

NORMAN Databases Workshop: UBA, Berlin, 8-9 June 2017



NORMAN Suspects meet the US EPA CompTox Chemistry Dashboard

Emma Schymanski, Eawag: Swiss Federal Institute of Aquatic Science and Technology Antony J. Williams & Andrew McEachran, United States Environmental Protection Agency





Chemistry Dashboard

Search a chemical by systematic name, synonym, CAS number, or InChlKey



Q





What is the CompTox Chemistry Dashboard?

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

Aims of US EPA's Computational Toxicity Program:

- Integrate advances in biology, chemistry, and computer science to...
- Prioritize chemicals based on potential human and ecological health risks

The CompTox Chemistry Dashboard:

- Provides access to ~747,000 chemical substances & associated data
- Years of investment/experience in data validation and curation
- Data include: (plus a LOT more ...)
 - Experimental and predicted physicochemical properties
 - ToxCast bioassay screening data
 - Product and functional use information and more
- O Search capabilities include:
 - Mass or formula-based searching
 - Rank-ordering of results via functional use statistics to prioritize chemicals





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



Chemistry Dashboard

Search a chemical by systematic name, synonym, CAS number, or InChlKey

Q

Single component search Ignore isotopes

See what people are saying, read the dashboard comments!

Need more? Use advanced search.

747 Thousand Chemicals



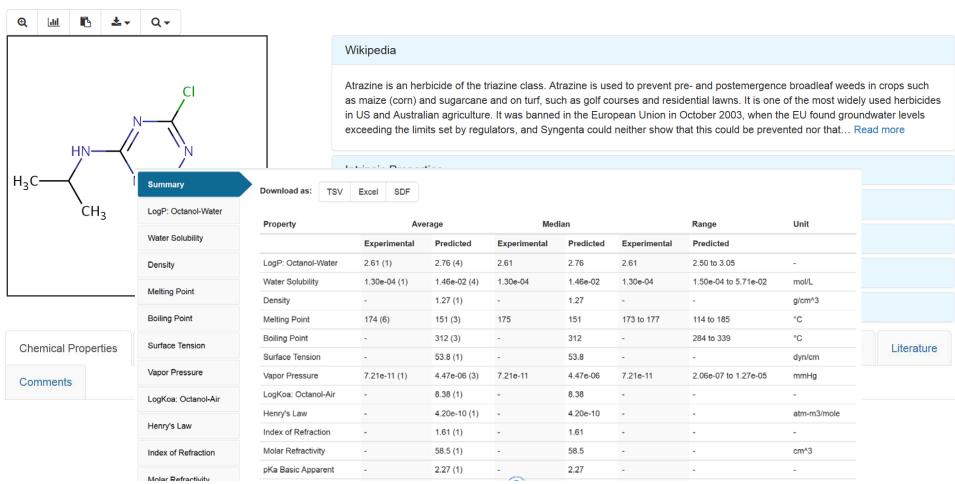


https://comptox.epa.gov/dashboard/

Atrazine

1912-249 | DTXSID9020112

Searched by Approved Name: Found 1 result for 'atrazine'.



