BIG DATA IN ENVIRONMENTAL SCIENCE AND TOXICOLOGY

superfund.tamu.edu

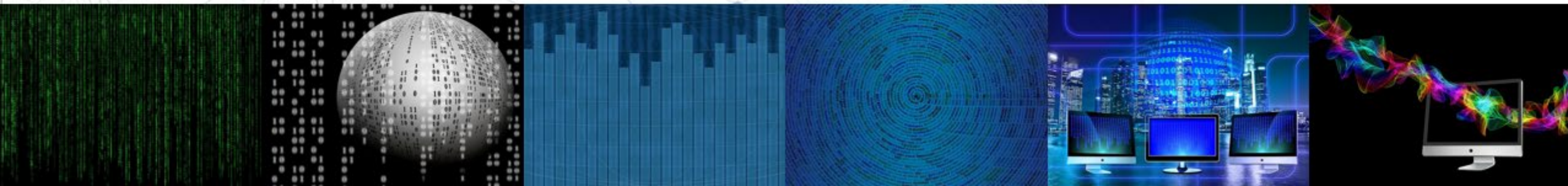
TEXAS A&M UNIVERSITY
Superfund Research Center

Welcome!

This Session will Begin at 2:00 pm Eastern US Time

HOW TO PLACE YOUR RESEARCH QUESTIONS OR RESULTS INTO THE CONTEXT OF THE "LEGACY" TOXICOLOGY DATA ?

Antony Williams - US Environmental Protection Agency

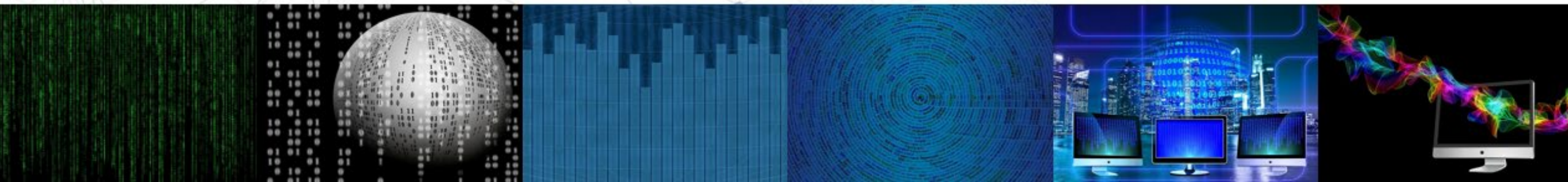


BIG DATA IN ENVIRONMENTAL SCIENCE AND TOXICOLOGY

TEXAS A&M UNIVERSITY
Superfund Research Center

- All participants are muted to enable the speaker to present without interruption.
- Please rename yourself and designate Full Name and Affiliation.
- Use the reaction icon at the bottom of your screen to raise your hand.

- This meeting will be recorded, and posted on the @tamusuperfund website https://superfund.tamu.edu/big-data-series-2021/ in the coming weeks.



http://www.orcid.org/0000-0002-2668-4821

EPA United States Environmental Protection Agency

How to Place Your Research Question or Results into the Context of Legacy Data

Antony John Williams

williams.antony@epa.gov

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

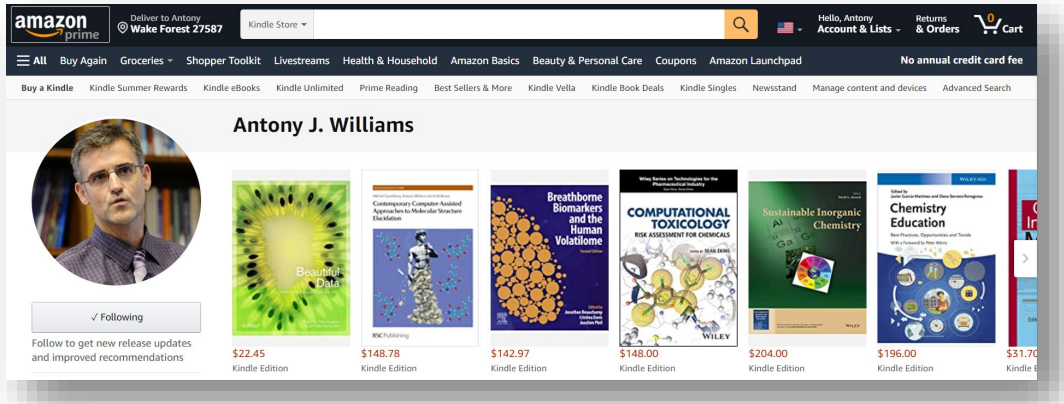
My background…

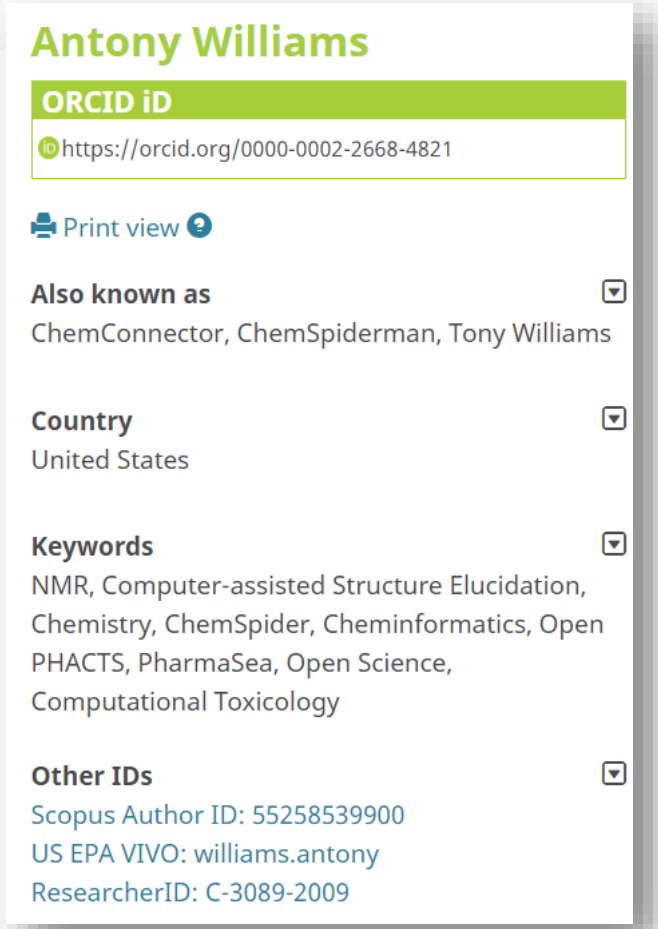
- Worked in
  - Government lab as postdoc
  - Academia as NMR Facility Manager
  - Fortune 500 Company as NMR Tech Leader
  - Small start-up company as product manager, marketing manager, Chief Science Officer
  - Consultant – chemistry informatics industry
  - Company owner – created ChemSpider
  - Royal Society of Chemistry (bought ChemSpider)
  - US-EPA – cheminformatician and “connector”

Who am I today?
https://orcid.org/0000-0002-2668-4821

EPA United States Environmental Protection Agency

- Computational chemist at the US-EPA – **scientist**
- Responsibility for cheminformatics projects, internal & external collaborations, "product marketing"– **cheminformatician**
- Work with a team of people developing software solutions – "**product & project manager**"
- Scientific publications, books, blogger – **author;** I am @ChemConnector – **social networker**





Abbreviations

- CompTox - Computational Toxicology
- DSSTox - Distributed Structure Searchable Toxicity DB
- CASRN - Chemical Abstracts Registry Number
- InChI - International Chemical Identifier
- QMRF – QSAR Model Report Format
- ToxVal - Toxicity Value Database
- OPERA - OPEn structure–activity Relationship App
- TEST - Toxicity Estimation Software Tool
- ToxCAST - Toxicity Forecaster
- SDF - Structure data file

