

Primer on New Version of CompTox Chemicals Dashboard

Antony John Williams

williams.antony@epa.gov

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

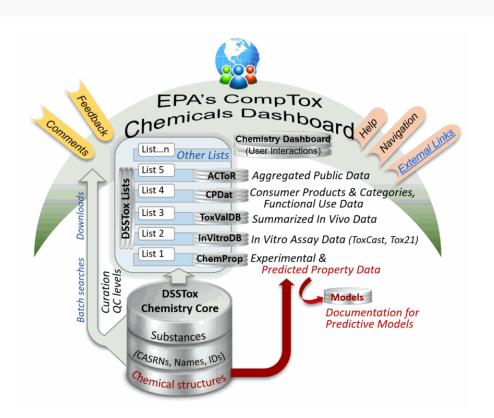
Communities of Practice, February 23rd 2022

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

The CompTox Chemicals Dashboard

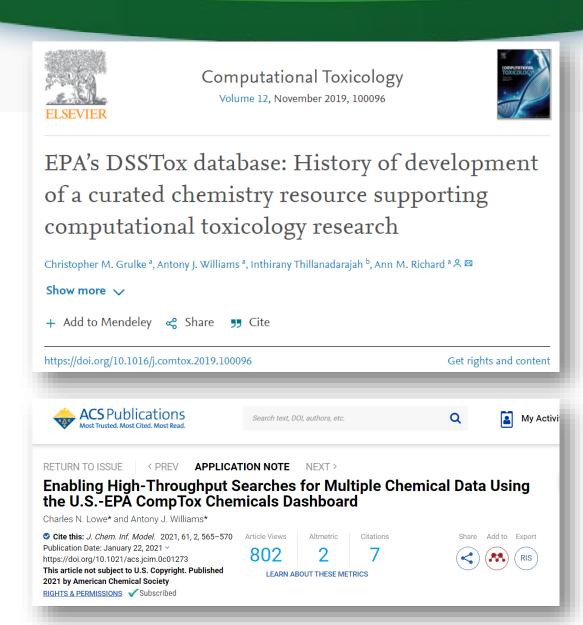


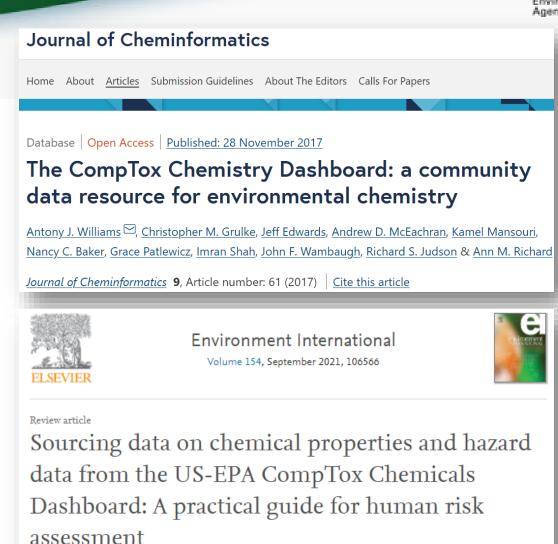
- Our primary tool for integrating and delivering our data to the community plus connect across agency resources
- Online since April 2016 with new releases
 - Spring and Fall every year (to coincide with conference seasons)
- Number of chemical substances has grown from 560k to 906k since first release and from ~40 to ~320 chemical lists
- Replaced multiple other Dashboards unifying data and making support easier



Some Related Publications of Interest







Antony J. Williams a R M, Jason C. Lambert a, Kris Thayer b, Jean-Lou C.M. Dorne C

Rearchitecting the Dashboard



- After 5 years and 10 releases of expanding data and functionality performance was degrading
- User experience and feedback encouraged redesign
- Multiple applications requiring underlying data needed unifying approach to facilitate development of other tools: e.g., GenRA, RapidTox

• The Dashboard was a successful "proof-of-concept" tool that required rearchitecting for ongoing support and performance

The Vision for the Rebuild



 VISION: Rearchitect the entire application using a data hub/data mart foundation, API to access and serve the data to the user interface

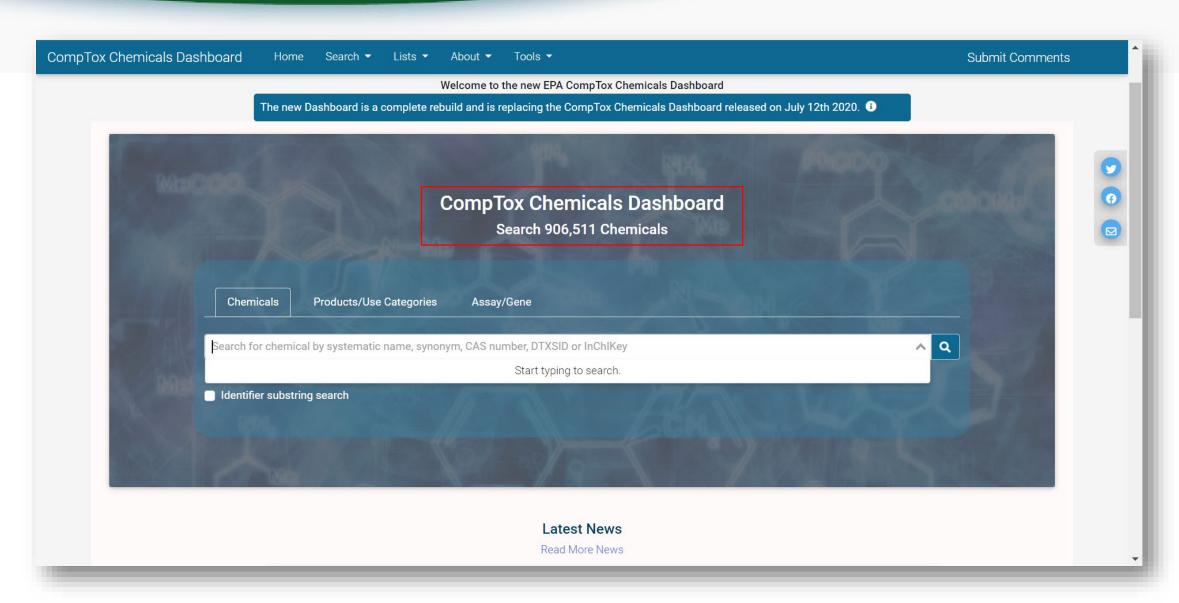
GUIDANCE: "1-to-1 mapping of new dashboard to old dashboard"

• EXPECTATIONS:

- Improved search performance
- Improved documentation and help manual
- Unify navigation especially in regards to tabular data handling
- Make future development easier with a new foundation