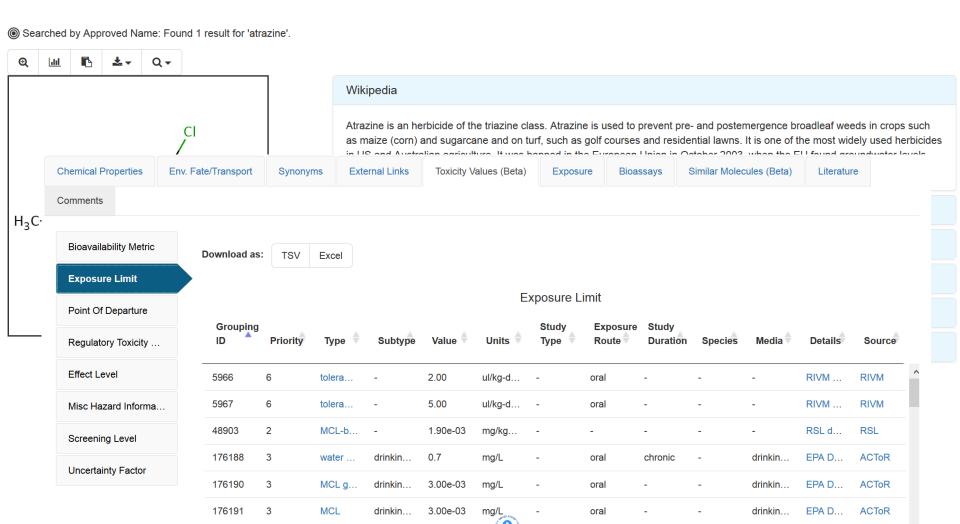




https://comptox.epa.gov/dashboard/

Atrazine

1912-249 DTXSID9020112



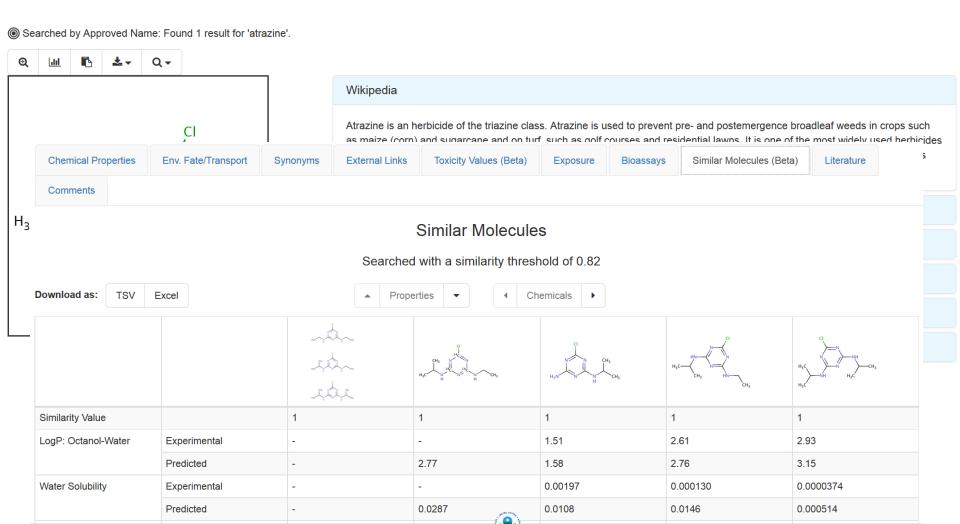




https://comptox.epa.gov/dashboard/

Atrazine

1912-249 DTXSID9020112







Why should the Dashboard meet NORMAN?

Common Goals

- Much of the CompTox data is open as is NORMAN data https://comptox.epa.gov/dashboard/downloads
- Increase access to data for use in other applications

Mutual Benefits

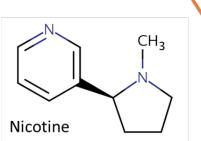
- We have access to data they don't have, within a large European network
- We have additional predictive values & initiatives that are of interest
- They have access to data that we don't have
- They have years of investment/experience in data validation and curation
- They have the ability to provide services currently way beyond our means





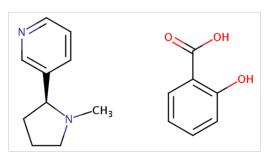
Schymanski & Williams, 2017, ES&T DOI: 10.1021/acs.est.7b01908

Collaboration on Chemical Curation



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

Tox: yes | Expo: yes | Bioassay: yes

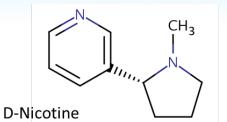


Benzoic acid, 2-hydroxy-, compd. with 3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)

OC(=O)C1=C(O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1 DTXSID5075319 | AIBWPBUAKCMKNS

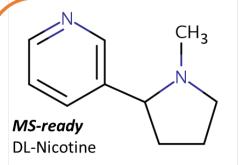
29790-52-1 | 300.1474 | 0.929 | 6

Tox: no | Expo: yes | Bioassay: no



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

Tox: **no** | Expo: **yes** | Bioassay: **yes**



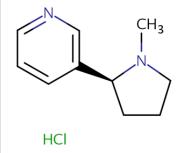
CN1CCCC1C1=CN=CC=C1 DTXSID3048154 | SNICXCGAKADSCV