Learning Objectives

- A *very short* overview of cheminformatics focused on
  - Chemical identifiers and some associated challenges
  - Molecular fingerprints
  - Molecular similarity
  - Structure-based modeling (QSAR/QSPR/QSUR)

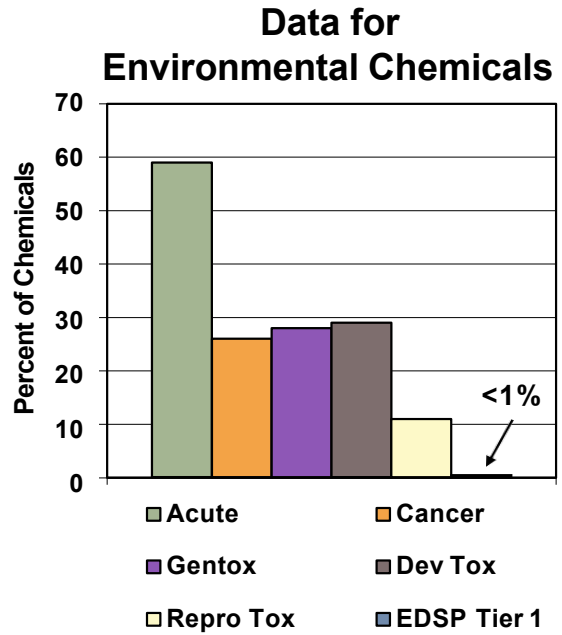
- An overview of the CompTox Chemicals Dashboard and how it can help to:
  - Search, source, visualize and download data for singleton or thousands of chemicals
  - Perform real-time prediction calculations and read-across
  - Navigate into dozens of other online resources that contain additional data

Problem: Too Many Chemicals and Too Few Resources

- Fast characterization of human and ecological risk posed by existing and emerging chemicals is a critical challenge
- Chemistry never stops. But there is sparse and distributed data…

CAS® A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

CAS REGISTRY® contains more than **171 million unique organic and inorganic chemical substances**, such as alloys, coordination compounds, minerals, mixtures, polymers and salts, and more than 68 million protein and DNA sequences



Modified from Judson *et al.*, EHP 2010

Solution

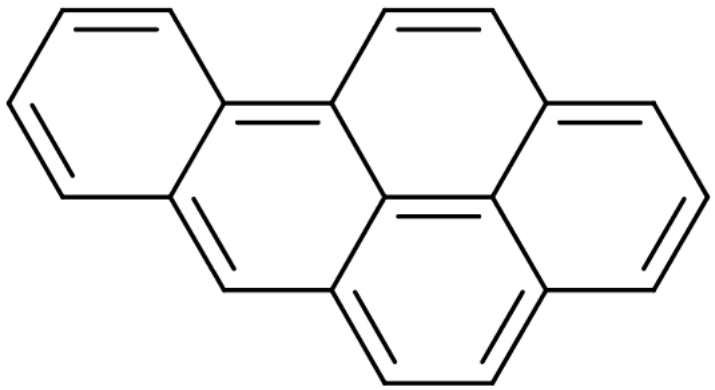
- Develop a “first-stop-shop” for environmental chemical data to support EPA and partner decision making:
  - **Centralized location** for relevant chemical data
  - Chemistry, exposure, hazard and dosimetry
  - Combination of existing data and predictive models
  - Publicly accessible, periodically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

Cheminformatics and the Dashboard

- Cheminformatics is the application of computer science and informatics-based approaches to:
  - Represent chemical structures, substances and reactions
  - Store chemistry-related data
  - Search for chemistry related data
  - Model data sets to provide predictive capabilities
  - Visualize and analyse chemistry related data

- The US-EPA uses cheminformatics (and bioinformatics) to manipulate, integrate, store, model and deliver access to our data. The CompTox Chemicals Dashboard is built on a solid cheminformatics foundation

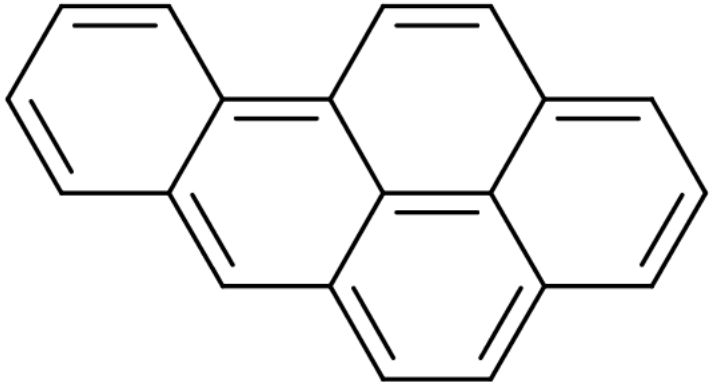
Types of Chemical Identifiers



- Structural Identifiers
- The visual depiction
- Multiple electronic formats
- InChI (Key): FMMWHPNWAFZXNH-UHFFFAOYSA-N

- Common Name: Benzo(a)pyrene
- Systematic Name: Benzo[pqr]tetraphene
- CAS Registry Number(s) : 50-32-8
- Lots of other “common names and trade names”

Information Associated with a Chemical Structure?



INTRINSIC PROPERTIES

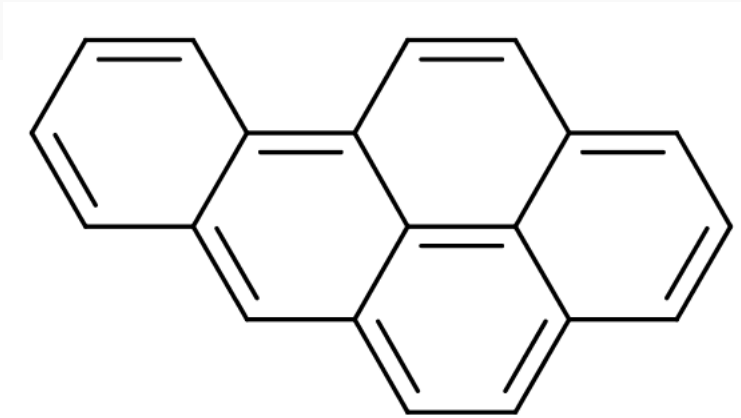
- Formula : $C_{20}H_{12}$
- Molecular weight: 252.316 g/mol
- Monoisotopic Mass: 252.093900 g/mol

MEASURED PROPERTIES

- LogKow 6.13
- Melting Pt177°C
- Boiling Pt 485°C
- ….and many more

How to Store a Chemical Structure

- Multiple approaches:
  - Names and identifiers
  - 2D or 3D structure “molfile”