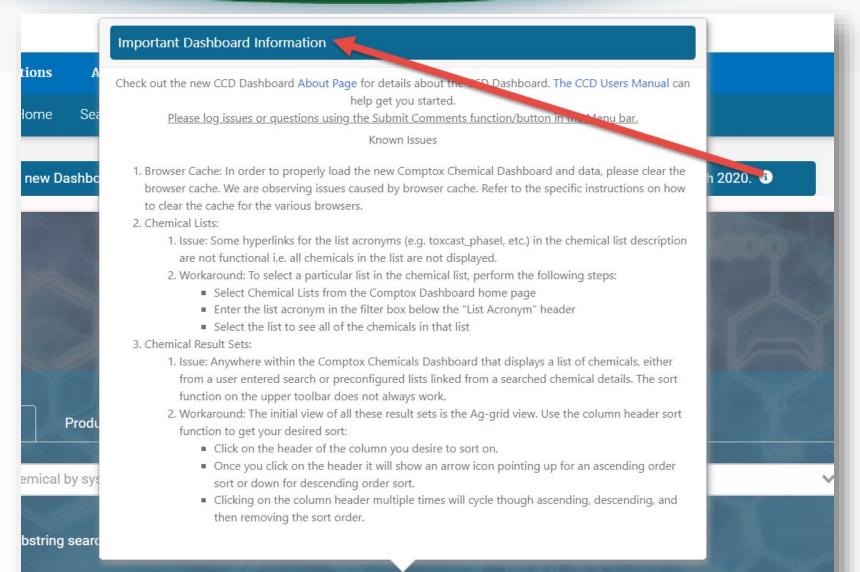
New Design - >906,000 chemicals





Updated information available





 On-click informational icon gives latest important information

Data updates across the board



- Data have been updated across the application
 - Chemistry data and lists
 - ToxVal
 - Exposure
 - Bioactivity
- Data versioning is under:

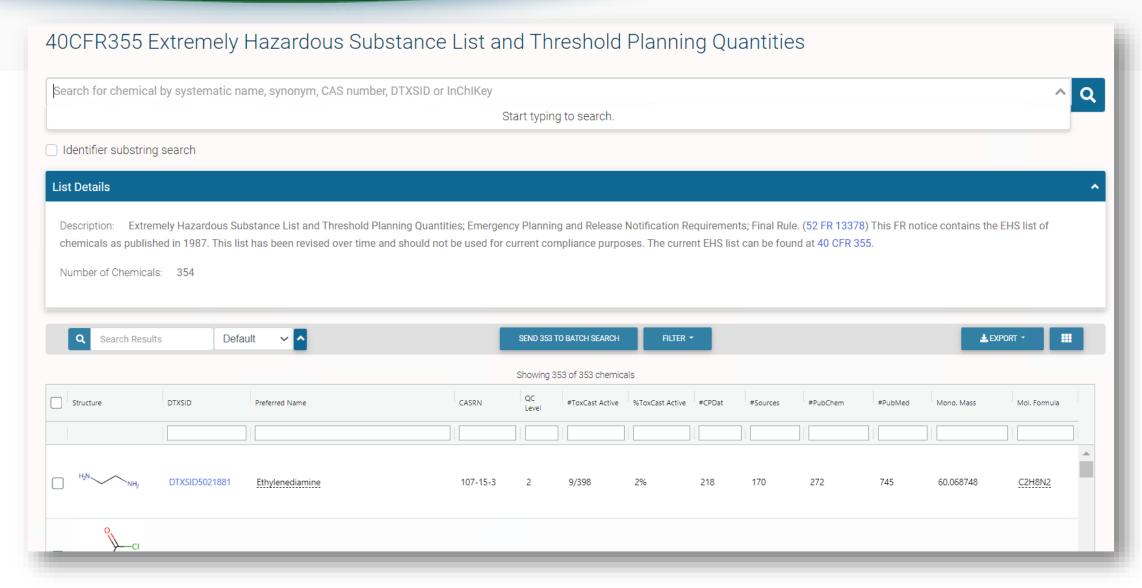
https://comptox.epa. gov/dashboard/about

Source Name	Version	Description	URL
ChemReg/ DSSTox	20210426_DSSTOX	Cheminformatics backbone of the CCTE's ToxCast and the multi-agency Tox21 HTS screening programs	https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database
			https://doi.org/10.23645/epacomptox.5491516.v5
Invitro	V.3.4	Data generated by the ToxCast and Tox21 in vitro high-throughput	https://www.epa.gov/chemical-research/toxicity-forecasting
		screening (HTS) programs	https://doi.org/10.23645/epacomptox.6062623.v5
ToxVal	V.9.1.1	Collection of animal (in vivo) toxicity study data,	https://gaftp.epa.gov/Comptox/Staff/rjudson/datasets/ToxValDB/
Rapid, Exposure and Dose Data (Factotum, SHEDS-	Factotum 9/11/202 SEEM3 HTTK	Provides data for exposure estimates for thousands of chemicals.	https://www.epa.gov/chemical-research/rapid-chemical-exposure-and dose-research#10
HT, SEEM, HTTK)			https://cran.r-project.org/web/packages/httk/index.html
			https://github.com/HumanExposure/SEEM3RPackage
ChemProp/ QSAR	20191118_ChemPROP	Capture measured or predicted property data associated with a particular source substance or list of	https://www.epa.gov/chemical-research/chemical-safety-analytics

New Table Capabilities

http://ccte-ccd.epa.gov/dashboard/chemical-lists/40CFR355





Find all chemicals with N2 in formula



CASRN	QC Level	#ToxCast Active	%ToxCast Active	#CPDat	#Sources	#PubChem	#PubMed	Mono. Mass	Mol. Formula
									N2
107-15-3	2	9/398	2%	218	170	272	745	60.068748	C2H8N2
77-81-6	2	-	-	-	54	36	271	162.055815	C5H11N2O2P
26419-73-8	2	-	-	-	38	15	-	234,049670	C8H14N2O2