22083-74-5 | **162.1157** | 0.953 | 9

Tox: yes | Expo: no | Bioassay: yes

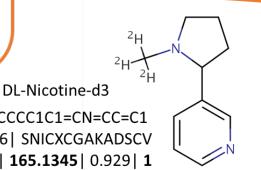
DTXSID | InChIKey 1st Block CAS | Monoiso. Mass | logP | Sources Data on: Toxicity | Exposure | Bioassays CH_3

LEGEND: Name, SMILES



Nicotine hydrochloride

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



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

Tox: no | Expo: no | Bioassay: no





Collaboration on Chemical Curation of Lists

1	,		
Pharmaceutical List with Consumption Data	SwissPharma_Table S2.csv	SwissPharma_TableS2_InChlKeys.txt	Singer et al. 2016. DOI: 10.1021/acs.est.5b03332
Swiss Insecticides, Fungicides and TPs	SwissPesticides_Table S1.csv	SwissPesticides_TableS1_InChlKeys.txt	Moschet et al. 2013.
NormaNEWS for retrospective screening of new emerging contaminants	NormaNEWS_V4_26042017.csv	NormaNEWS_V4_InChIKeys.txt	NormaNEWS list provided by Nikiforos Alygizakis, Saer Samanipour and Kevin Thomas
Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006)	Merged_CosmeticProducts_04052017.csv	Merged_CosmeticProducts_04052017_InChlKeys.txt	The scientific committee on cosmetic products and non-food products Intended for consumers - SCCNFP/0389/00 Final and Commission Decision 2006/257/EC amending the Decision 96/335/EC. Provided by Peter von der Ohe, UBA, curated by Reza Aalizadeh, University of Athens
PFAS Highly fluorinated substances list: KEMI	PFAS_Market_Keml_EPA_1Feb2017.xlsx ~2,600 PFAS	Curation in progress: coming soon	Appendix 2 from Swedish Chemicals Agency KEMI Report 7/15. Provided by Stellan Fischer, KEMI
NORMAN Priority List 2015	NORMAN_PriorityList_2016.csv Further curation in progress	NORMAN_PriorityList_2016_InChIKeys.txt	Priority substances from NORMAN WG-1 (Prioritisation), provided by Valeria Dulio
French Monitoring	French_List_08052017.csv Further curation in progress	FrenchList_UniqueInChlKeys_08052017.txt	Provided by Valeria Dulio, curated by Reza Aalizadeh, University of Athens
KEMI Market List 24,	KEMI_MarketList_12052017_MSready.xlsx 883 Substances (Ex	KEMI_MarketList_12052017_MSready_InChlKeys.txt OO, Hazard Scores)	Provided by Stellan Fischer, KEMI including Hazard and Exposure scores, documented here. Curated by Reza Aalizadeh, University of Athens.
TSCA Surfactants	Coming soon		Provided by Lee Ferguson, sourced from James Little





KEMI PFAS List (Stellan Fischer)

PFAS Highly fluorinated substances list: KEMI PFAS_Market_KemI_EPA_1Feb2017.xlsx

~2,600 PFAS

Curation in progress: coming soon

Appendix 2 from Swedish Chemicals Agency KEMI Report 7/15. Provided by Stellan Fischer, KEMI

