

# Integrating Mass Spectrometry Non-Targeted Analysis and Computational Toxicology to Characterize Chemicals

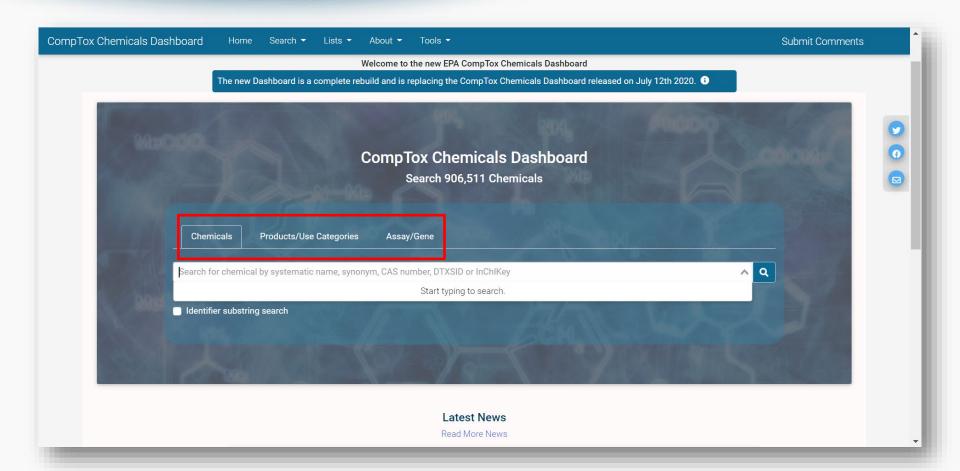
## **Antony Williams**

Center for Computational Toxicology and Exposure, US-EPA, RTP, NC

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

## CompTox Chemicals Dashboard >906k chemicals





## The Charge for the Dashboard

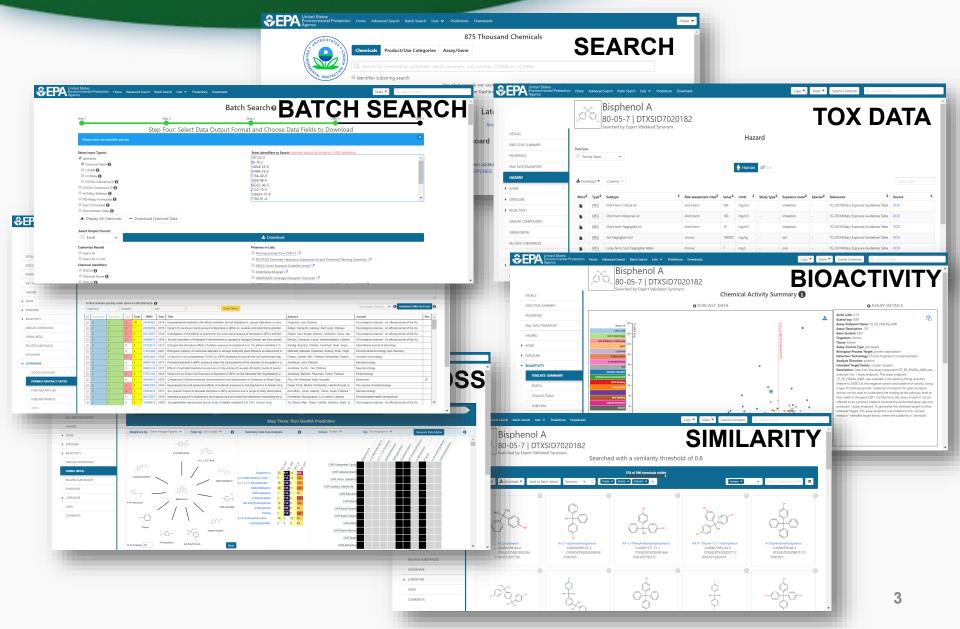


- Develop a "first-stop-shop" for environmental chemical data to support EPA and partner decision making:
  - Centralized location for relevant chemical data
  - Chemistry, exposure, hazard and dosimetry
  - Combination of existing data and predictive models
  - Publicly accessible, periodically updated, curated
- Easy access to data improves efficiency and ultimately accelerates chemical risk assessment

## CompTox Chemicals Dashboard

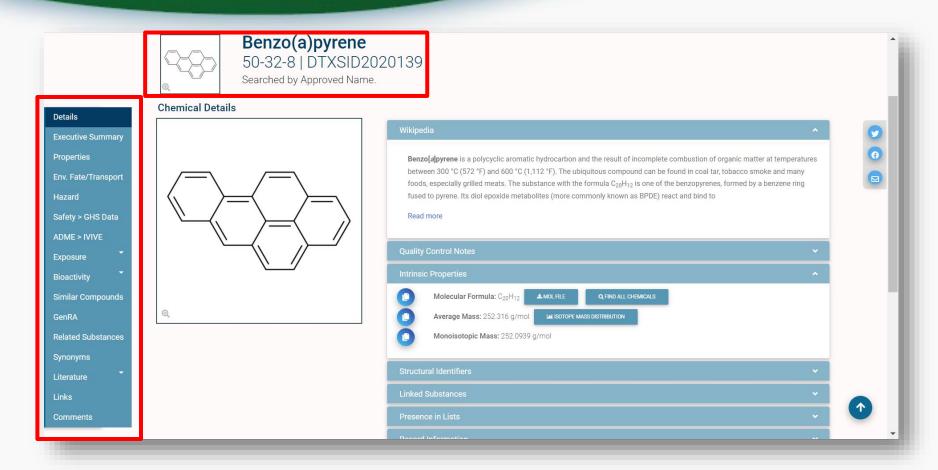






## **Detailed Chemical Pages**

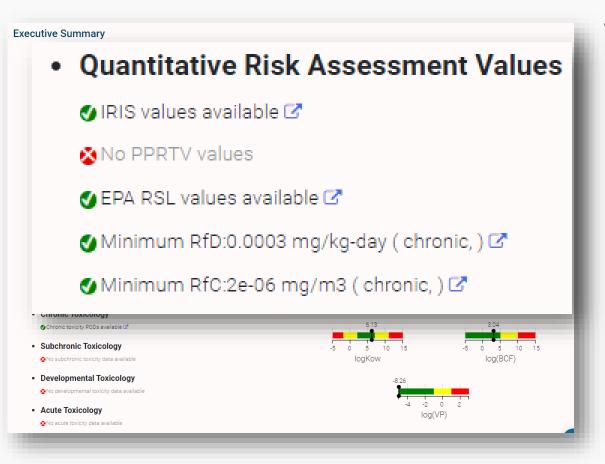




 Chemical page: Wikipedia snippet when available, intrinsic properties, structural identifiers, linked substances

