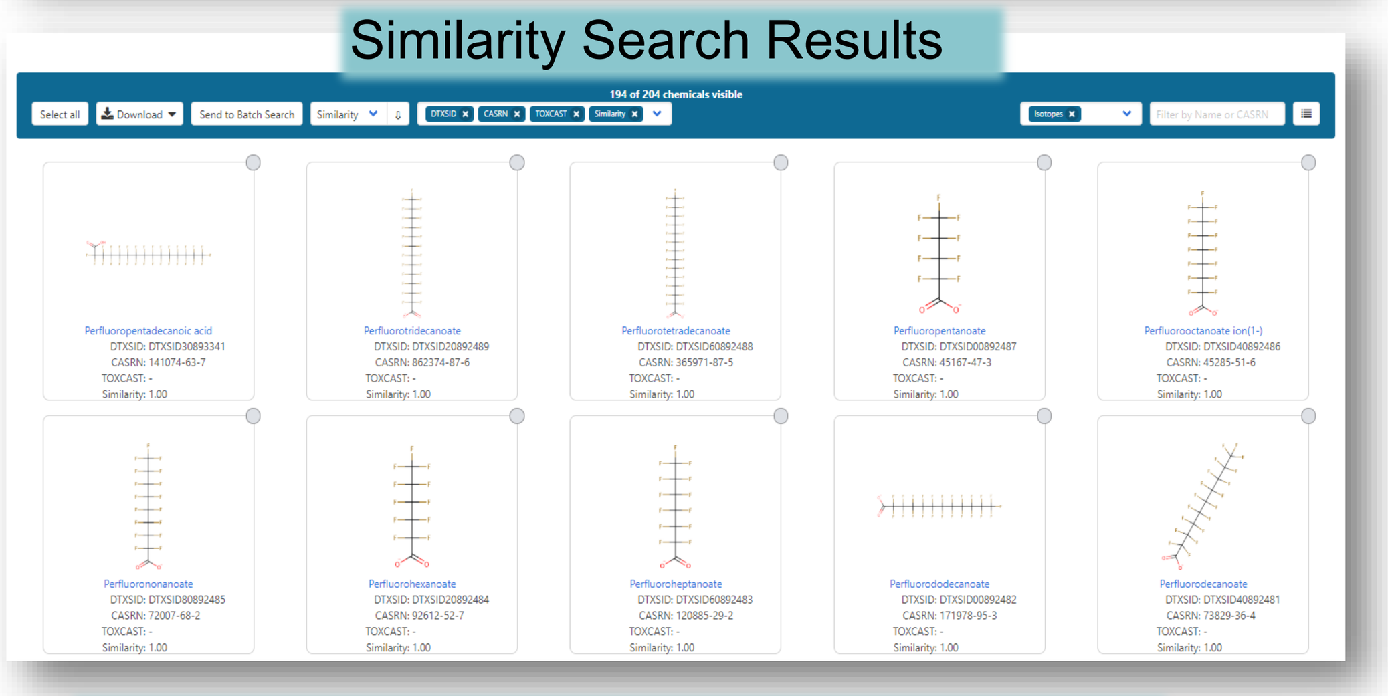
CompTox Chemicals Dashboard

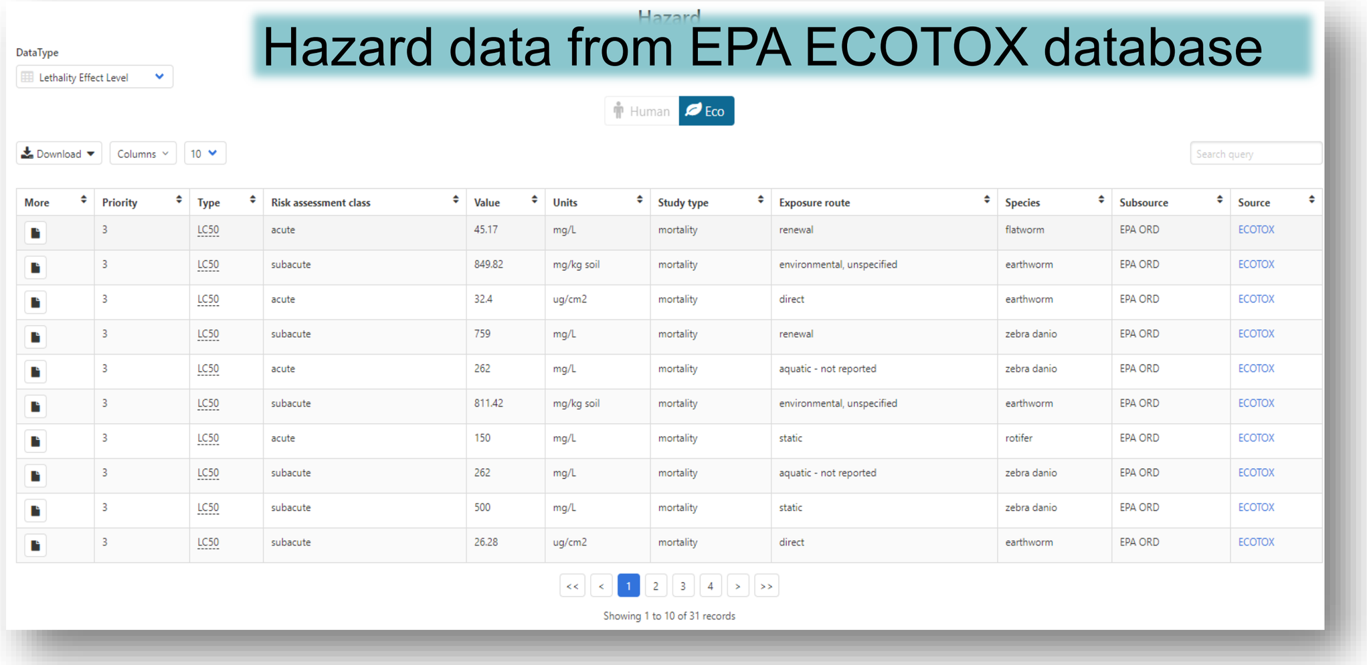
The CompTox Chemicals Dashboard provides an intuitive web-based interface for navigating various types of data associated with a list of ~875,000 DSSTox chemical substances. Data include, where available, experimental and predicted property data (including OPERA QSAR models[4]), hazard data (from the underlying EPA ToxVal database), exposure data (including product and use categories, and exposure predictions) and *in vitro* bioactivity data associated with ToxCast and Tox21 screening. Other modules include a literature module that accesses PubMed for abstract-sifting[5], a Generalized Read-Across (GenRA) module[6], and a batch-search capability for searching up to thousands of chemical names, CASRN or other chemical identifiers (InChI Strings or Keys). The screenshots below illustrate some capabilities across PFAS space.