	_		. \						
CASno	0	CASnr	ECno	DTXSID	PREFERRED NAME	CASRN	SMILES		
422-6	53-9	422639	207-020-0	DTXSID9059969	1,1-Propanediol, 2,2,3,3,3-pentafluoro	422-63-9	OC(O)C(F)(F)C(F)(F)F	
375-8	88-2	375882	206-799-4	DTXSID9059919	Heptane, 1-bromo-1,1,2,2,3,3,4,4,5,5,6	375-88-2	FC(F)(F)C(F)	(F)C(F)(F)C(F)(F)
375-6	52-2	375622	206-790-5	DTXSID9059917	Pentanoyl fluoride, nonafluoro-	375-62-2	FC(=O)C(F)(F)C(F)(F)C(F)	(F)C
375-1	.6-6	375166	206-785-8	DTXSID9959915	Butanoyl chloride, heptafluoro-	375-16-6	FC(F)(F)C(F)	(F)C(F)(F)C(C	CI)=C
375-0	8-00	375008	206-781-6	DTXSID9059913	Butanenitrile, heptafluoro-	375-00-8	FC(F)(F)C(F)	(F)C(F)(F)C#I	N
356-8	36-5	356865	206-608-4	DTXSID9059884	2,2,3 3,3-Pentafluoropropyl acrylate	356-86-5	FC(F)(F)C(F)	(F)COC(=O)C	:=C
356-2	27-4	356274	206-602-1	DTXSID9059882	Ethyl heptafluorobutyrate	356-27-4	CCOC(=O)C(F)(F)C(F)(F)C	C(F)(
338-8	33-0	338830	206-420-2	DTXSID9059834	1-Propal amine, 1,1,2,2,3,3,3-heptaflu	338-83-0	FC(F)(F)C(F)	(F)C(F)(F)N(0	C(F)(
335-9	9-9	335999	206-406-6	DTXSID9059832	1-Heptanol, 2,2,3,3,4,4,5,5,6,6,7,7-doc	335-99-9	OCC(F)(F)C(F)(F)C(F)(F)C	(F)(F
33 35 89				Norman Net	twork PFAS (KEMI Report	t)			F)((F) (F)
57 76 38			Search S	FISHFLUORO Cher	micals	Q			(F)
42 73	List Details								C(F
85 30	and-alternativ	ves.pdf target='_bla	ank'>Appendix 2	-	p://www.kemi.se/en/global/rapporter/2015/report-7-15-oc ency Report 7/15 on the occurrence and use of highl s provided by Stellan Fisher.				F)(C(F F)(





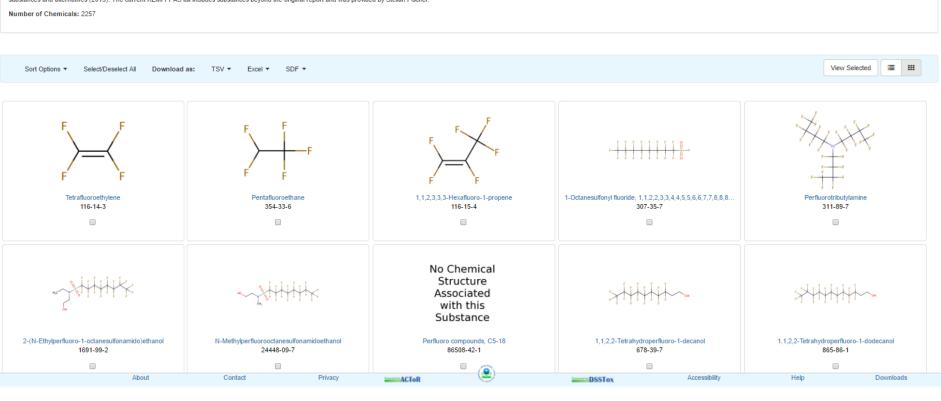
KEMI PFAS List (Stellan Fischer)

Norman Network PFAS (KEMI)

Search SFISHFLUORO Chemicals Q

List Details

Description: This list of perfluorinated substances originated from Appendix 2 from Swedish Chemicals Agency Report 7/15 (available at http://www.kemi.se/en/global/rapporter/2015/report-7-15-occurrence-and-use-of-highly-fluorinated-substances-and-alternatives.pdf) on the occurrence and use of highly fluorinated substances and alternatives (2015). The current KEMI PFAS list includes substances beyond the original report and was provided by Stellan Fischer.