## "Executive Summary"

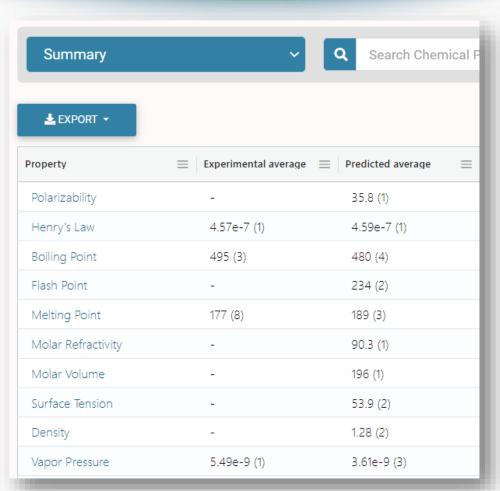




- Overview of toxicityrelated info
  - Quantitative values
  - Physchem. and Fate & Transport
  - Adverse Outcome Pathway links
  - In vitro bioactivity summary plot

## Experimental and Predicted Data





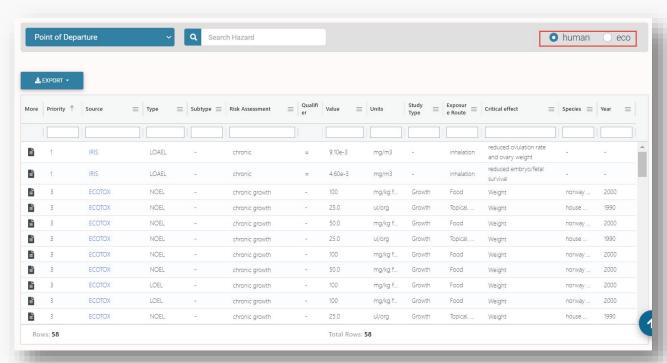
- Physchem and Fate & Transport experimental and predicted data
- Data can be downloaded as Excel, TSV and CSV files

## **Chemical Hazard Data**



#### ToxVal Database

- >50k chemicals
- >770k tox. values
- >30 sources of data
- ~5k journals cited
- ~70k citations



## Safety Data



#### **GHS Data**

Print Page

PUBCHEM > BENZO[A]PYRENE > LABORATORY CHEMICAL SAFETY SUMMARY (LCSS) > GHS CLASSIFICATION

CID 2336

#### Benzo[a]pyrene

#### **GHS Classification**

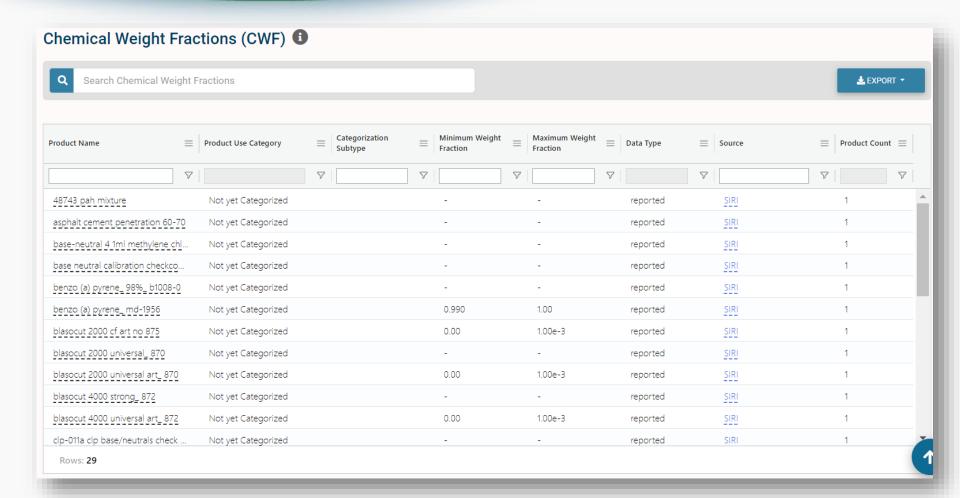
(?

#### Showing 6 of 6

Pictogram(s)	Irritant Health Hazard Environmental Hazard
Signal	<u>Danger</u>
	H317: May cause an allergic skin reaction [Warning Sensitization, Skin]
	H340: May cause genetic defects [Danger Germ cell mutagenicity]
GHS Hazard Statements	H350: May cause cancer [Danger Carcinogenicity]
	H360FD: May damage fertility; May damage the unborn child [Danger Reproductive toxicity]
	H400: Very toxic to aquatic life [Warning Hazardous to the aquatic environment, acute hazard]
	H410: Very toxic to aquatic life with long lasting effects [Warning Hazardous to the aquatic environment, long-term hazard]
Precautionary Statement	P201, P202, P261, P272, P273, P280, P281, P302+P352, P308+P313, P321, P333+P313, P363, P391, P405, and P501
Codes	(The corresponding statement to each P-code can be found at the GHS Classification page.)