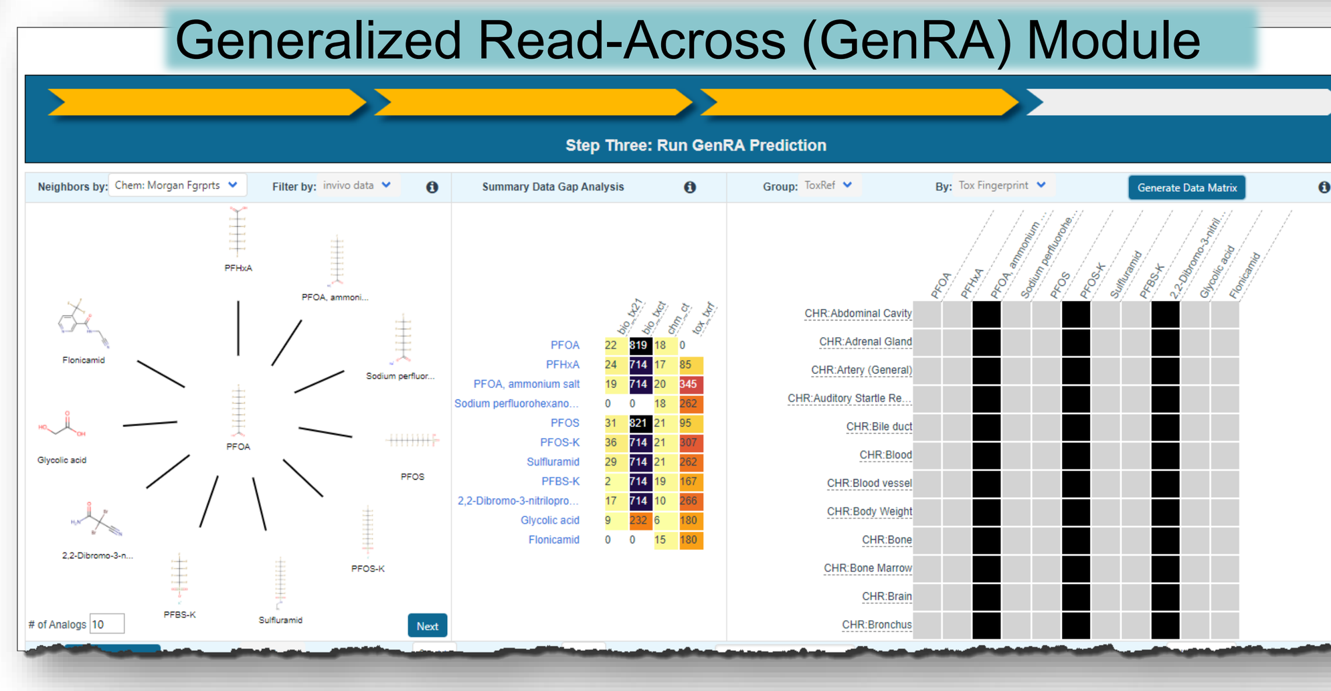




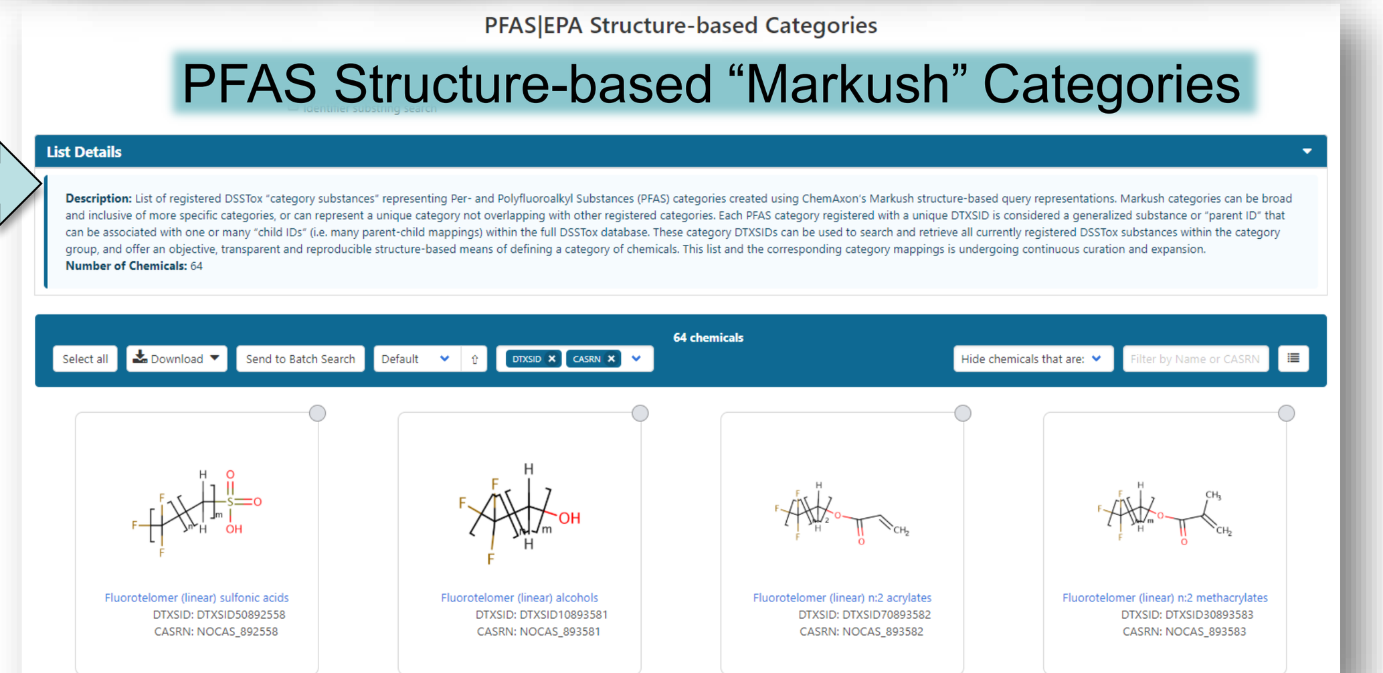








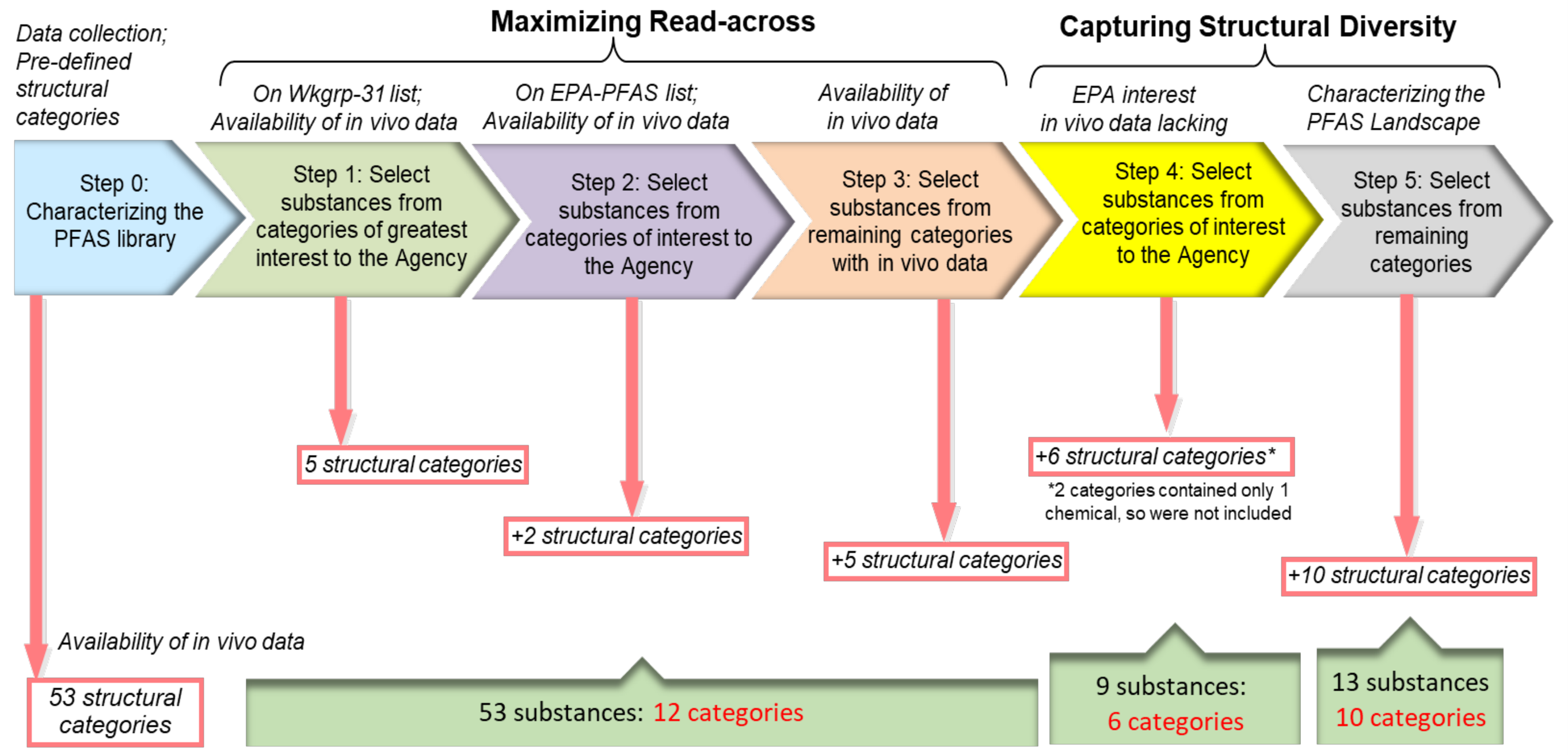
A list of registered PFAS "category substances" was created using Markush structure-based query representations. These categories can be broad and inclusive of more specific categories, or can represent a unique category not overlapping with other registered categories. This approach offers an objective, transparent and reproducible structure-based means of defining a category of chemicals.