Find "aniline" in the names Substructure searching is better

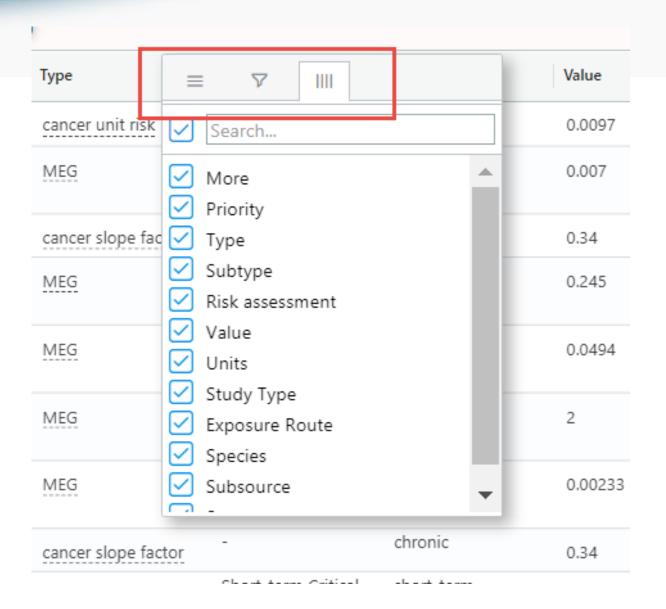


Structure	DTXSID	Preferred Name 🔽
		aniline
□ F NH₂	DTXSID9024512	3-(Trifluoromethyl)aniline
H ₂ N CH ₃	DTXSID5043847	2,4,6-Trimethylaniline
☐ H ₂ N—	DTXSID8020090	Aniline

Capabilities to switch columns on/off



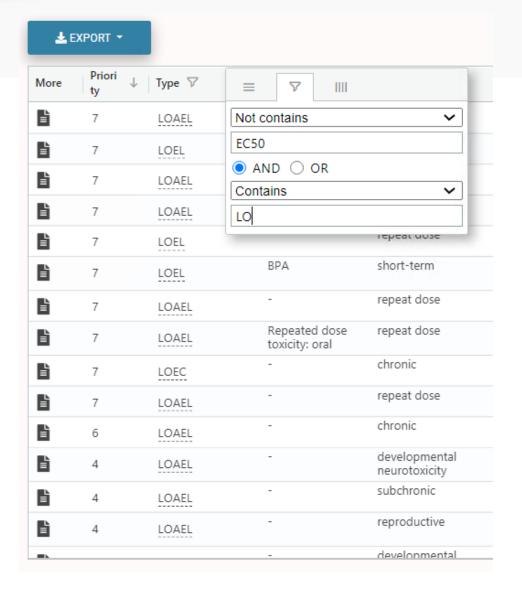
 Table functionality for selecting columns for view is built in



Capabilities to Filter the Table



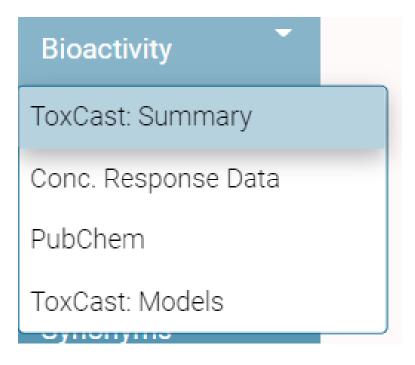
- Filtering on the tables is possible
- It is not available on every column on every table and in many cases probably doesn't make sense
- While using please highlight where it might be of value



Bioactivity Refresh



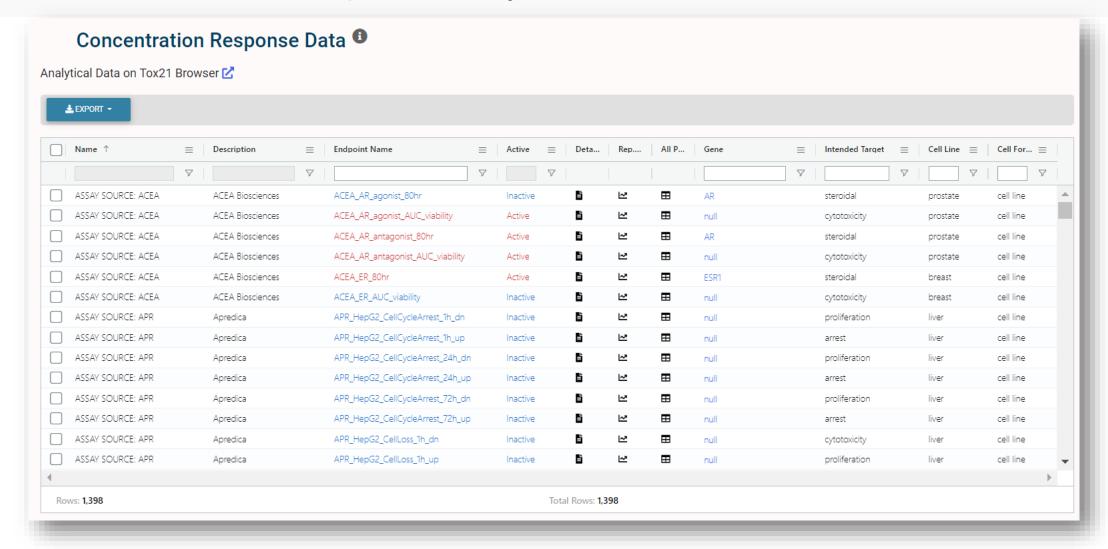
- We have collapsed EDSP21 and ToxCast/Tox21
- Concentration Response Data views have been rebuilt
- New display of ToxCast Summary view and Conc. Response curves