## Sources of Exposure to Chemicals





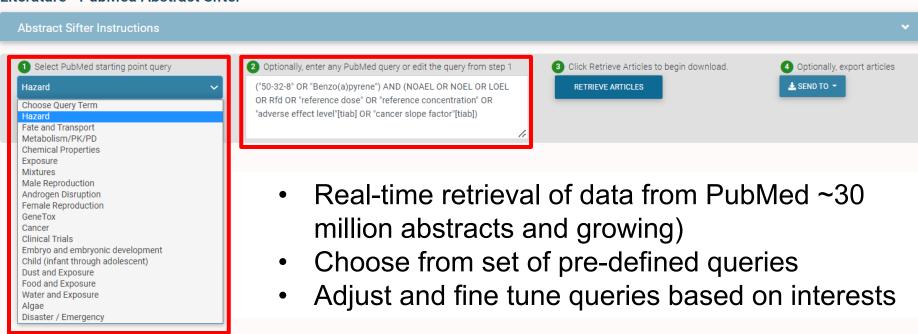


# Searching Literature and the Internet

## Literature Searching



#### Literature - PubMed Abstract Sifter



## Markup of abstracts



To find articles quickly, ent		metabol		uacis.	Clear Terms			Download / Send to V Download	d Sifter for I	Excel	0	
	Atrazine	metabol	Total	PMID	Year	Title Authors	J	Journal	R	Rev	Rev	_
	19	10	29	16527233	2006	Determination of atrazine and its metabolites in mouse urine and	Ross; Filipov		Analytical biochemistry			
	13	10	23	16656648	1967	Atrazine metabolism and herbicidal selectivity.	Shimabukuro		Plant physiology			
	13	10	23	8481106	1993	In vitro percutaneous absorption and metabolism in man of 2-chl	Ademola; Sedik; Wester; Maibach		Archives of toxicology			
	19	3	22	33421427	2021	Transcriptomic profiling of atrazine phytotoxicity and comparativ	Qu; Mei; Liu; Zhao; Liu; Li; Huang; Zhu		Environmental research			
	19	3	22	23102724	2012	Fate of atrazine in switchgrass-soil column system.	Albright; Murphy; Anderson; Coats		Chemosphere			
	15	7	22	11476505	2001	Anaerobic degradation of atrazine and metolachlor and metaboli	Seybold; Mersie; McNamee		Journal of environmental quality			
	16	5	21	20830925	2010	Enhanced degradation and soil depth effects on the fate of atrazi Krutz; Shaner, Zablotowicz			Journal of environmental quality			
	16	5	21	14674556	2003	Infiltration and adsorption of dissolved atrazine and atrazine met	ion of dissolved atrazine and atrazine met Krutz; Senseman; Dozier; Hoffman; Tierney		Journal of environmental quality			
	19	2	21	2761262	1989	Testosterone metabolism in neuroendocrine organs in male rats	Babić-Gojmerac; Kniewald; Kniewald		Journal of steroid biochemistry			
	20	0	20	33254405	2020	Removal of atrazine in catalytic degradation solutions by microal	Hu; Xu; Sun; Zhu; Sun; Zhao; Hu		Ecotoxicology and environmental safety			
	17	3	20	24062064	2013	Biodegradation of atrazine by Rhodococcus sp. BCH2 to N-isopr	Kolekar; Phugare; Jadhav		Environmental science and pollution research	h intern		
	18	2	20	21121649	2010	Metabolism and persistence of atrazine in several field soils with	Jablonowski; Hamacher; Martinazzo; Langen	n; Köpp	Journal of agricultural and food chemistry			
	20	0	20	18848368	2008	Nitrogen limited biobarriers remove atrazine from contaminated	Hunter; Shaner		Journal of contaminant hydrology			
	20	0	20	16595379	2006	Mixed-effect models for evaluating multiple measures of atrazine	Hines; Deddens; Lu; Fenske; Striley		Journal of occupational and environmental hy	ygiene		
	12	8	20	16349478	2005	Molecular basis of a bacterial consortium: interspecies catabolis	de Souza; Newcombe; Alvey; Crowley; Hay;	Sadow	Applied and environmental microbiology			
	11	8	19	16656991	1968	Atrazine metabolism in resistant corn and sorghum	Shimabukuro		Plant physiology			

#### Determination of atrazine and its metabolites in mouse urine and plasma by LC-MS analysis.

Atrazine is a herbicide widely used on agricultural commodities. Existing analytical methods to analyze atrazine and its metabolites in biological matrices have various drawbacks. Thus, further development of such methods will be needed to correlate the growing number of toxicological effects associated with atrazine exposure with the concentrations of this compound and its metabolites in plasma, urine, and tissues. The purpose of this study was to develop a broad and sensitive LC-MS method for the analysis of atrazine and its metabolites in mouse urine and plasma. We were able to simultaneously measure atrazine and its metabolites, which include didealkyl atrazine, desisopropyl atrazine, desethyl atrazine, desisopropyl atraz

## What's the best way to search the internet for chemical data?

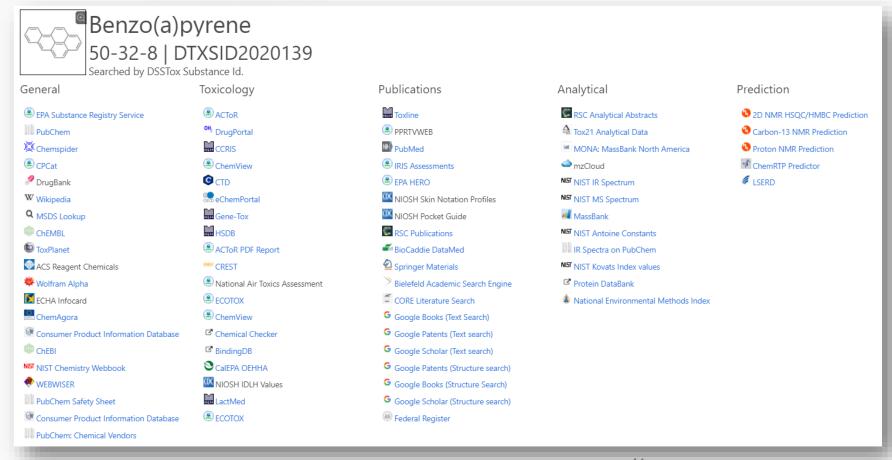


- We know how complex identifiers are...
  - CASRN(s)
  - Hundreds of names (maybe)
  - SMILES
  - InChis
  - EINECS, EC numbers

What can WE do to help navigate the internet?

## External Links – Also use Identifiers Names, CASRN, PubChem IDs, InChls...

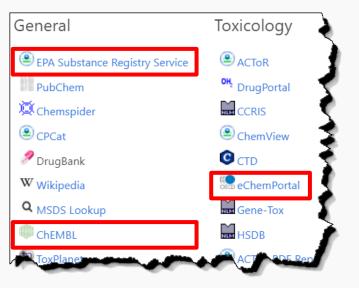


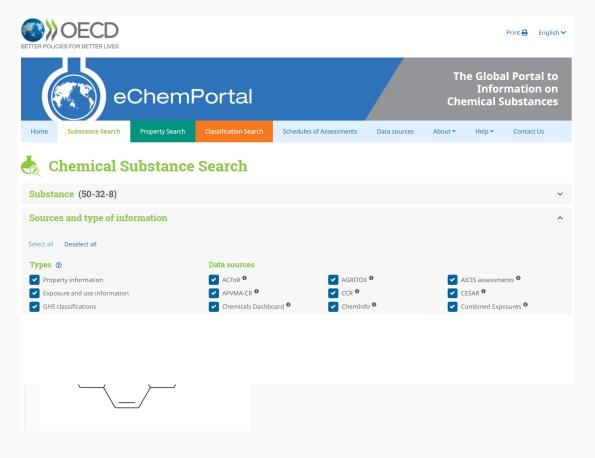


## **External Links**



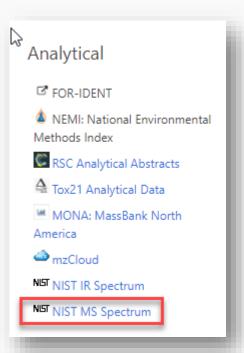
 Links to ~90 websites providing access to additional data on the chemical of interest