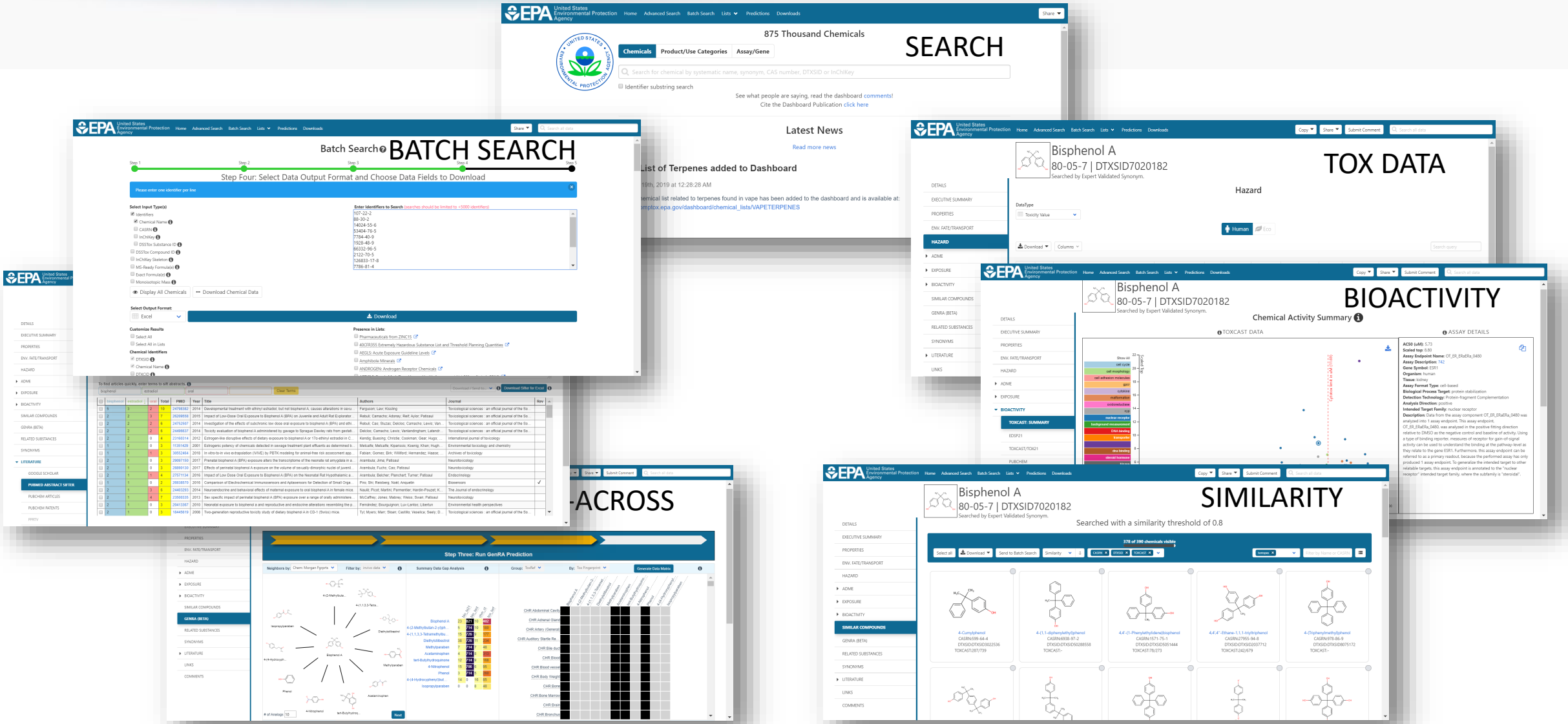
```
  Mrv1533009301517202D

  0  0  0     0  0            999 V3000
M  V30 BEGIN CTAB
M  V30 COUNTS 20 24 0 0 0
M  V30 BEGIN ATOM
M  V30 1 C 5.3801 0 0 0
M  V30 2 C 6.9201 0 0 0
M  V30 3 C 7.6901 -1.33 0 0
M  V30 4 C 9.2302 -1.33 0 0
M  V30 5 C 10.0003 -2.67 0 0
M  V30 6 C 9.2302 -4.0001 0 0
M  V30 7 C 7.6901 -4.0001 0 0
M  V30 8 C 6.9201 -5.3301 0 0
M  V30 9 C 5.3801 -5.3301 0 0
M  V30 10 C 4.6201 -4.0001 0 0
M  V30 11 C 5.3801 -2.67 0 0
M  V30 12 C 6.9201 -2.67 0 0
M  V30 13 C 4.6201 -1.33 0 0
M  V30 14 C 3.0801 -1.33 0 0
M  V30 15 C 2.31 -2.67 0 0
M  V30 16 C 3.0801 -4.0001 0 0
M  V30 17 C 0.77 -2.67 0 0
M  V30 18 C 0 -1.33 0 0
M  V30 19 C 0.77 0 0 0
M  V30 20 C 2.31 0 0 0
M  V30 END ATOM
M  V30 BEGIN BOND
M  V30 1 2 1 2
M  V30 2 1 1 13
M  V30 3 1 2 3
M  V30 4 2 3 4
M  V30 5 1 3 12
M  V30 6 1 4
```

- SMILES:
  - c1cc2c3ccc4cccc5ccc(cc2cc1)c3c45
  - C1=CC2=CC3=CC=C4C=CC=C5C=CC(=C2C=C1)C3=C45
  - and many other variants….
- InChI=1S/C20H12/c1-2-7-17-15(4-1)12-16-9-8-13-5-3-6-14-10-11-18(17)20(16)19(13)14/h1-12H
- InChIKey: FMMWHPNWAFZXNH-UHFFFAOYSA-N

If We Database Chemical Structures…

- …then we can search the dataset by inherent **structural** properties
  - Formula
  - Mass
  - Substructure
  - Structural similarity
- …we can **integrate** other info into the database for retrieval
- …available data, both experimental and predicted, is a click away
- …data can be downloaded, distributed and shared
- …linking out to other resources enabled by adopting specific standards
- …structure collections, with associated data, are available for modeling

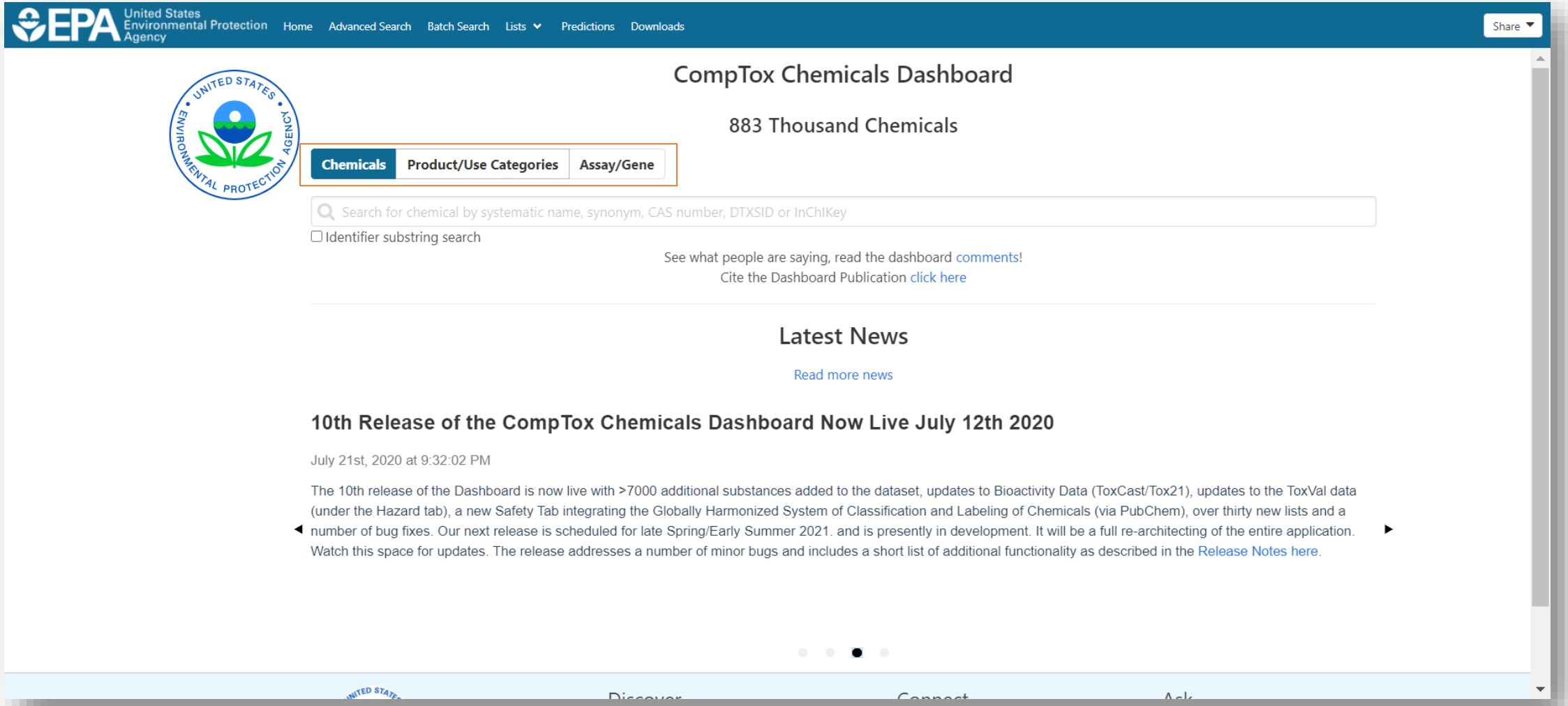
CompTox Chemicals Dashboard
https://comptox.epa.gov/dashboard