Bioactivity Refresh



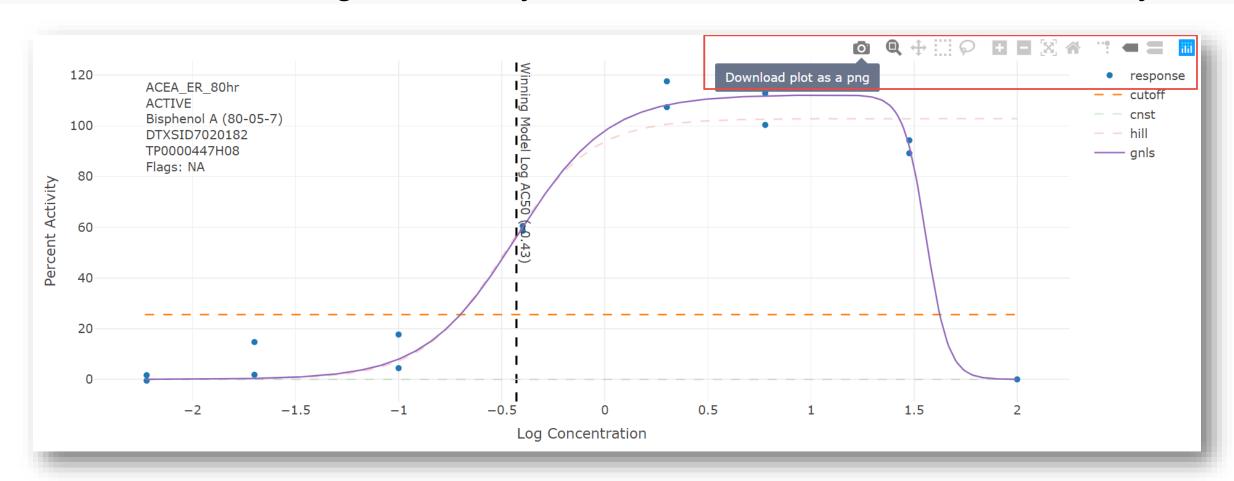
Concentration Response fully reworked



More flexible interface for data viewing Based on Plotly

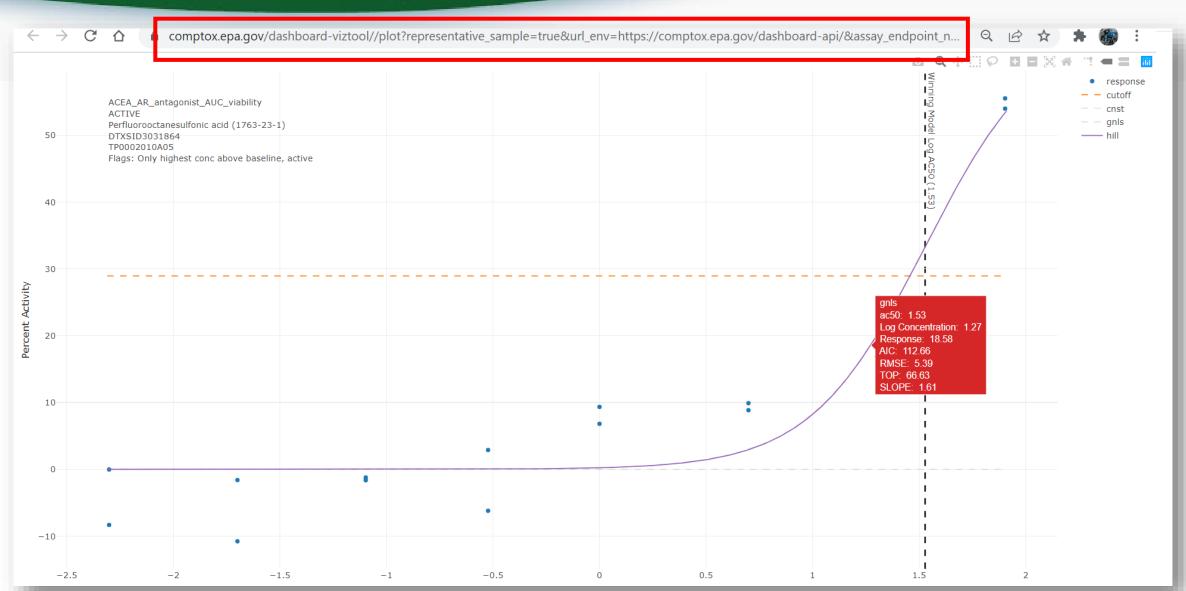


All the advantages of Plotly visualization now built into bioactivity



Direct URL access – useful for referencing in publications

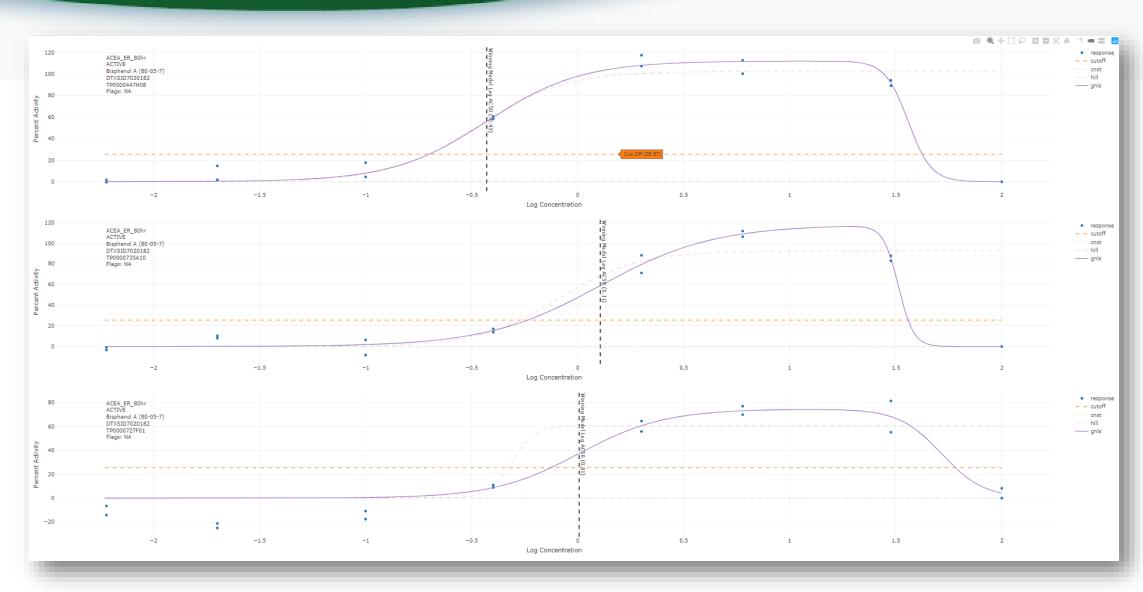




Multiple plots available via URL

https://comptox.epa.gov/dashboard-viztool//plot?representative_sample=false&url_env=https://comptox.epa.gov/dashboard-api/&assay_endpoint_nm=ACEA_ER_80hr&dsstox_id=DTXSID7020182