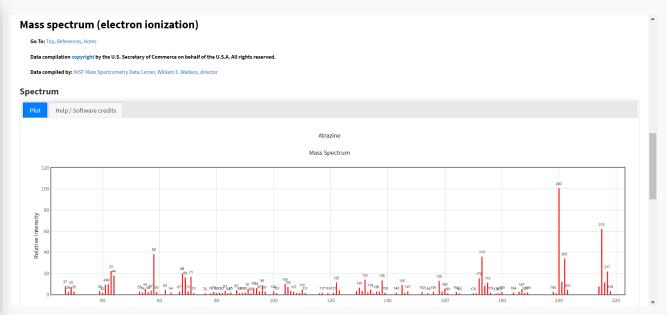




## NIST WebBook https://webbook.nist.gov/chemistry/

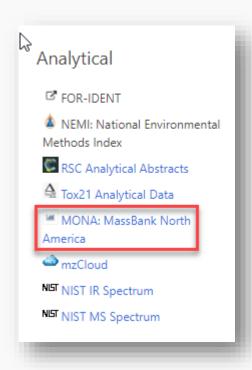


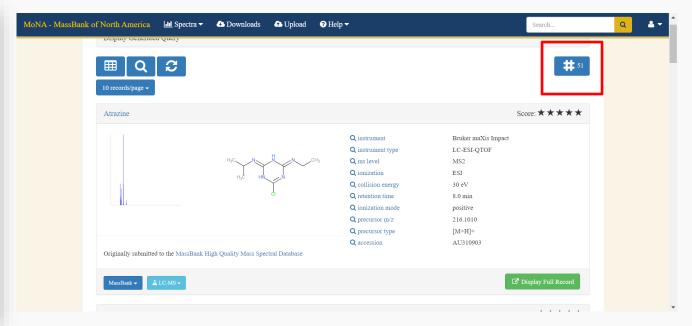




## MassBank of North America https://mona.fiehnlab.ucdavis.edu









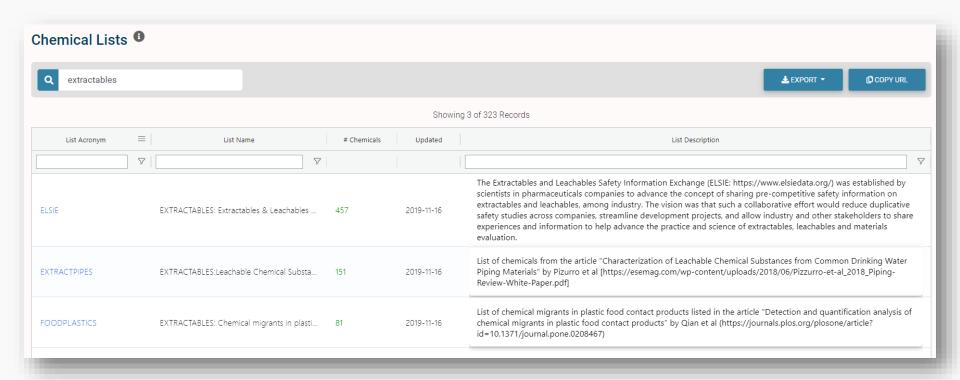
## Chemical Lists of Interest...

### Lists of Extractables and Leachables

https://comptox.epa.gov/dashboard/chemical-lists?filtered=&search=extractables



- Chemical lists with extractables & leachables
- Expands with literature extraction



### The ELSIE Database







# EXTRACTABLES: Extractables & Leachables Safety Information Exchange (ELSIE) Bearch for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey Start typing to search. Identifier substring search List Details Description: The Extractables and Leachables Safety Information Exchange (ELSIE: https://www.elsiedata.org/) was established by scientists in pharmaceuticals companies to advance the concept of sharing precompetitive safety information on extractables and leachables, among industry. The vision was that such a collaborative effort would reduce duplicative safety studies across companies, streamline development projects, and allow industry and other stakeholders to share experiences and information to help advance the practice and science of extractables, leachables and materials evaluation. Number of Chemicals: 457

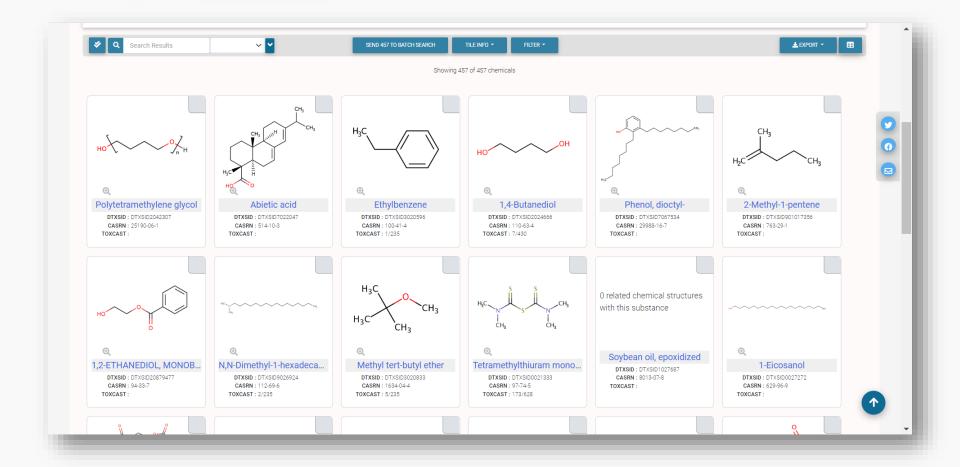
## Extractables



### **EXTRACTABLES:**Leachable Chemical Substances from Common Drinking Water Piping Materials Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey Start typing to search. Identifier substring search **List Details** Description: List of chemicals from the article "Characterization of Leachable Chemical Substances from Common Drinking Water Piping Materials" by Pizurro et al [https://esemag.com/wpcontent/uploads/2018/06/Pizzurro-et-al\_2018\_Piping-Review-White-Paper.pdf] Number of Chemicals: 151 **EXTRACTABLES: Chemical migrants in plastic food contact products** Q Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey Start typing to search. Identifier substring search **List Details** Description: List of chemical migrants in plastic food contact products listed in the article "Detection and quantification analysis of chemical migrants in plastic food contact products" by Qian et al (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0208467) Number of Chemicals: 81

## Download data, send to batch search for data harvesting







# "MS-ready" structures

McEachran et al. J Cheminform (2018) 10:45 https://doi.org/10.1186/s13321-018-0299-2

Journal of Cheminformatics

#### **METHODOLOGY**

**Open Access** 

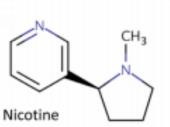
"MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies

Andrew D. McEachran<sup>1,2\*</sup>, Kamel Mansouri<sup>1,2,3</sup>, Chris Grulke<sup>2</sup>, Emma L. Schymanski<sup>4</sup>, Christoph Ruttkies<sup>5</sup> and Antony J. Williams<sup>2\*</sup>

## Overview of MS-Ready Structures

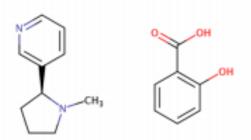


- All structure-based chemical substances are algorithmically processed to
  - Split multicomponent chemicals into individual structures
  - Desalt and neutralize individual structures
  - Remove stereochemical bonds from all chemicals
- MS-Ready structures are then mapped to original substances to provide a path between chemicals detected by mass spectrometry to original substances



CN1CCC[C@H]1C1=CN=CC=C1 DTXSID1020930| SNICXCGAKADSCV 54-11-5 | **162.1157** | 0.929 | **72** 

Tox: yes | Expo: yes | Bioassay: yes

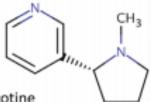


Benzoic acid, 2-hydroxy-, compd. with 3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)

OC(=O)C1=C(O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1

DTXSID5075319 | AIBWPBUAKCMKNS

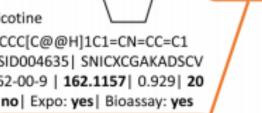
29790-52-1 | 300.1474 | 0.929 | 6 Tox: no | Expo: yes | Bioassay: no

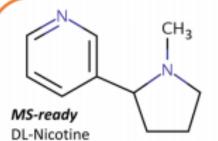


D-Nicotine

CN1CCC[C@@H]1C1=CN=CC=C1 DTXSID004635 | SNICXCGAKADSCV 25162-00-9 | 162.1157 | 0.929 | 20

Tox: no | Expo: yes | Bioassay: yes



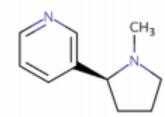


CN1CCCC1C1=CN=CC=C1 DTXSID3048154 | SNICXCGAKADSCV 22083-74-5 | 162.1157 | 0.953 | 9

Tox: yes | Expo: no | Bioassay: yes

LEGEND: Name, SMILES DTXSID | InChlKey 1st Block

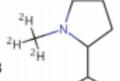
CAS | Monoiso. Mass | logP | Sources Data on: Toxicity | Exposure | Bioassays



HCI

Nicotine hydrochloride

Cl.CN1CCC[C@H]1C1=CN=CC=C1 DTXSID602093 | HDJBTCAJIMNXEW 2820-51-1 | 198.0924 | 0.929 | 9 Tox: no | Expo: yes | Bioassay: yes