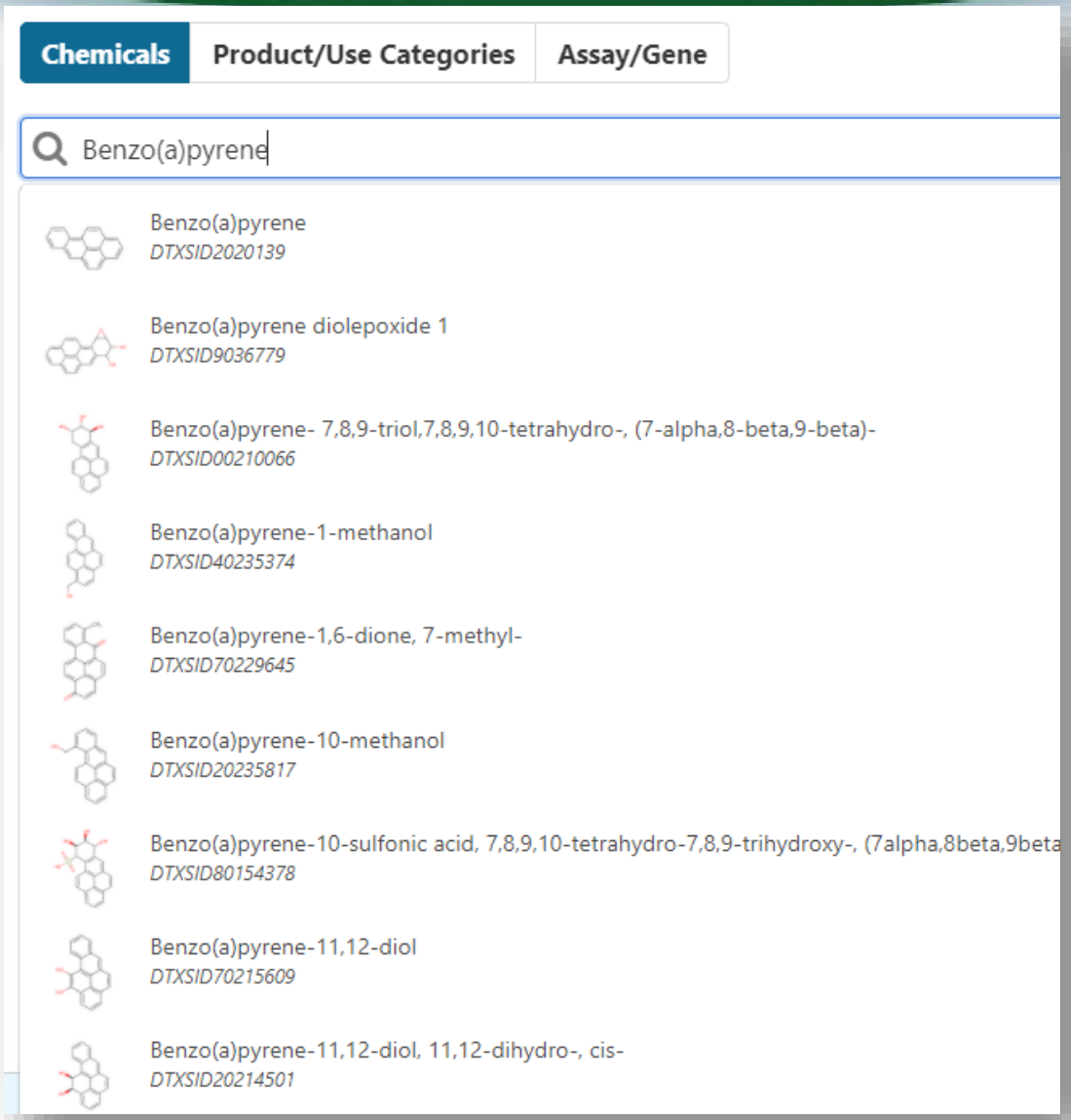
CompTox Chemicals Dashboard

EPA United States Environmental Protection Agency

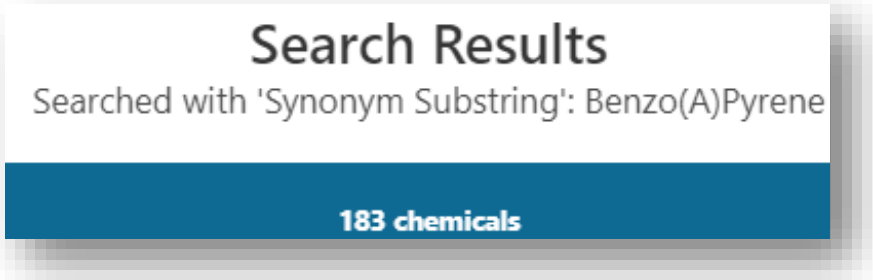
883k Chemical Substances



BASIC Search

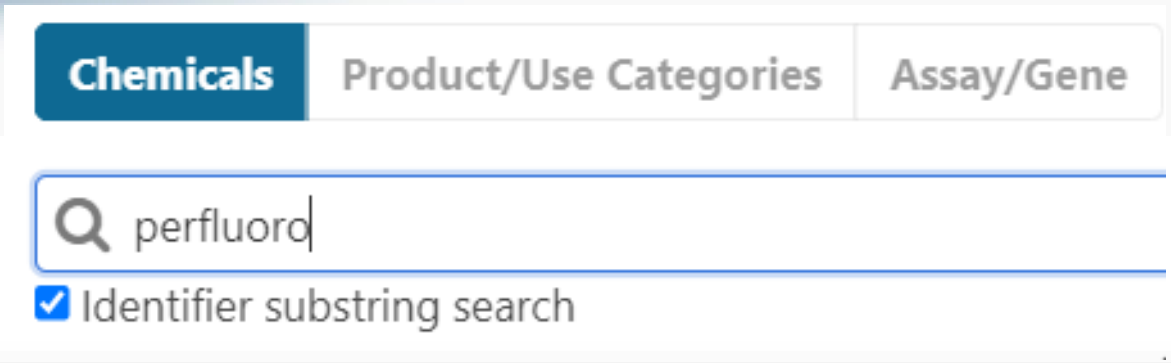


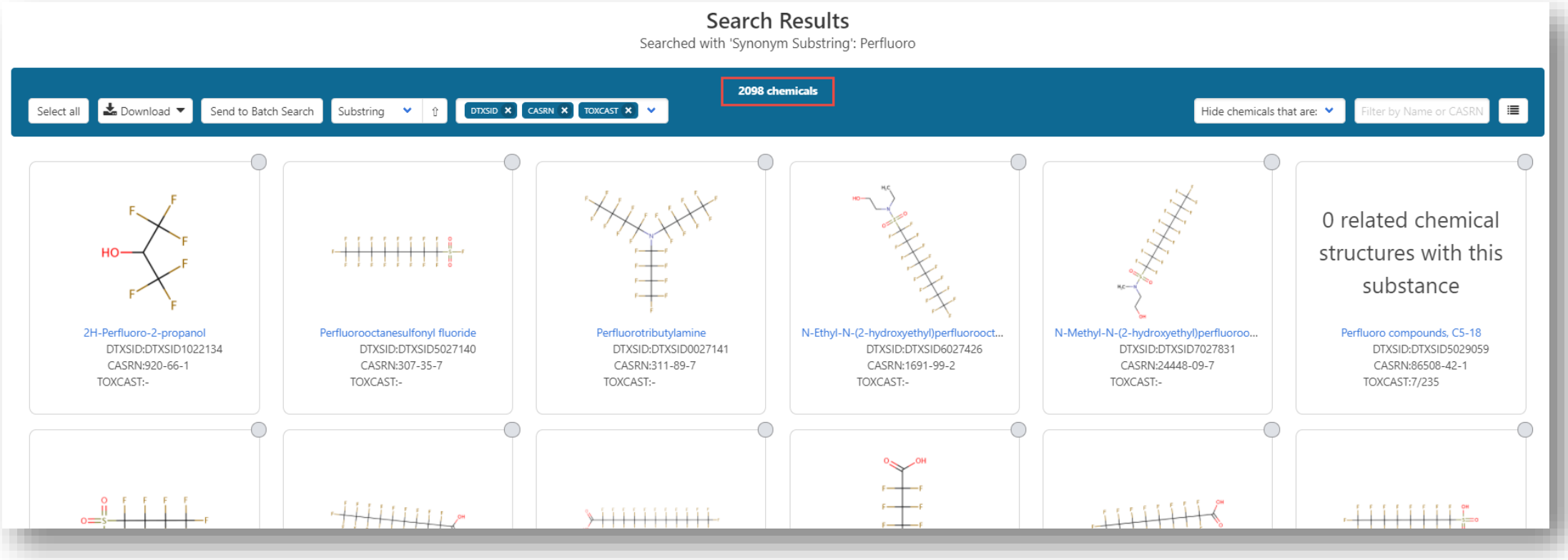
- Type ahead search using Names, synonyms and CASRNs
- Millions of identifiers
- Substring search



Search for classes of chemicals

- Examples: “perfluoro”

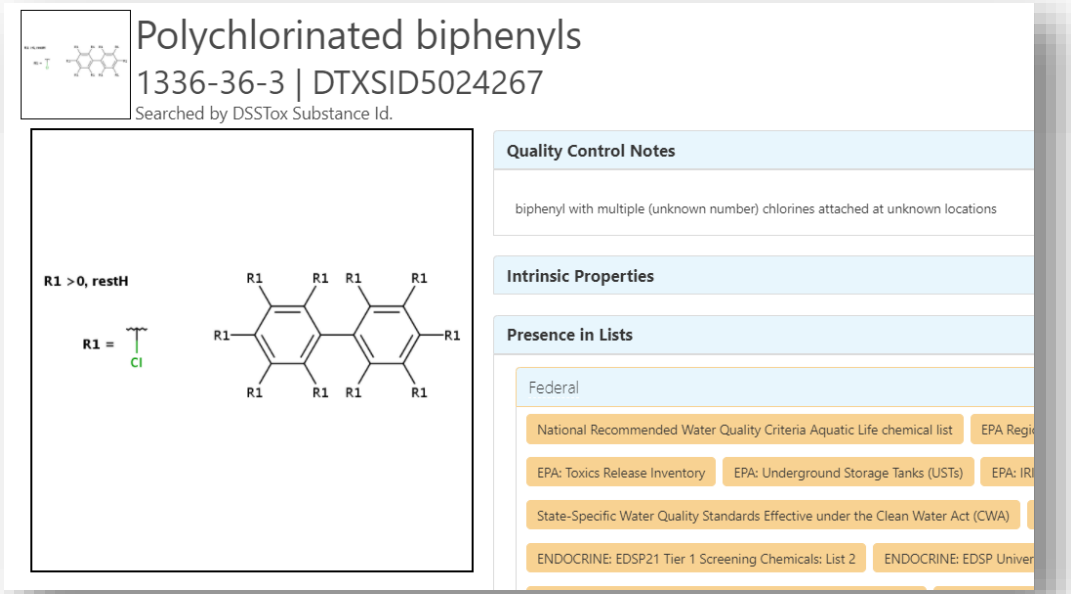




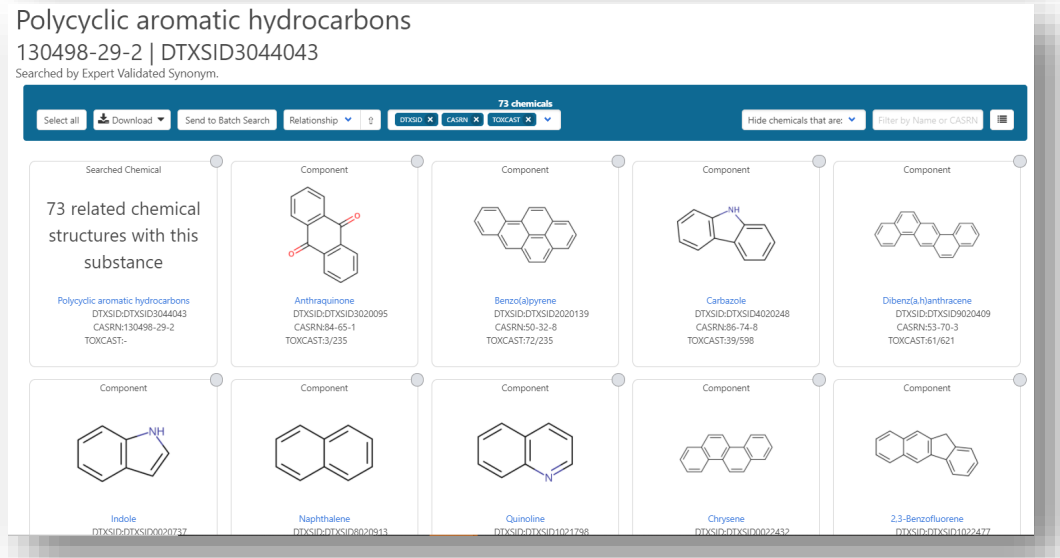
Search for classes of chemicals

Examples:

- PCBs (or polychlorinated biphenyls)



- Polycyclic aromatic hydrocarbons



QUESTION 1

- How many “conazoles” are in the dashboard?

29 56 125 321

Challenges with Nomenclature

- Be CAREFUL with names! There is a LOT of confusion in the public domain. CHOOSE sources wisely!
- There are MANY public databases but not many are curated
- All public databases have value but not many curate data

- Example: METHANE on PubChem https://pubchem.ncbi.nlm.nih.gov/compound/297

CAS Registry Numbers on PubChem

EPA United States Environmental Protection Agency

2.3.1 CAS

74-82-8

- CAMEO Chemicals; CAS Common Chemistry; ChemIDplus; DrugBank; EPA Chemicals under the TSCA; EPA DSSTox; European Chemicals Agency (ECH...

8006-14-2

- CAMEO Chemicals; EPA Chemicals under the TSCA; EPA DSSTox; European Chemicals Agency (ECHA)