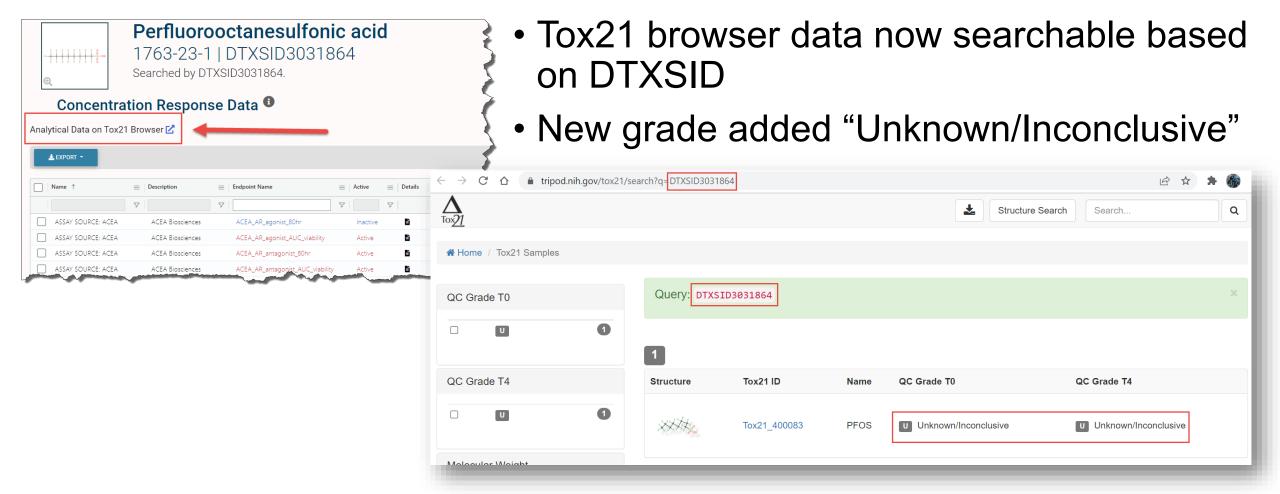




Analytical QC Data updated

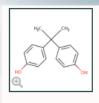


 Previous Analytical QC data was out of date so now linked to the latest data update on the Tox21 browser



External Links Updated and Expanded





Bisphenol A

80-05-7 | DTXSID7020182

Searched by DTXSID7020182.

General

- ACS Reagent Chemicals
- CAMEO Chemicals
- m ChEBI
- ChemAgora
- m ChEMBL
- T Chemspider
- Consumer Product Information Database
- © CPCat
- DrugBank
- **ECHA Brief Profile**
- **ECHA** Infocard
- EPA Substance Registry Service
- Q MSDS Lookup
- MIOSH Chemical Safety Cards
- NIST Chemistry Webbook
- mm PubChem
- PubChem 3D conformer download
- PubChem 3D Structure Display
- PubChem: Chemical Vendors
- PubChem Safety Sheet
- (A) State-Specific Water Quality Standards
- ToxPlanet
- WEBWISER
- Wikidata
- **W** Wikipedia
- 🜞 Wolfram Alpha

Toxicology

- ACTOR
- ACTOR PDF Report
- BindingDB
- CalEPA OEHHA
- Chemical Checker
- (ChemView
- CTD
- OH, DrugPortal
- eChemPortal
- ECOTOX
- National Air Toxics Assessment
- MIOSH IDLH Values

Publications

- >> Bielefeld Academic Search Engine
- BioCaddie DataMed
- CORE Literature Search
- Federal Register
- G Google Books (Structure Search)
- G Google Books (Text Search)
- G Google Patents (Structure search)
- G Google Patents (Text search)
- G Google Scholar (Structure search)
- G Google Scholar (Text search)
- (IRIS Assessments
- MIOSH Pocket Guide
- NIOSH Skin Notation Profiles
- PPRTVWEB
- PubMed
- Regulations.gov
- RSC Publications
- Springer Materials

Analytical

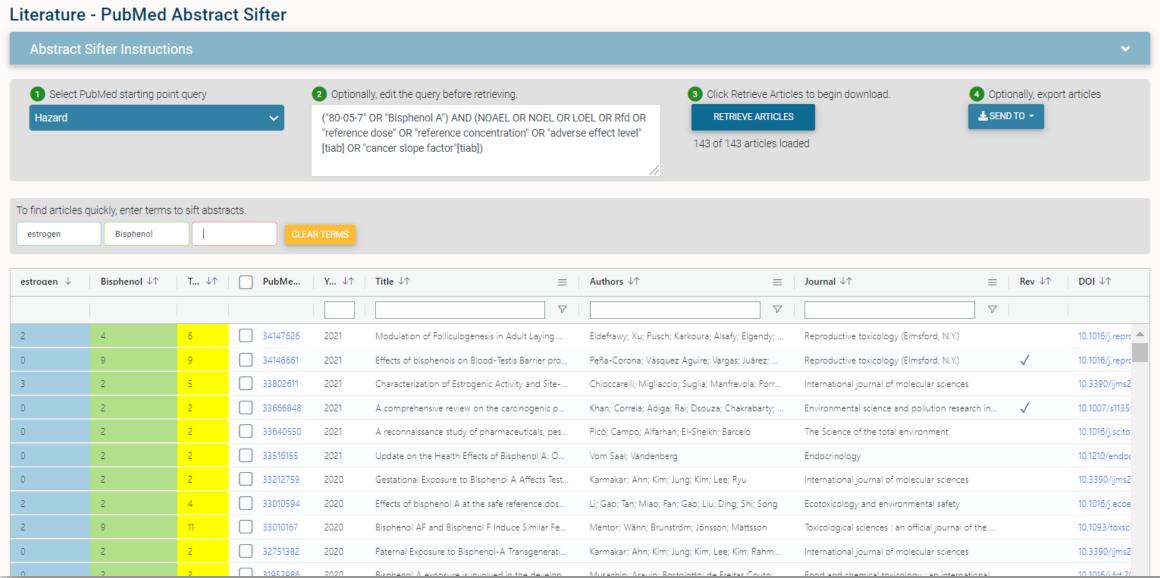
- IR Spectra on PubChem
- MassBank
- MONA: MassBank North America
- amzCloud
- National Environmental Methods Index
- NIST Antoine Constants
- NIST IR Spectrum
- NIST Kovats Index values
- NIST MS Spectrum
- Protein DataBank
- RSC Analytical Abstracts
- A Tox21 Analytical Data

Prediction

- 2D NMR HSQC/HMBC Prediction
- Carbon-13 NMR Prediction
- ChemRTP Predictor
- Proton NMR Prediction
- Link decay is an ongoing issue and URLs need updating
- New resources added by request