DL-Nicotine-d3

[2H]C([2H])([2H])N1CCCC1C1=CN=CC=C1 DTXSID80442666| SNICXCGAKADSCV 69980-24-1 | 165.1345 | 0.929 | 1

Tox: no | Expo: no | Bioassay: no



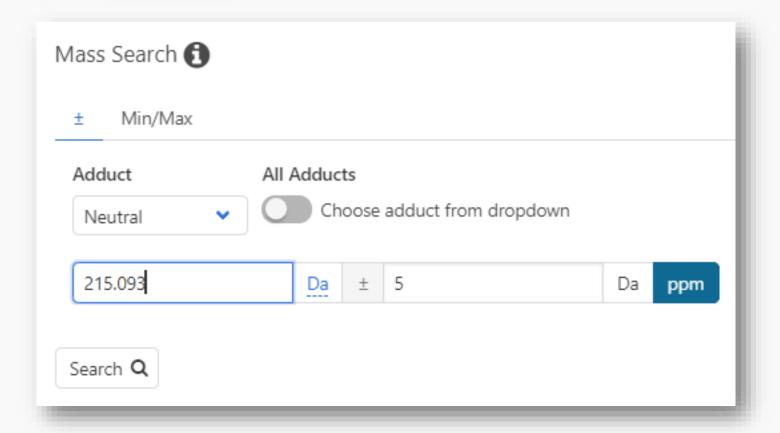




# Mass and Formula Searching

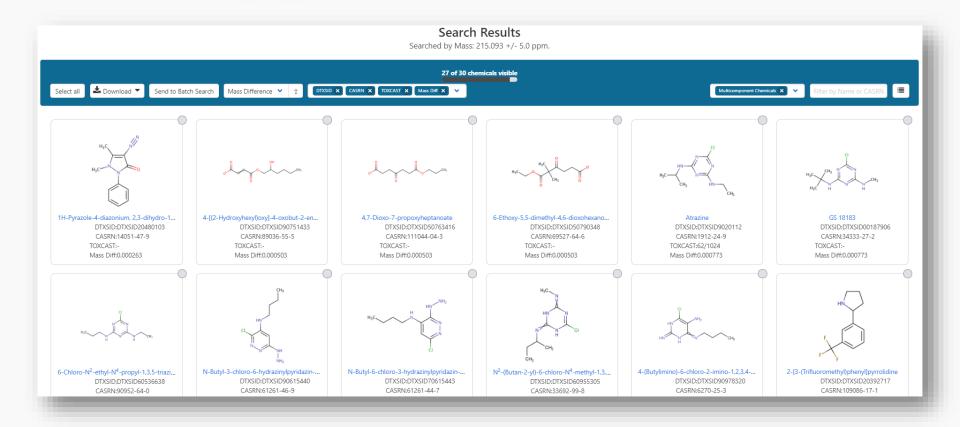
### Advanced Searches: Mass Search





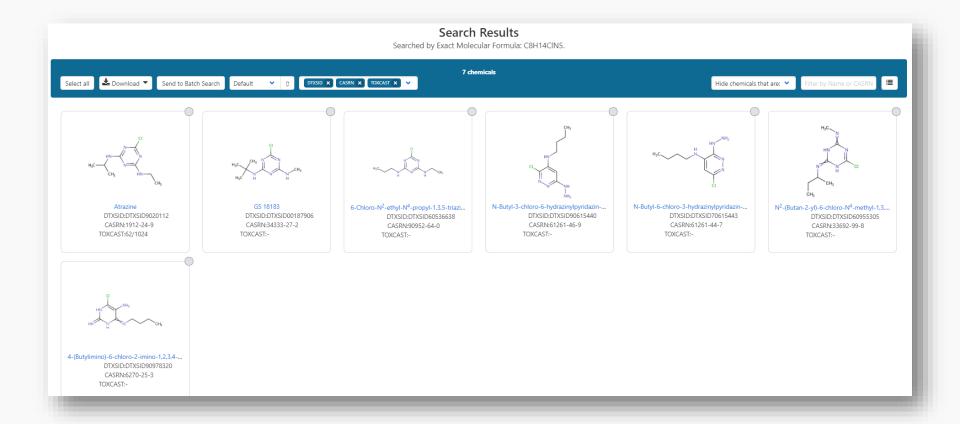
### Advanced Searches: Mass Search





## Advanced Searches: Formula Search







## Batch Searching

## Batch Searching



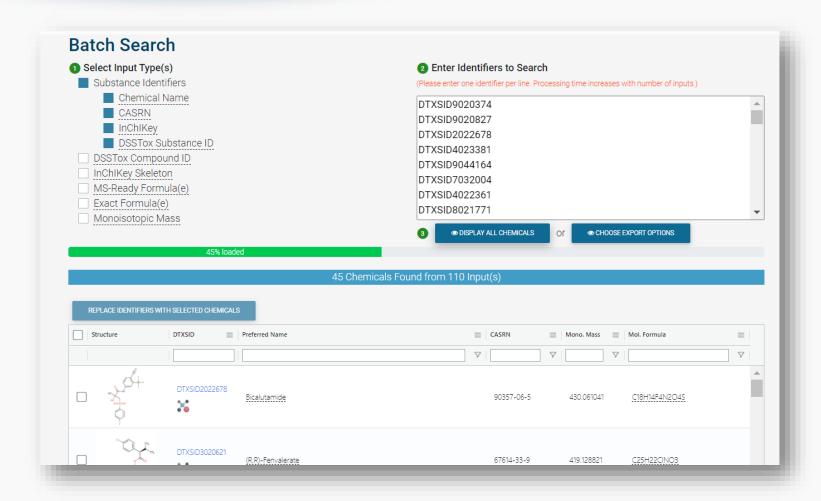
 Singleton searches are useful but we work with thousands of masses and formulae!

## Typical questions

- What is the list of chemicals for the formula C<sub>x</sub>H<sub>y</sub>O<sub>z</sub>
- What is the list of chemicals for a mass +/- error
- Can I get chemical lists in Excel files? In SDF files?
- Can I include properties in the download file?

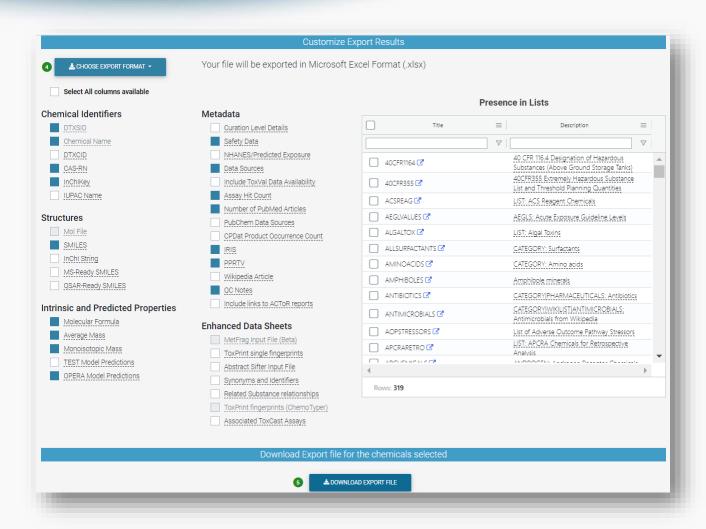
## Batch Search





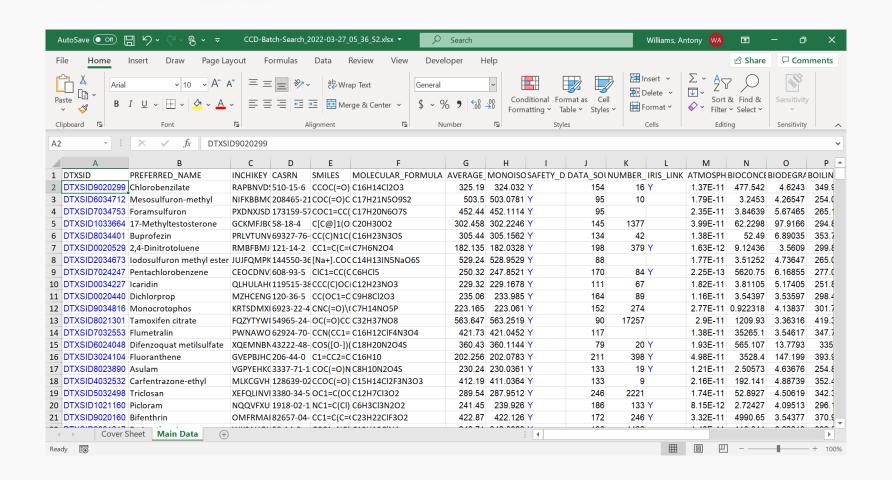
## Batch Search – Excel, CSV, SDF file





### **Batch Search**





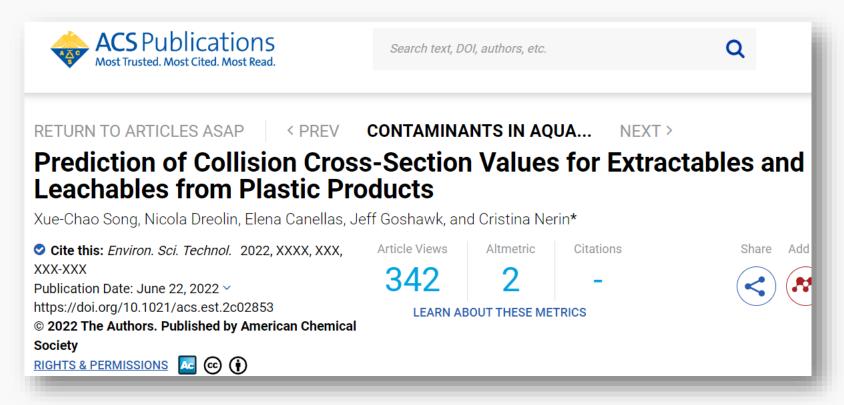


# Work in Progress

# New lists added regularly In progress



 New chemicals registered daily and released with each new version of the Dashboard