7440-44-0

- ChemIDplus

7782-40-3

- ChemIDplus

7782-42-5

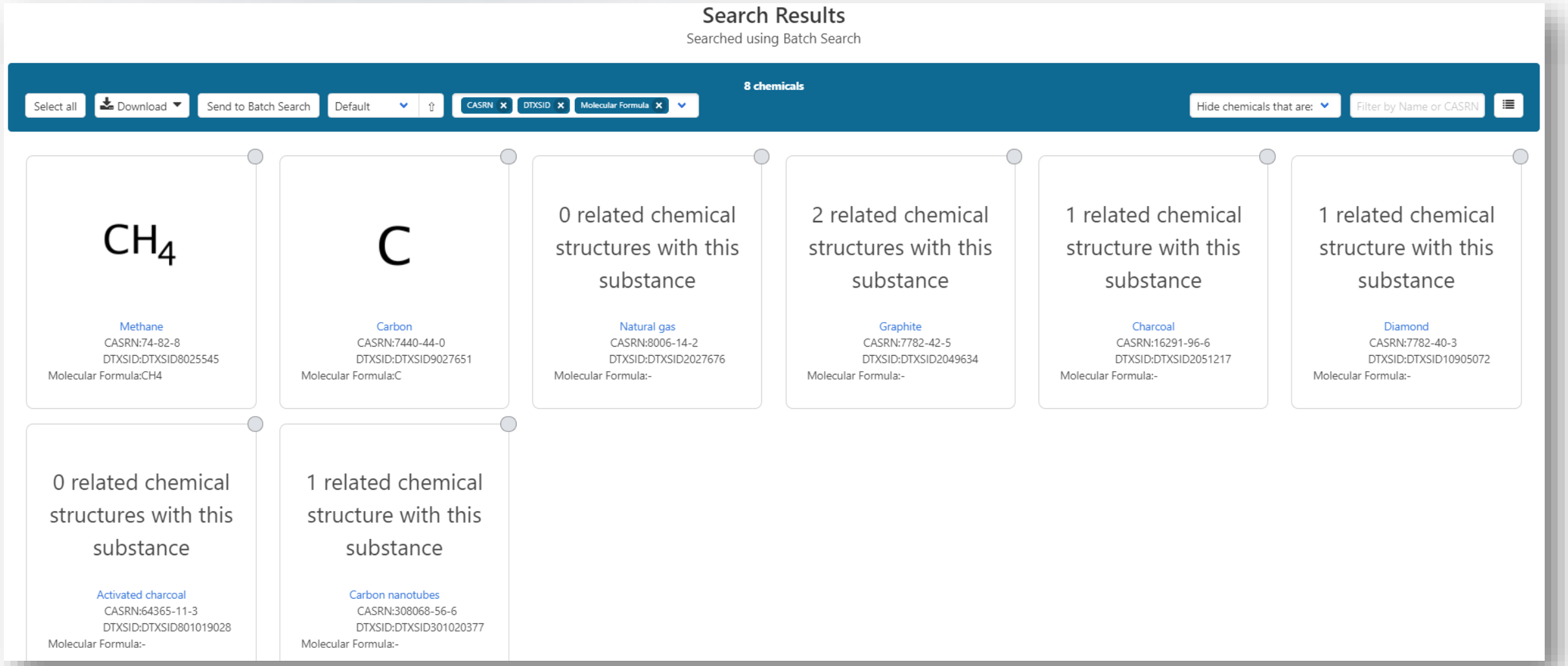
- ChemIDplus

16291-96-6

- ChemIDplus

64365-11-3

CASRN lookup on the dashboard



Methane is Diamond and Nanotubes?

- These are all Depositor Names for Methane ☹

2.4.2 Depositor-Supplied Synonyms

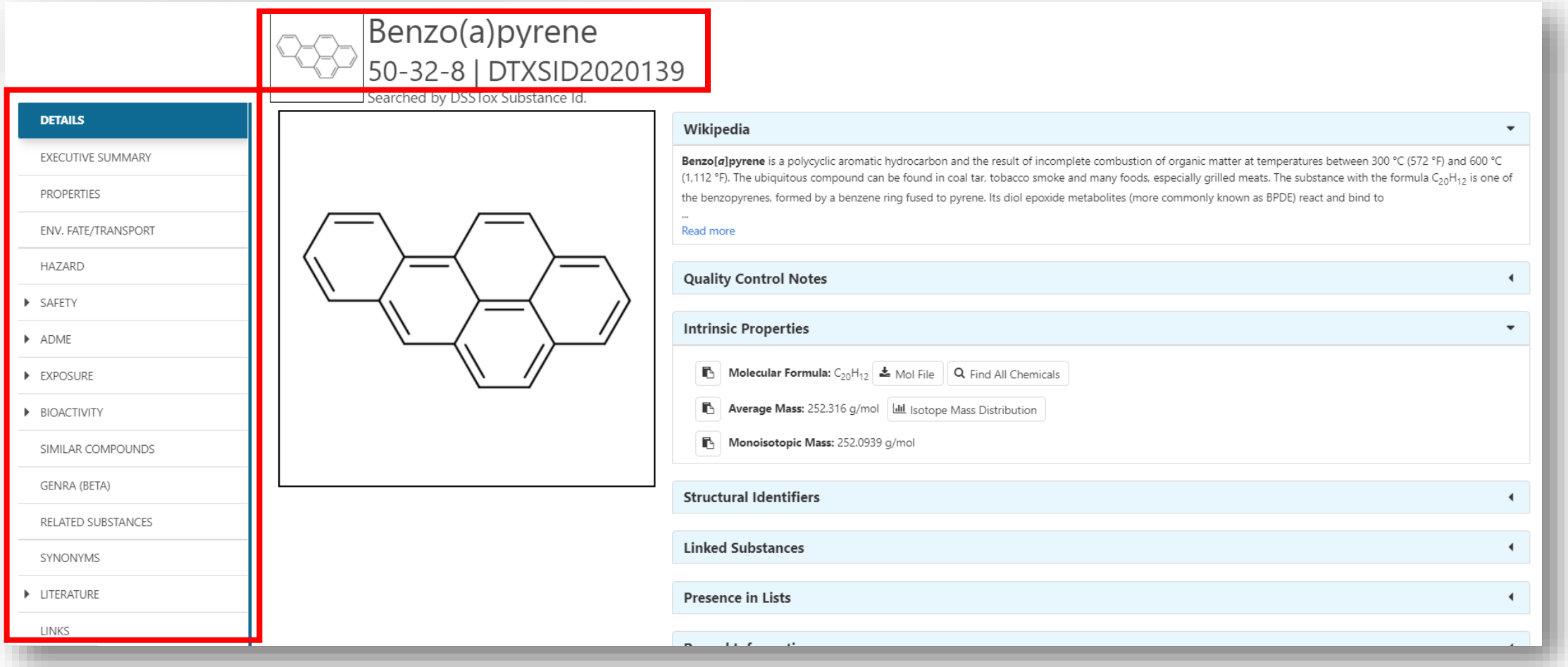
UN 1971 (Salt/Mix)	Fullerene soot, (as produced)	Carbon Nanotube sponges XFCN01	DTXSID9027651
UN 1972 (Salt/Mix)	MWNTs Butyl acetate suspension	Carbon Nanotube sponges XFCN07	Graphite electrode, rotrode disc
Activated carbon, pellets 3mm	QuadraPure C, 0.3-0.8mm	Carbon Nanotube sponges XFCN08	Carbon conductive cement adhesive
Graphene quantum dots(Powder)	6GRV67N0U2	Carbon, activated, -4+8 mesh	Conductive Flexible TPU Filament
Multiwall Nanotubes 5-15 nm	GO quantum dots yellow(Powder)	Carbon, activated, 2mm & down	GO quantum dots yellow(1mg/ml)
GO quantum dots(C: 1mg/ml)	Graphene electric aqueous slurry	CHEMBL2106049	Graphite powder, -20+84 mesh
MWNTs ethyl acetate suspension	Graphene powder Physical methods	Diamond Synthesized, 95% Nano	Carbon black, Super P Conductive
Reduced Graphene Oxide@ SnO2	Reduced Graphene Oxide@ Co3O4	Diethyl Cyanomethyl Phosphonate	DTXSID50179391
Carbon nanotubes aqueous slurry	Carbon Conductive Adhesive Tapes	DTXSID8025545	NanoIntegris metallic SWCNTs70%

QUESTION 1

- How many “conazoles” are in the dashboard?