Abstract Sifter Refresh



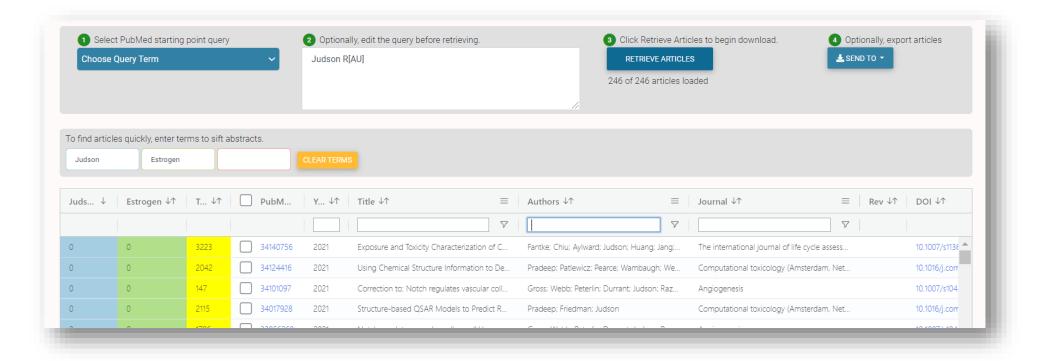


Abstract Sifter as a TOOL



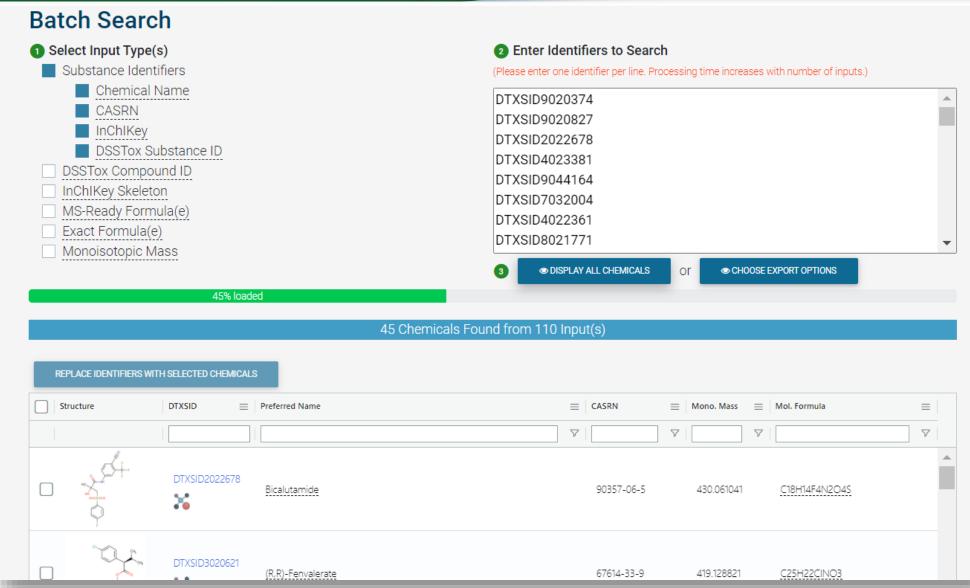
Literature - PubMed Abstract Sifter





Batch Search Refresh





Batch Search – List Filtering

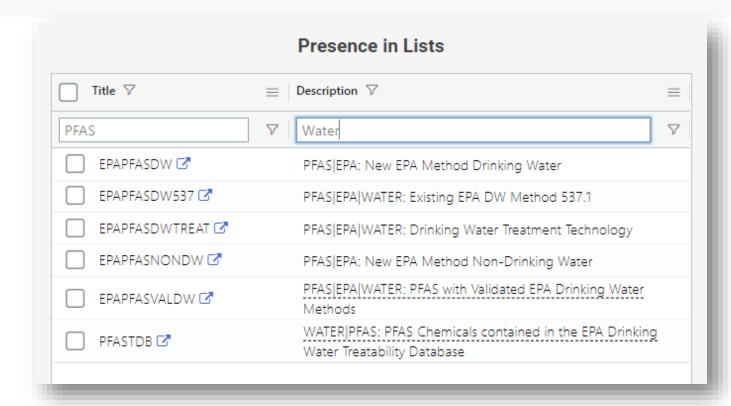


	Customize E	Export Results		
♣ CHOOSE EXPORT FORMAT ▼	Your file will be exported in Microsoft E	Excel Format (.xlsx)		
Select All columns available				
hemical Identifiers	Metadata		Presence in Lists	
DTXSID	Curation Level Details	Title	■ Description	=
Chemical Name	Safety Data		7	∇
DTXCID	NHANES/Predicted Exposure		40 CFR 116.4 Designation of Hazardous Substances	
CAS-RN	Data Sources	40CTK1104 C	(Above Ground Storage Tanks)	
InChiKey	Include ToxVal Data Availability	☐ 40CFR355 🗹	40CFR355 Extremely Hazardous Substance List and	
IUPAC Name	Assay Hit Count	☐ ACSREAG ☑	Threshold Planning Quantities LIST: ACS Reagent Chemicals	
	Number of PubMed Articles			
ructures	PubChem Data Sources	☐ AEGLVALUES ☑	AEGLS: Acute Exposure Guideline Levels	
Mol File	CPDat Product Occurrence Count	ALGALTOX 🗗	LIST: Algal Toxins	
SMILES	IRIS	☐ ALLSURFACTANTS ☑	CATEGORY: Surfactants	
InChI String	PPRTV	■ AMINOACIDS ☑	CATEGORY: Amino acids	
MS-Ready SMILES	Wikipedia Article	■ AMPHIBOLES	Amphibole minerals	
QSAR-Ready SMILES	QC Notes	■ ANTIBIOTICS 🗗	CATEGORY PHARMACEUTICALS: Antibiotics	
trinsic and Predicted Properties	Include links to ACToR reports	☐ ANTIMICROBIALS ☑	CATEGORY WIKILIST ANTIMICROBIALS: Antimicrobials from Wikipedia	IS
Molecular Formula	Enhanced Data Sheets	□ AOPSTRESSORS ☑	List of Adverse Outcome Pathway Stressors	
Average Mass	MetFrag Input File (Beta)	☐ APCRARETRO ☑	LIST: APCRA Chemicals for Retrospective Analysis	
Monoisotopic Mass	ToxPrint single fingerprints	☐ ARCHEMICALS ☑	ANDROGEN: Androgen Receptor Chemicals	
TEST Model Predictions	Abstract Sifter Input File	4	/ NO CENT / No Ogen Neceptor Enermons	•
OPERA Model Predictions	Synonyms and Identifiers	, and		
	Related Substance relationships	Rows: 319		
	ToxPrint fingerprints (ChemoTyper)			
	Associated ToxCast Assays			

List Filtering



- List filtering is very beneficial
- Now ~320 lists so new functionality quickly filters
- Select the lists to push flags into the export file



Batch Search Output File... includes new "Cover Sheet"



2	WARNING	DO NOT COPY / PASTE THIS DATA Some search terms returned multiple values, copy/paste will result in misaligned data
3	Search datestamp	2022-02-23 12:18:17
4	Search term count	21
5	Found count	15
6	Not found count	6
7	Duplicate count	0
8		