NormaNEWS





Q

F)(F F)(F F)(f) (F) =C(=C(

C=C

NormaNEWS for retrospective screening of new emerging	NormaNEWS_V4_26042017.csv	,	NormaNEWS list provided by Nikiforos Alygizakis, Saer Samanipour and Kevin Thomas

				_					
INPUT	DTXSID	PREFERRED	NAME	CASRN	IUPAC NANSMILES				
DTXSID40881097	DTXSID40881097	C11-LAS		NOCAS_881097	- CCCCCC	(CCCC)C1=	=CC=C(C=C	C1)S(O)(=O))=O
DTXSID30860093	DTXSID30860093	4-(Dodecan-6-	-yl)benze	23003-92-1	4-(Dodecan CCCCCC	(CCCCC)C	1=CC=C(C=	=C1)S(O)(=0	O)=O
DTXSID80881096	DTXSID80881096	C13-LAS		NOCAS_881096	CCCCCC	CCC(CCC)	C1=CC=C(C=C1)S(O)((=O)=O
DTXSID20881095	DTXSID20881095	C14-LAS		NOCAS_881095	- CCCCCC	CCCC(CCC	C)C1=CC=C	(C=C1)S(O)(=O)=O
DTXSID60881094	DTXSID60881094	SPA-8C		NOCAS_881094	- CCCC(CCC	CC(O)=O)C	1=CC=C(C=	C1)S(O)(=0	O=(C
DTXSID50865484	DTXSID50865484	10-hydroxycar	bazepine	29331-92-8	10-Hydroxy-NC(=O)N1C	2=CC=CC	=C2CC(O)C	2=CC=CC=	=C12
DTXSID00881093	DTXSID00881093	Desacetyl dilti	azem	42399-40-6	- [H][C@]1(S	C2=C(C=C	C=C2)N(CC	N(C)C)C(=0	O)[C@@H
D-									

NormaNEWS: Norman Early Warning System

Search NORMANEWS Chemicals

List Details

D,

D. D.

D. D.

D, D,

D,

D.

D,

D.

D.

D.

Description: The Norman Early Warning System (NormaNEWS) is a pilot network designed to investigate the spatial and temporal distribution of newly identified contaminants of emerging concern in environmental samples through performing retrospective suspect screening on HRMS data acquired using different instrumental platforms and data processing software. The NormaNEWS pilot study was performed through recruiting eight reference laboratories with available archived HRMS data with the goal of exploring the potential of an early warning network to rapidly establish the occurrence of newly-identified contaminants of emerging concern across Europe and beyond, through the use of retrospective suspect screening employing HRMS. The pilot study was referred to as the Norman Early Warning System, abbreviated to NormaNEWS.

Number of Chemicals: 131



NormaNEWS





y Dashboard I NORMANEWS

NORMANews

Search NORMANEWS Chemicals

Q

List Details

Description: The Norman Early Warning System (NormaNEWS: http://www.norman-network.com/?q=node/244) is a pilot network designed to investigate the spatial and temporal distribution of newly identified contaminants of emerging concern in environmental samples through performing retrospective suspect screening on HRMS data acquired using different instrumental platforms and data processing software. The NormaNEWS pilot study was performed through recruiting eight reference laboratories with available archived HRMS data with the goal of exploring the potential of an early warning network to rapidly establish the occurrence of newly-identified contaminants of emerging concern across Europe and beyond, through the use of retrospective suspect screening employing HRMS. The pilot study was referred to as the Norman Early Warning System, abbreviated to NormanNEWS.

Number of Chemicals: 131

Select/Deselect All

Download as:

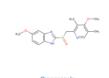
TSV ▼

1.2.3-Benzotriazole 95-14-7

Bisphenol A 80-05-7

View Selected

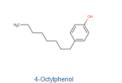
N-Methyl-2-pyrrolidone 872-50-4



Gabapentin

60142-96-3

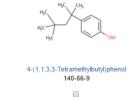
Omeprazole 73590-58-6

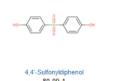


Atrazine

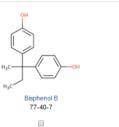
1912-24-9

1806-26-4





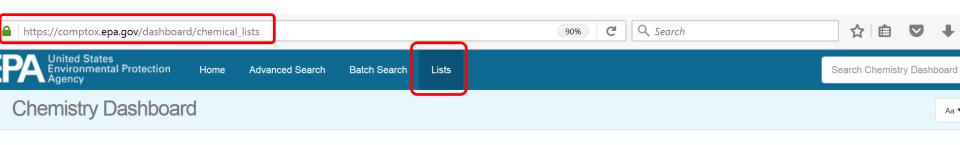
80-09-1







An overview of all the lists ...