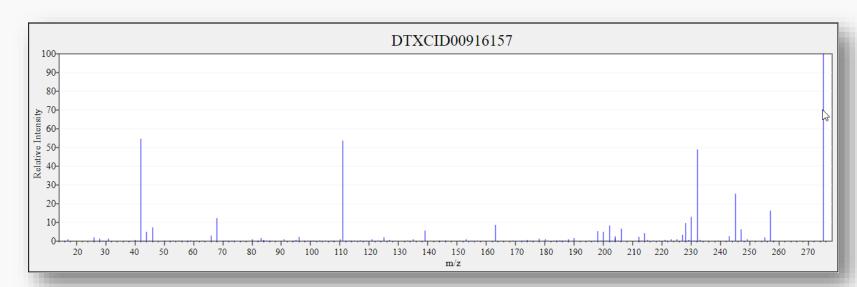
#### Predicted Mass Spectra

http://cfmid.wishartlab.com/



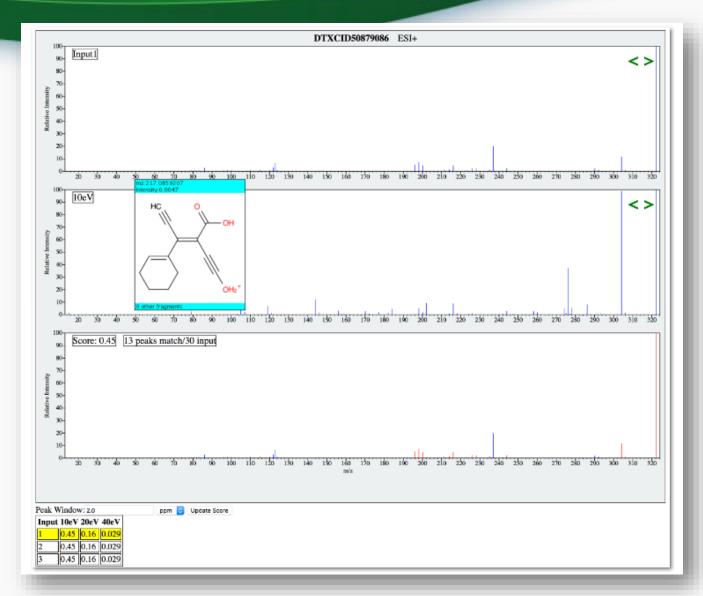


- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



### Spectral Viewer Comparison





### Already published...



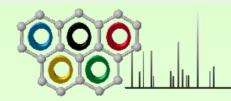
Analytical and Bioanalytical Chemistry https://doi.org/10.1007/s00216-019-02351-7

## In silico MS/MS spectra for identifying unknowns: a critical examination using CFM-ID algorithms and ENTACT mixture samples

Alex Chao <sup>1,2</sup> • Hussein Al-Ghoul <sup>1,2</sup> • Andrew D. McEachran <sup>1,3</sup> • Ilya Balabin <sup>4</sup> • Tom Transue <sup>4</sup> • Tommy Cathey <sup>4</sup> • Jarod N. Grossman <sup>2,3</sup> • Randolph Singh <sup>1,5</sup> • Elin M. Ulrich <sup>2</sup> • Antony J. Williams <sup>6</sup> • Jon R. Sobus <sup>2</sup>

Received: 4 October 2019 / Revised: 27 November 2019 / Accepted: 11 December 2019 © The Author(s) 2019







#### **CASMI 2017**

Important Dates
Contest Rules
Challenge Data
Solutions
Preliminary results
Results
About the Team

**CASMI 2016** 

**CASMI 2014** 

**CASMI 2013** 

**CASMI 2012** 

## Critical Assessment of Small Molecule Identification

The experimental and computational mass spectrometry communities are invited to participate in the fifth round of an open contest on the identification of small molecules from mass spectrometry data.

This year the contest will test the applicability of MS and MS/MS on natural products chemistry identifications. With 45 (Category 1) and up to 243 (Categories 2&3) natural products challenges - including a few tricky ones - there's something for everyone!

CASMI 2017 is organised by Dr. Dejan Nikolic (University of Illinois at Chicago, USA), Dr. Nir Shahaf (Weizmann Institute of Science, Israel), Dr. Emma Schymanski (Eawag, Switzerland) and Dr. Steffen Neumann (IPB Halle, Germany).

#### **Mailing lists**

BL LILL CACATE III LILL CACATE







Article

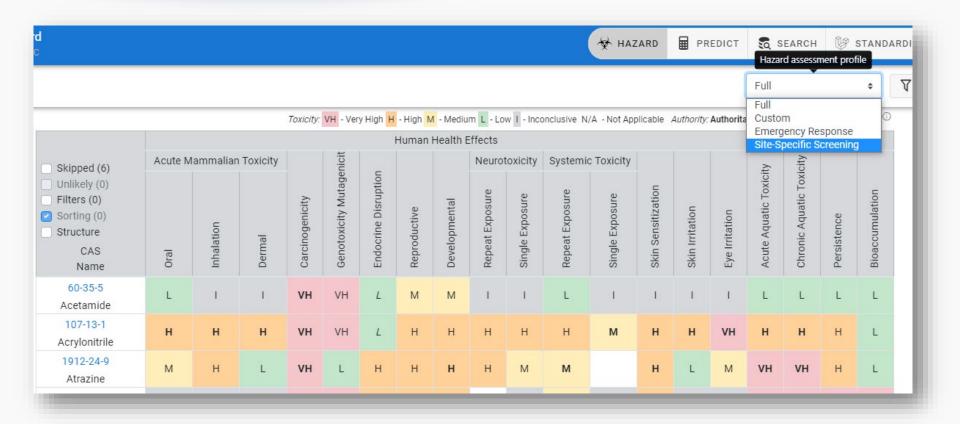
## Revisiting Five Years of CASMI Contests with EPA Identification Tools

**Table 2.** Percentage of the total number of compounds from each CASMI contest year that were ranked in the top 5 by Competitive Fragmentation Modeling for Metabolite Identification (CFM-ID) only and by the summation of CFM-ID and DSSTox Data Source Counts (DS), alongside the percentage in the top 5 reported by the contest years' winning entry. Complete ranking results are provided in Supplemental File S1.