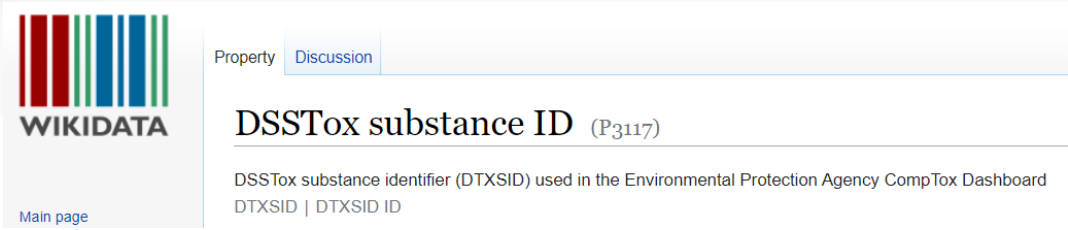
29 56 **125** 321

Detailed Chemical Pages
One more identifier – the DTXSID

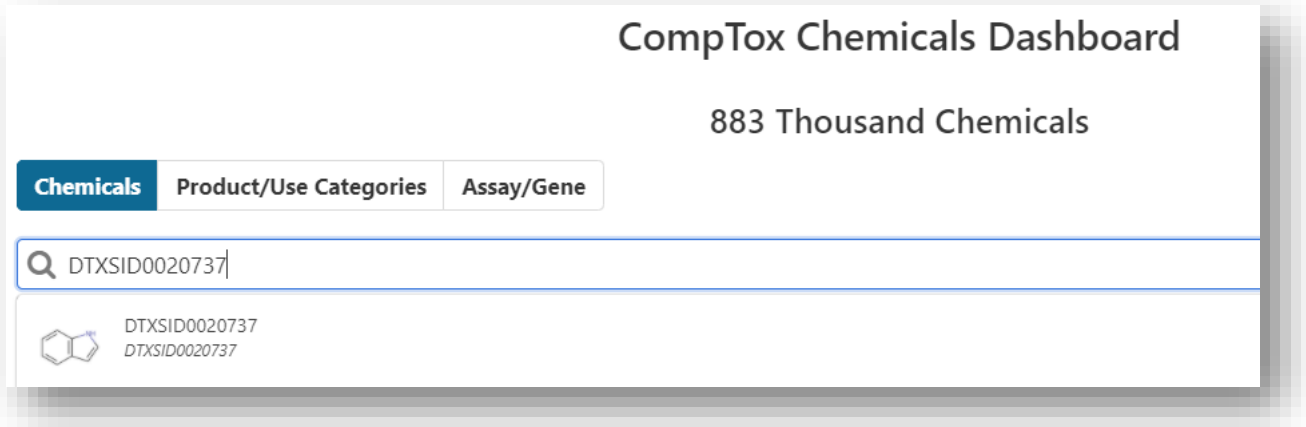


- Chemical page: Wikipedia snippet when available, intrinsic properties, structural identifiers, linked substances

DTXSID as a substance identifier

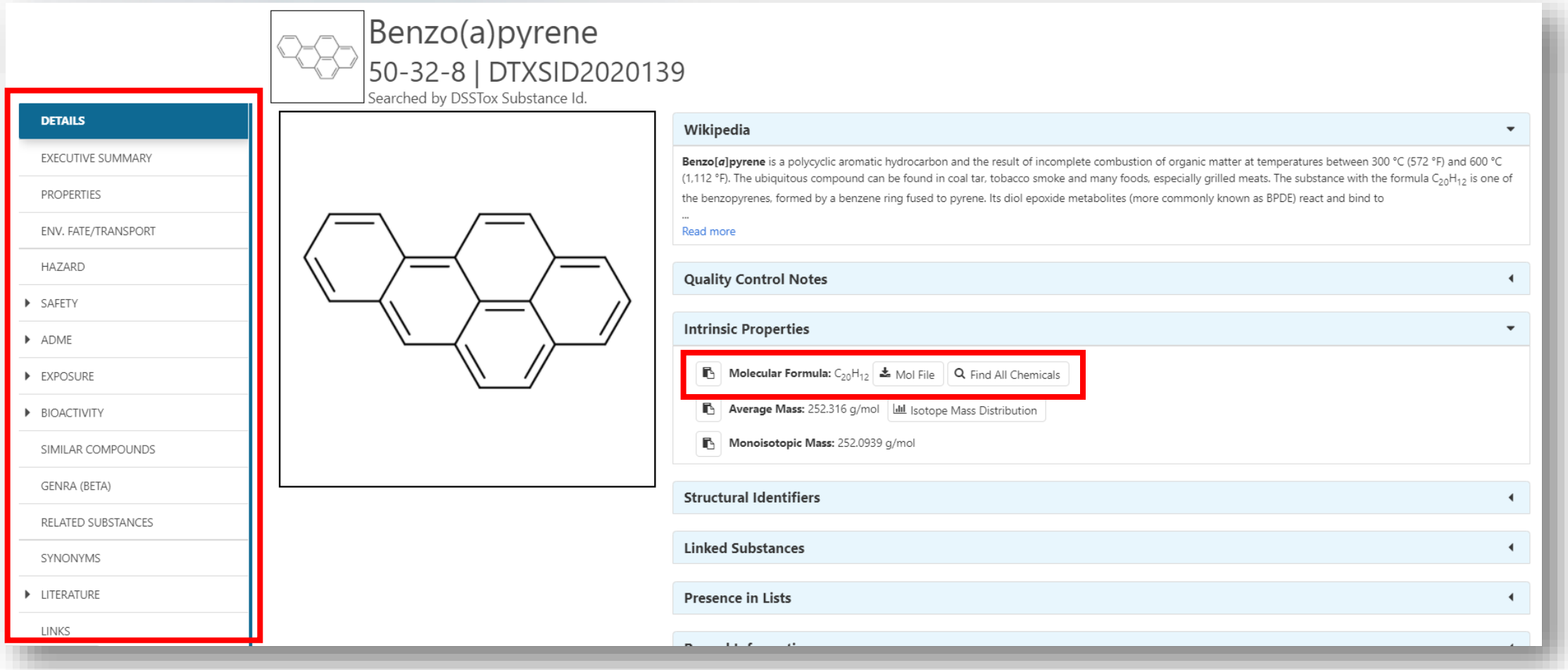


- Blessed now by Wikipedia
- In MANY public databases – ChemSpider, PubChem, eChemPortal
- Increasingly used across all EPA databases
- Easy to use/search



- URL linking: https://comptox.epa.gov/dashboard/DTXSID0020737
- ANY database can link directly via DTXSIDs…more later..

Detailed Chemical Pages
Easy Navigation



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

From the Chemical Details Page… all chemicals with same FORMULA

EPA United States Environmental Protection Agency

Intrinsic Properties

Molecular Formula: $C_{20}H_{12}$ | Mol File | Find All Chemicals

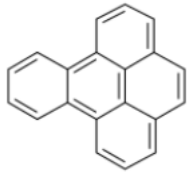
Search Results

Searched by Exact Molecular Formula: C20H12.

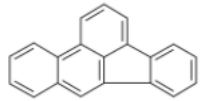
27 chemicals

Download | Send to Batch Search | Default | CASRN | DTXSID | Hide chemicals that are: | Filter by Name or C

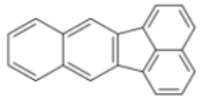
Benzo(a)pyrene
CASRN:50-32-8
DTXSID:DTXSID2020139



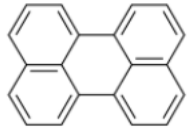
Benzo(e)pyrene
CASRN:192-97-2
DTXSID:DTXSID3023764



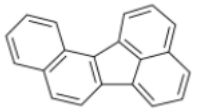
Benzo(b)fluoranthene
CASRN:205-99-2
DTXSID:DTXSID0023907



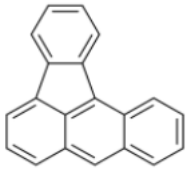
Benzo(k)fluoranthene
CASRN:207-08-9
DTXSID:DTXSID0023909



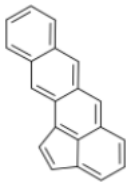
Perylene
CASRN:198-55-0
DTXSID:DTXSID4047753



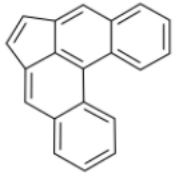
Benzo(j)fluoranthene
CASRN:205-82-3
DTXSID:DTXSID8052691