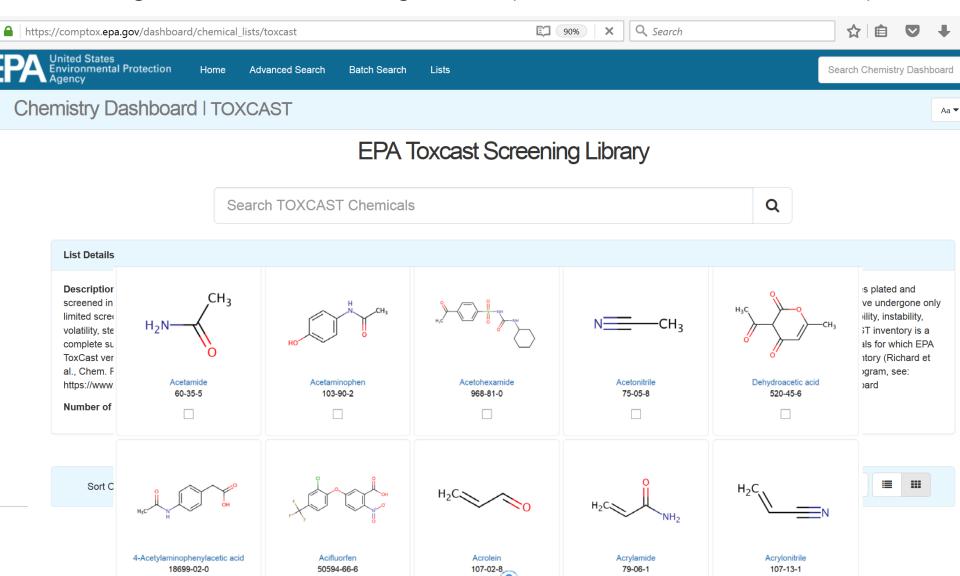
Select List

List Name	Number of Chemicals	List Description
CHEMINV: EPA Chemical Inventory for ToxCast (20170203)	5231	CHEMINV is full list of unique DSSTox substances mapped to historical chemical inventory of physical samples registered by EPA's ToxCast Chemical Contractor (Evotec) since launch of ToxCast program in 2007.
DNT Screening Library	1476	DNTSCREEN is a list of chemicals that is being used in mediumand high-throughput in vitro and zebrafish assays.
EPA Toxcast Screening Library	4736	TOXCAST includes all EPA-provided chemicals for which screening data have been generated in the ToxCast research program since 2007.
Norman Network PFAS (KEMI)	2257	Perfluorinated substances from a Swedish Chemicals Agency Report (provided by Stellan Fischer) on the occurrence and use of highly fluorinated substances.
NORMANews	131	The NORMAN Early Warning System (NormaNEWS) is a collaborative activity run by the NORMAN Network to investigate newly identified contaminants of emerging concern via retrospective screening on HRMS data.
Tox21 Screening Library More lists become availa	able with every release	TOX21SL is list of unique substances in Tox21 multi-federal agency screening library, contributed by the EPA, National Toxicology Program (NTP), and National Center for Advances in Translational Science (NCATS).