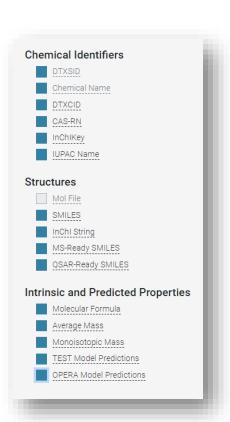
DTXSID	PREFERRED_NAME	EPAPFASDW	EPAPFASDW537	EPAPFASDWTREAT	EPAPFASNONDW	EPAPFASVALDW	PFASTDB
DTXSID4059916	Perfluorobutanoic acid	Υ	-	Υ	Υ	Y	Υ
DTXSID70191136	Perfluoro-3-methoxypropanoic acid	Υ	-	-	-	Y	Υ
DTXSID6067331	6:2 Fluorotelomer sulfonic acid	Υ	-	-	Υ	Y	Υ
DTXSID6062599	Perfluoropentanoic acid	Υ	-	-	Υ	Υ	Υ
DTXSID3031862	Perfluorohexanoic acid	Υ	Y	Υ	Υ	Y	Υ
DTXSID1037303	Perfluoroheptanoic acid	Υ	Υ	-	Υ	Y	Υ
DTXSID00192353	8:2 Fluorotelomer sulfonic acid	Υ	-	-	Υ	Y	Υ
DTXSID60500450	Perfluoro(4-methoxybutanoic acid)	Υ	-	-	_	Y	Υ
DTXSID8031865	Perfluorooctanoic acid	Υ	Υ	Υ	Υ	Y	-
DTXSID30382063	Perfluoro-3,6-dioxaheptanoic acid	Υ	-	-	_	Y	-
DTXSID8031863	Perfluorononanoic acid	Υ	Υ	Υ	Υ	Y	Υ
DTXSID8062600	Perfluoropentanesulfonic acid	Υ	-	-	Υ	Y	-
DTXSID5030030	Perfluorobutanesulfonic acid	Υ	Υ	Υ	Υ	Y	Υ
DTXSID5062760	2-(N-Ethylperfluorooctanesulfonamido)acetic acid	_	Υ	-	Υ	Y	Υ
DTXSID3031860	Perfluorodecanoic acid	Υ	Υ	Υ	Υ	Υ	Y

Performance Improvement



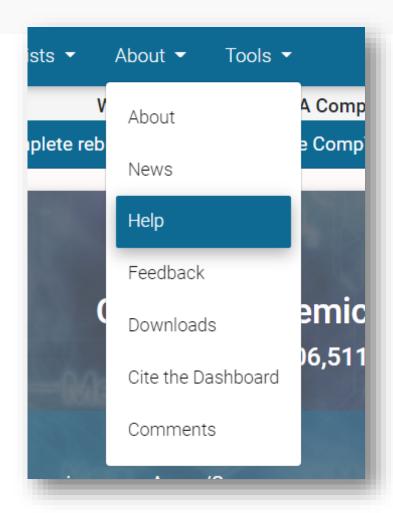
- The new architecture brings significant performance enhancements especially for list loading and batch searching
- List Loading: 5000 chemicals: 7secs vs 21secs

- Batch Search: 5000 chemicals: 2.5 secs vs 81 secs!
- Search limitation lifted from 5000 to 10,000 inputs



New Online Help





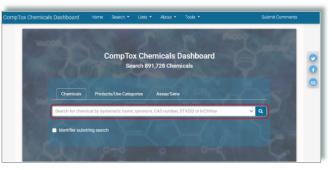


CCD Help Pages Searching CCD

For information about the CCD Batch Search tool, see <u>Batch Search</u>.

Simple search

Simple (single chemical) searches can be done from either the <u>Dashboard homepage</u>, or from the search box present at the top right of most Dashboard pages. For chemicals, simple searches are based on preferred name, synonym, CASRN, DSSTOX substance identifier (DTXSID), InChIKey, or IUPAC name. From the homepage, there are also tabs above the search box to search by <u>product use category</u> or <u>assay / gene</u>. Product use and assay / gene can only be searched from the homepage.



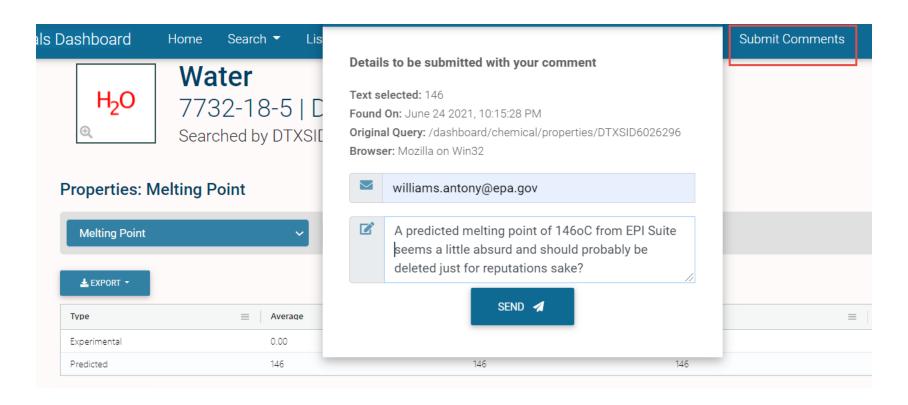
 $CompTox\ Chemicals\ Dashboard\ home\ page, \ \underline{https://ccte-ccd.epa.gov}, with\ search\ box\ marked\ in\ red.\ Note\ tabs\ for\ product\ use\ category\ and\ assay\ /\ gene\ above\ the\ search\ box.$

SEPA STATE TO A STATE OF THE ST

How to report comments and bugs



- Please note that Submit Comments is how we want feedback for CHEMICAL LEVEL detail. Select relevant text and submit.
- All comments will go directly into Jira for tracking and metrics