Chemical Category-Based Prioritization Approach for Selecting PFAS Chemicals for High-throughput Testing

The development of a PFAS screening library and the process by which a subset of 75 PFAS substances were selected for testing in new approach methods was recently reported[7]. Absent chemical catalogs for PFAS, an initial scoping for potentially procurable PFAS relied on the use of DSSTox PFAS structure lists. Commercial availability of candidate PFAS, together with computed properties, knowledge of existing *in vivo* data and inclusion in specific categories provided useful information to support the expert review selection process. The two prioritized sets of 75 PFAS chemicals currently undergoing testing (EPAPFAS75S1) or slated for testing (EPAPFASSET2) are available on the dashboard.

Future PFAS candidates will be selected from the EPA's complete, DMSO solubilized PFAS inventory (EPAPFASINV), now totaling 430 unique DSSTox substances. This inventory and the list of 43 PFASs procured but found DMSO-insoluble (EPAPFASINSOL) as well as the final EPAPFAS75S1 and EPAPFAS75S2 lists are available for download at https://comptox.epa.gov/dashboard/chemical_lists/pfasmaster.



Acknowledgements

We thank the NCCT software development team for their development of the Dashboard and underlying infrastructure, Stellan Fischer from the Swedish Chemicals Agency for his support with the KEMI list of chemicals and Kamel Mansouri for including new PFAS data in the OPERA model results. Special thanks to Inthirany Thillainadarajah and Brian Meyer, for expert review and structure-annotation of the PFAS chemical lists compiled for this study.

References

1. Appendix 2 from the Swedish Chemicals Agency Report on PFAS Chemicals - https://www.kemi.se/en/global/rapporter/2015/report-7-15-occurrence-and-use-of-highly-fluorinated-substances-and-alternatives.pdf
2. OECD New Comprehensive Global Database of PFASs available from http://www.oecd.org/chemicalsafety/portal-perfluorinated-chemicals/
3. Williams *et al.,* The CompTox Chemicals Dashboard, Journal of Cheminformatics, 2017, 9:61, https://doi.org/10.1186/s13321-017-0247-6
4. Mansouri *et al.,* OPERA models for predicting physicochemical properties and environmental fate endpoints, Journal of Cheminformatics 2018, 10:10, https://doi.org/10.1186/s13321-018-0263-1
5. Baker *et al.,* Abstract Sifter: a comprehensive front-end system to PubMed, F1000Research 2017, 6(Chem Inf Sci):2164, https://doi.org/10.12688/f1000research.12865.1
6. Helman et al., Generalized Read-Across (GenRA): A workflow implemented in the EPA CompTox Chemicals Dashboard, ALTEX, 2019, https://doi.org/10.14573/altex.1811292
7. Patlewicz *et al.,* A Chemical Category-Based Prioritization Approach for Selecting 75 Per- and Polyfluoroalkyl Substances (PFAS) for Tiered Toxicity and Toxicokinetic Testing, Environmental Health Perspectives 127(1) January 2019, https://doi.org/10.1289/EHP4555

Innovative Research for a Sustainable Future