Obtaining all chemicals from a given list (downloadable as XLS or SDF)





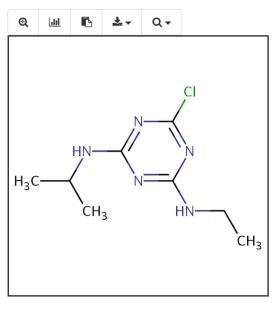


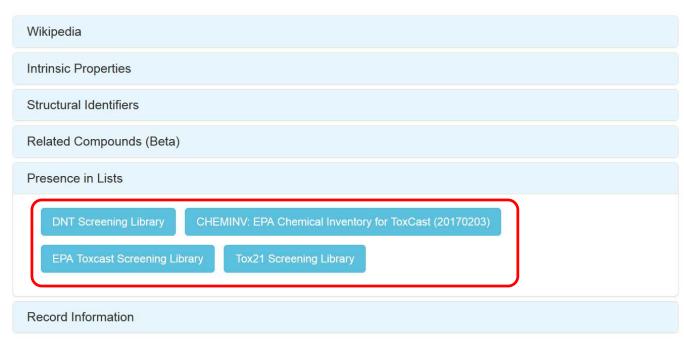
Individual Chemical Level

Atrazine

1912-249 DTXSID9020112

Searched by Approved Name: Found 1 result for 'atrazine'.







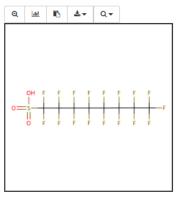


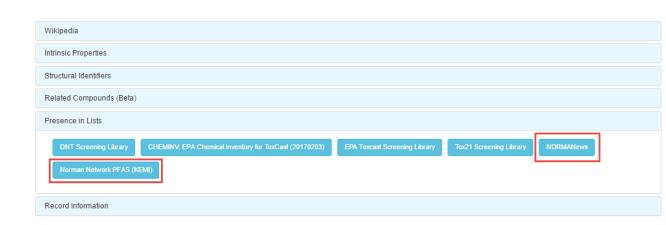
Individual Chemical Level ... example from NORMAN lists...

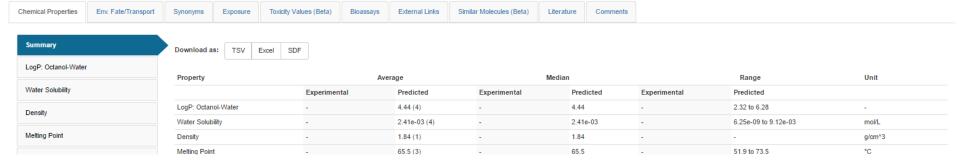
PFOS

1763-23-1 | DTXSID3031864

Searched by Approved Name: Found 1 result for 'PFOS'.







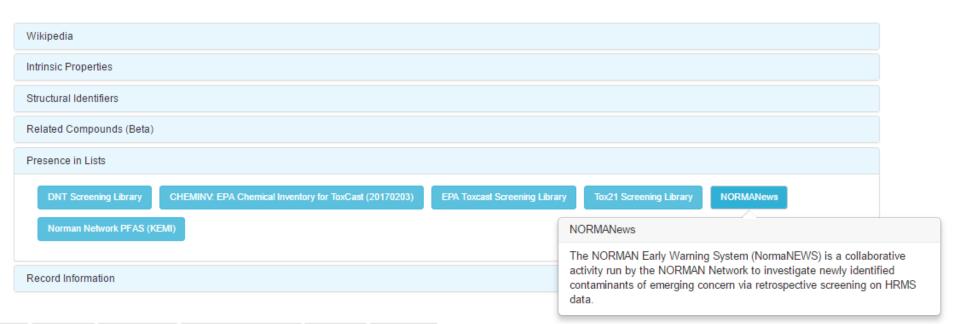




NORMAN Lists in the Dashboard ...

Coming soon ... (or just released ...)

Norman Network PFAS (KEMI Report) Perfluorinated substances from a Swedish Chemicals Agency Report (provided by Stellan Fisher) on the occurrence and use of highly fluorinated substances.