CASMI Year	CFM-ID Only	CFM-ID + DS	Winners' Results <sup>1</sup>	Total in DB/Total in Dataset <sup>2</sup>
2012	36%	64%	36%	14/14
2013	81%	88%	88%	16/16
2014	57%	76%	71%	42/42
2016-training	63%	96%		312/312
2016-challenge	66%	94%	81%	208/208
2017	59%	53%	$74\%$ $^{3}$	227/243

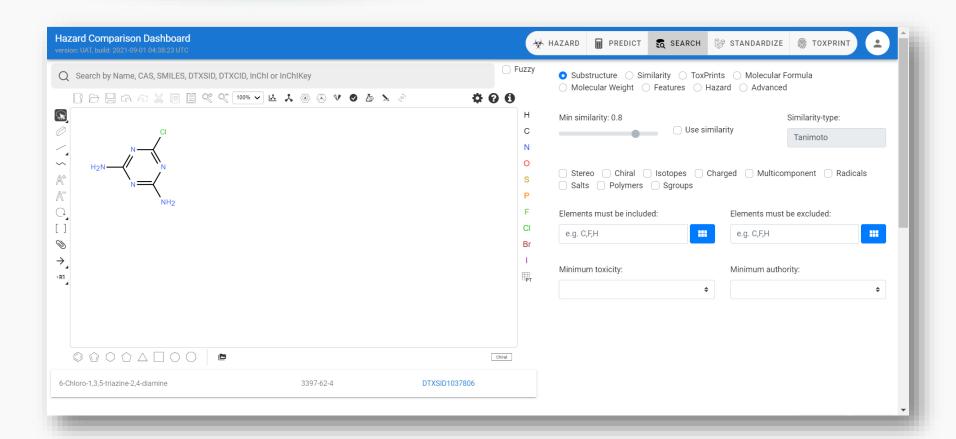
#### Hazard Module





#### Structure/Substructure/Similarity





#### Database of Public Domain MS Data



- Metadata ranking of candidates based on mass/formula searching was a starting point
- Searching experimental spectra against in silico predictions was next.
- Searching experimental spectra against experimental spectra is the next phase.
- Assembling experimental mass spec data from the internet, homogenize formats and database – includes curation

#### Database of MS Methods



- There are hundreds of methods distributed across the EPA website – that can be enabled by cheminformatics
  - Focus: Aggregate MS method documents and extract chemicals to make methods structure searchable
  - Vision: Search by structure/substructure/similarity to find existing method(s) as a starting point

#### Work in Progress





Presently working with scientists from NIST

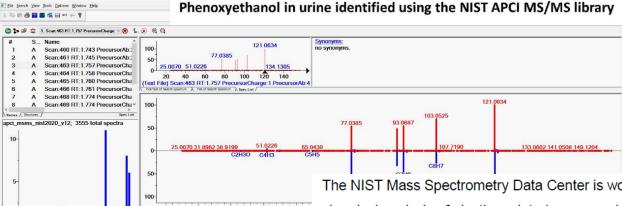
Unified mass spectrometry approach for identifying plastics- related contaminant: Chemical coverage, applications and web resources				
<i>Mar 23, 2022 7:00pm - Mar 23, 2022 9:00pm</i>				
Overview Comments				
Description A comprehensive list of plastic monomers, additives, and processing aids containing more than 10,000 compounds has been published recently (Environ. Sci. Technol. 2021, 55, 9339-9351). According to ISO standards, USP compliant testing, and FDA/EPA				

- Dataset to be released as a list
- Spectra will be added to our experimental spectral database for searching

# Collaboration with NIST > 10,000 plastic related chemicals



We have identified phenoxyethanol, several phthalates and phthalate metabolites, and bisphenol A in pooled human urine samples (NIST/SRMs), together with other contaminants in need of further analysis. A similar study was conducted for extracts from orthopedic casts using 16 polymer extracts.



DotProd

Name

2-Phen

2-Phen Styrene

Irganox

2-Phen

Benzoid

Benzoir

C8H10O2 (apci\_msms\_nist2020\_v12) 2-Phenoxyetha

Score

**■ 896** 

⊞ 654

628

**⊞** 565

m 243

The NIST Mass Spectrometry Data Center is working on a comprehensive approach to the chemical analysis of plastics-related compounds (PRC) using mass spectrometry as a contribution to the NIST circular economy program (<a href="https://www.nist.gov/circular-economy">https://www.nist.gov/circular-economy</a>).



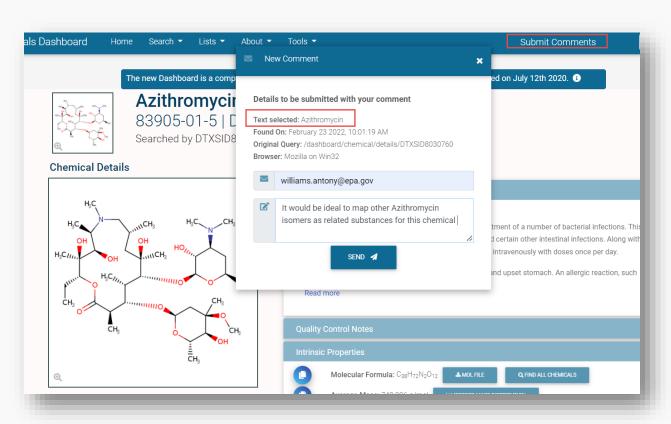


Due to the versatility of plastics, many manufactured products and environmental pollutants are associated with PRC. 10,547 compounds including monomers, additives, and processing aids of the plastics industry (Environ. Sci. Technol. 2021, 55,9339-9351) and three types of ionizations has been used in this work to obtain mass spectra and build standard libraries of PRC.

#### Request for participation



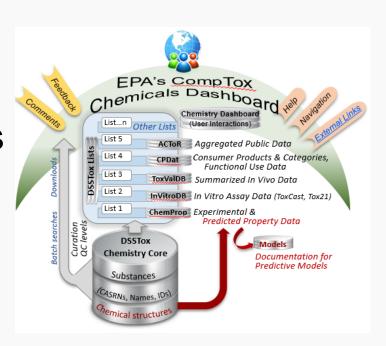
- Please point us to relevant datasets and articles
- Nothing is perfect so please flag issues



#### Conclusion



- Dashboard access to data for ~900,000 chemicals, next release will update to 1.2 million substances
- Extractable and leachables lists continue to expand
- "MS-Ready data" facilitates structure identification
- Data continues to grow with ongoing curation activities
- Proof-of-concept developments will release in future versions
- Do you want to learn more?



#### Contact

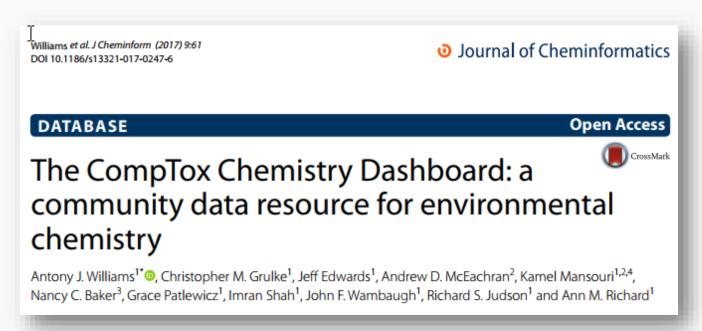


#### **Antony Williams**

CCTE, US EPA Office of Research and Development,

Williams.Antony@epa.gov

ORCID: https://orcid.org/0000-0002-2668-4821



https://doi.org/10.1186/s13321-017-0247-6