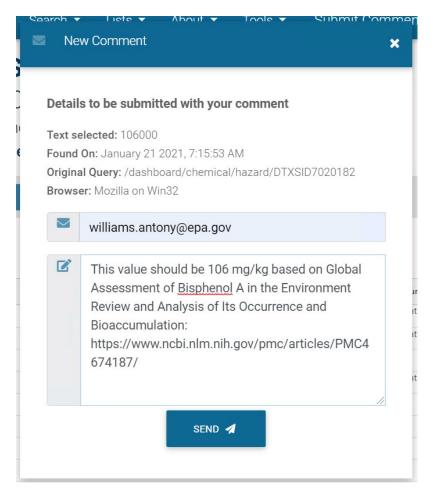


Highlight text and describe issue



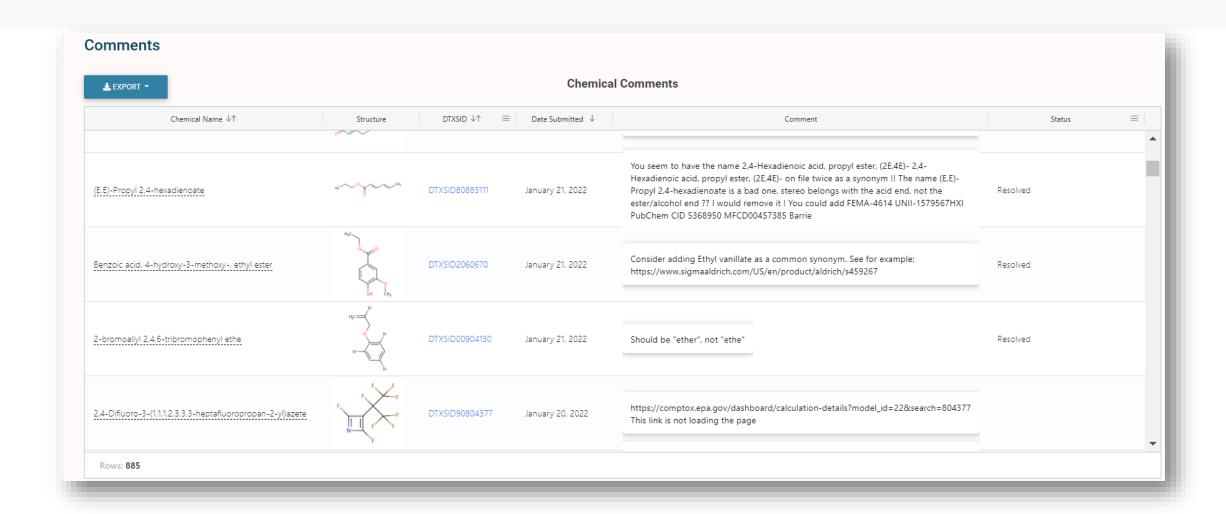
 Highlight any part of the screen and submit comments will capture the context, URL, date and time etc.

ubtype	Risk assessment	Value	Units Study Type
Short-term Criti	short-term	500	mg/m3
Short-term Negl	short-term	15	mg/m3
Long-Term, 5L/	chronic	7	mg/L -
Short-term Mar	short-term	100	mg/m3
Soil Negligible S	chronic	106000	mg/kg
-	chronic	0.05	mg/k
_	chronic		_



Comments are all viewable https://comptox.epa.gov/dashboard/comments





Work in Progress



Data updates in preparation – >1.2M chemicals release next

CAS Common Chemistry™ expands collection of publicly available chemical information

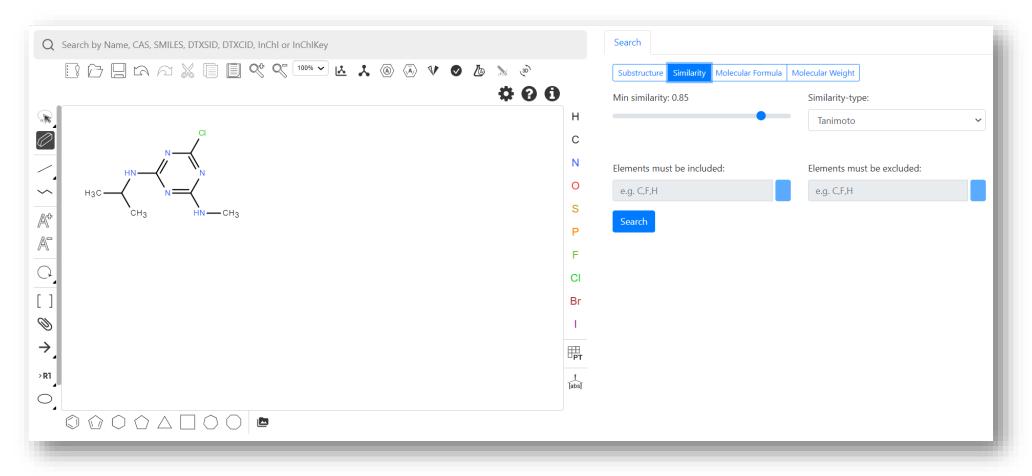
COLUMBUS, Ohio, March 17, 2021 — CAS, a division of the American Chemical Society (ACS) that specializes in scientific information solutions, has expanded the <u>CAS Common Chemistry resource</u>. To strengthen the accuracy of publicly available scientific information, CAS Common Chemistry now provides authoritative information on nearly 500,000 substances from CAS REGISTRY®. The collection represents substances commonly found in consumer products, on regulatory lists and as part of introductory chemistry curricula.

- Incorporation of some of the Common ChemistryTM data, specifically without structures
- Addition of structures is ongoing methodical curation
- Increased efforts mapping parents to metabolites and degradants and polymers to monomers
- 10s of new lists added and existing lists updated where possible

Work in Progress



• Structure/substructure/similarity search module in development



Access the CCD-API

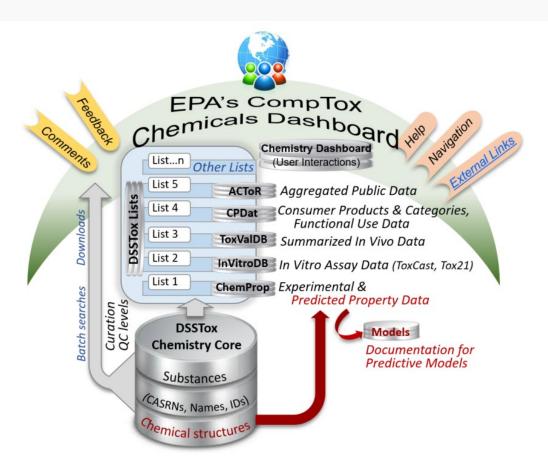


- We WANT you to use the underlying data
- API (application programming interface) servicing the Dashboard will go public this year

 Previous users of ACToR web services are encouraged to use the CompTox Chemicals Dashboard API (CCD-API) when available

Summary and Conclusion





- A much-needed rearchitecture has been successfully delivered
- This sets us up for more frequent data releases and is a foundation for new functionality

 Thank you for all of your feedback and support to this point

Acknowledgments



Feedback and follow-up is welcomed! Your questions help

Use the comments system to provide feedback

 The dashboard is based on the efforts of so many, those still here and those that have gone



 Thanks to all data providers, developers infrastructure staff, management staff for support and guidance

Transition of Product Owners



- AJW has been Dashboard product owner for 6 years & 11 releases
- A whiteboard idea went proof-of-concept to production application

- The NEW product owner is Dr Nisha Sipes.
 Assistant Center Director for Research
 Translations & Program/Regulatory Support
 for the US EPA Center for Computational
 Toxicology and Exposure (CCTE)
- Sipes.Nisha@epa.gov



New Tools in Development





- **Literature Mining** Tool (Abstract Sifter)
- **Downloadable Data**
- GenRA

- The center continues to work on new tools available via the portal https://comptox.epa.gov/
- The latest tool added is the new GenRA release...