Future work ... integrating DTXSIDs into NORMAN Lists

Undefined mixtures (UVCBs)

Cleaning up lists to remove errors

Mol_ID	Name	EDITED NAMES FOR INPUT INTO SEARCH	CAS_RN	Merged DTXSIDs	DTXSID Based on Name	Preferred Name
SA8750	By-Product	By-Product	NA	-	-	NO_MATCH
stpQQR1546	C10-DATS C10-Dialkyl tet	C10-DATS C10-Dialkyl tetralin sulfonate 8	NA	ب	- /	NO_MATCH
SA2074	C10-LAS	C10-LAS	NA	-		NO_MATCH
stpQQR1582	C10LAS C10-linear alkylbe	C10LAS C10-linear alkylbenzyl sulfonate 4	NA	-	-	NO_MATCH
SA14931	C10-phosphonic	C10-phosphonic	NΛ		-	NO_MATCH
StpBB815	C12-15 ALKYL BENZOATE	C12-15 ALKYL BENZOATE	68411-27-8		-	NO_MATCH
SA13282	C12-AE5S	C12-AE53	NA	<u>)</u>	-	NO_MATCH
stpQQR1548	C12-LAS C12-linear alkyl k	C12-LAS C12-linear alkyl benzene sulfonat	NA	-	-	NO_MATCH
stpQQR690	C14-SAS (TENTATIVE) teti	C14-SAS (TENTATIVE) tetradecane-7-sulfo	NA	-	-	NO_MATCH
stpQQR1557	C16EOx C16EO2 C16-alco	C16EOx C16EO2 C16-alcohol polyethoxyla	NA	-	-	NO_MATCH
stpQQR1556	C18EOx C18EO2 C18-alco	C18EOx C18EO2 C18-alcohol polyethoxyla	4439-32-1	-	-	NO_MATCH
SA14932	C4-phosphonic	C4-phosphonic	NA	}	-	NO_MATCH
SA14929	C6-phosphonic	C6-phosphonic	NA	J.	-	NO_MATCH
stpQQR1583	C7SPC C7-sulfophenyl car	C7SPC C7-sulfophenyl carboxylates 4-(de	NA	-	-	NO_MATCH
SA14930	C8-phosphonic	C8-phosphonic	NA	-	-	NO_MATCH
stpQQR1547	C8-SPC C8-Sulfophenyl ca	C8-SPC C8-Sulfophenyl carboxylic acid 4-(NA	-	-	NO_MATCH
stpQQR1576	CA5PE2C 7-{4-[2-(carboxy	CA5PE2C 7-{4-[2-(carboxymethoxy)ethoxy	NA	-	-	NO_MATCH
stpQQR1578	CA6PE2	CA6PE2	NA	-	-	NO_MATCH
stpQQR1577	CA6PE2C	CA6PE2C	NA	-	-	NO_MATCH
stpQQR1575	CASPE2C	CASPE2C	NA	-	-	NO_MATCH
SA9863	cacotheline	cacotheline	561-20-6	-	-	NO_MATCH
SAn15715	Caerulomycin A	Caerulomycin A	21802-37-9	-		NO_MATCH
SA5151	cafedrine	cafedrine	58166-83-9	}-		NO_MATCH

(many) more registrations...





Handling of Undefined Mixtures

C10-12 chloroalkanes

Q

https://comptox.epa.gov/dashboard/dsstoxdb/results?utf8=\langle &search=C10-12+chloroalkanes

C10-12 chloroalkanes

108171-26-2 | DTXSID10872316

1 Searched by DSSTox_Substance_Id: Found 1 result for 'DTXSID10872316'.

Presence in Lists

Record Information

Quality Control Notes

Related Chemicals

Found 3 chemicals





Acknowledgements

...after almost 1000 emails, dozens of phone calls, a LOT of extremely valuable input and uncountable hours of late night, after hours work, a HUGE thanks goes to Tony Williams and his team at the Dashboard





Stellan Fischer, KEMI



Kevin Thomas (UQ), Saer Samanipour (NIVA), Nikiforos Alygizakis (EI),





Andrew McEachran, ORISE Fellow, US EPA

Questions?



Dashboard:

https://comptox.epa.gov/

Contact:

emma.schymanski@eawag.ch











Mixtures (work in progress)

Mixture identification and curation

- Step 1: Identify all mixture components
- Step 2: Create SMILES, split into components, retain "non-salts"
- Step 3: Create MS-ready identifiers, compare retained components
 - Single structure => all the same, treat as individual
 - Otherwise save as Level 7 for validation ...