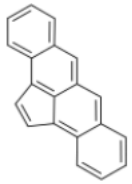
Benzo[a]fluoranthene
CASRN:203-33-8
DTXSID:DTXSID4059756



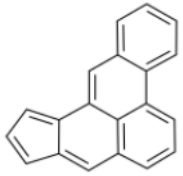
Cyclopenta(de)naphthacene
CASRN:16683-64-0
DTXSID:DTXSID80168197



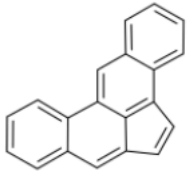
Benz(a)acephenanthrylene
CASRN:192-28-9
DTXSID:DTXSID70172748



Cyclopenta(fg)naphthacene
CASRN:19770-52-6
DTXSID:DTXSID40173469



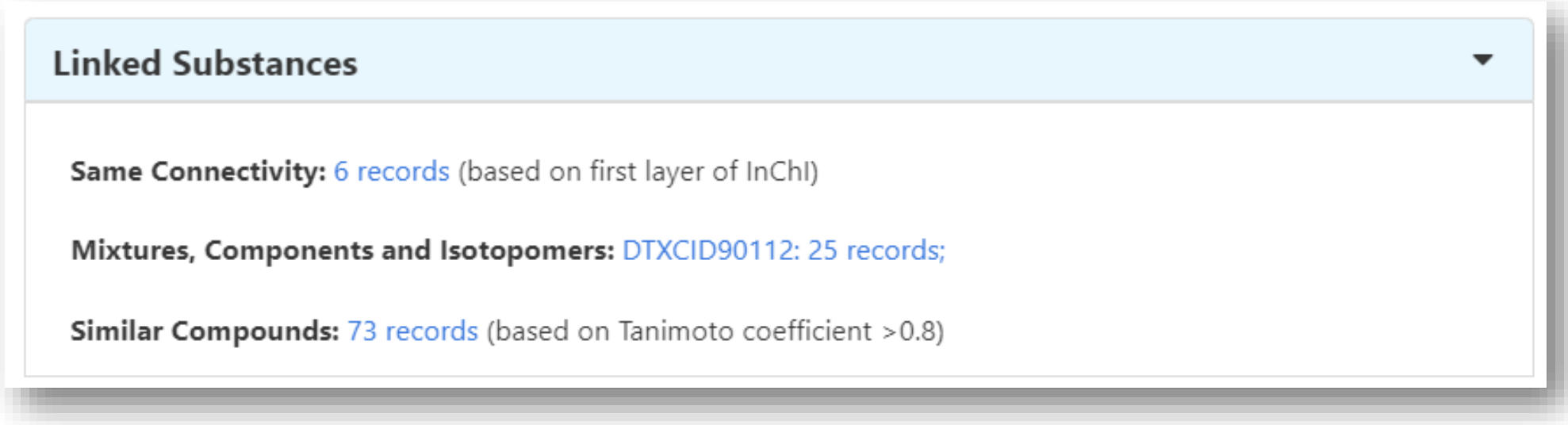
Benzo(de)cyclopent(a)anthracene
CASRN:198-46-9
DTXSID:DTXSID60173507



Benz(e)aceanthrylene
CASRN:199-54-2
DTXSID:DTXSID30173675

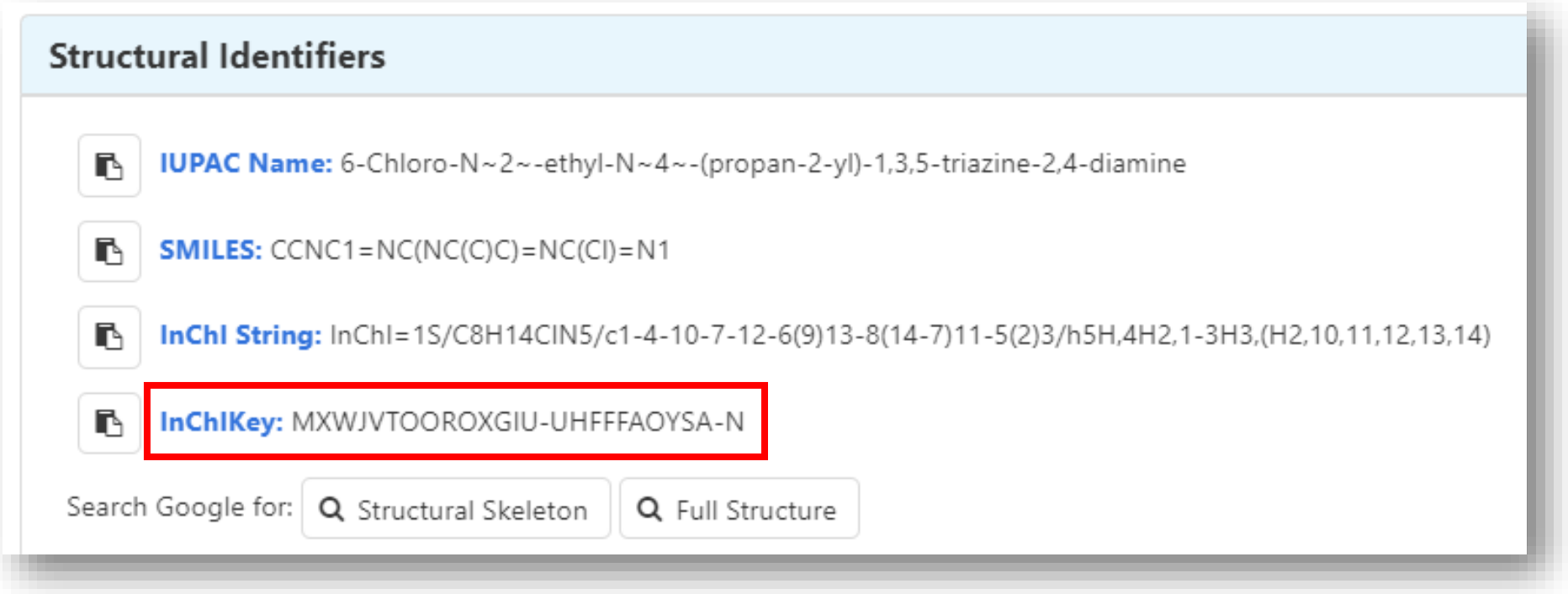
How many chemicals are associated through LINKED SUBSTANCES?

- Atrazine, is a herbicide – in MANY commercial products
- The dashboard has salt forms, isotopically labelled forms, multicomponent forms
- How do we identify what they are???



A little more about the InChI

- An InChIKey is made up of two blocks…
  - Block 1 – “the connectivity” of atoms and bonds
  - Block 2 – isotopes, charge, stereo



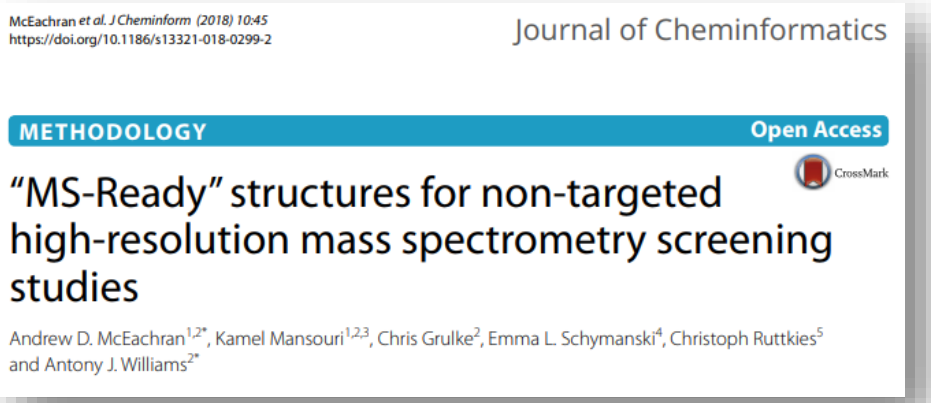
- The InChIKey is VERY USEFUL

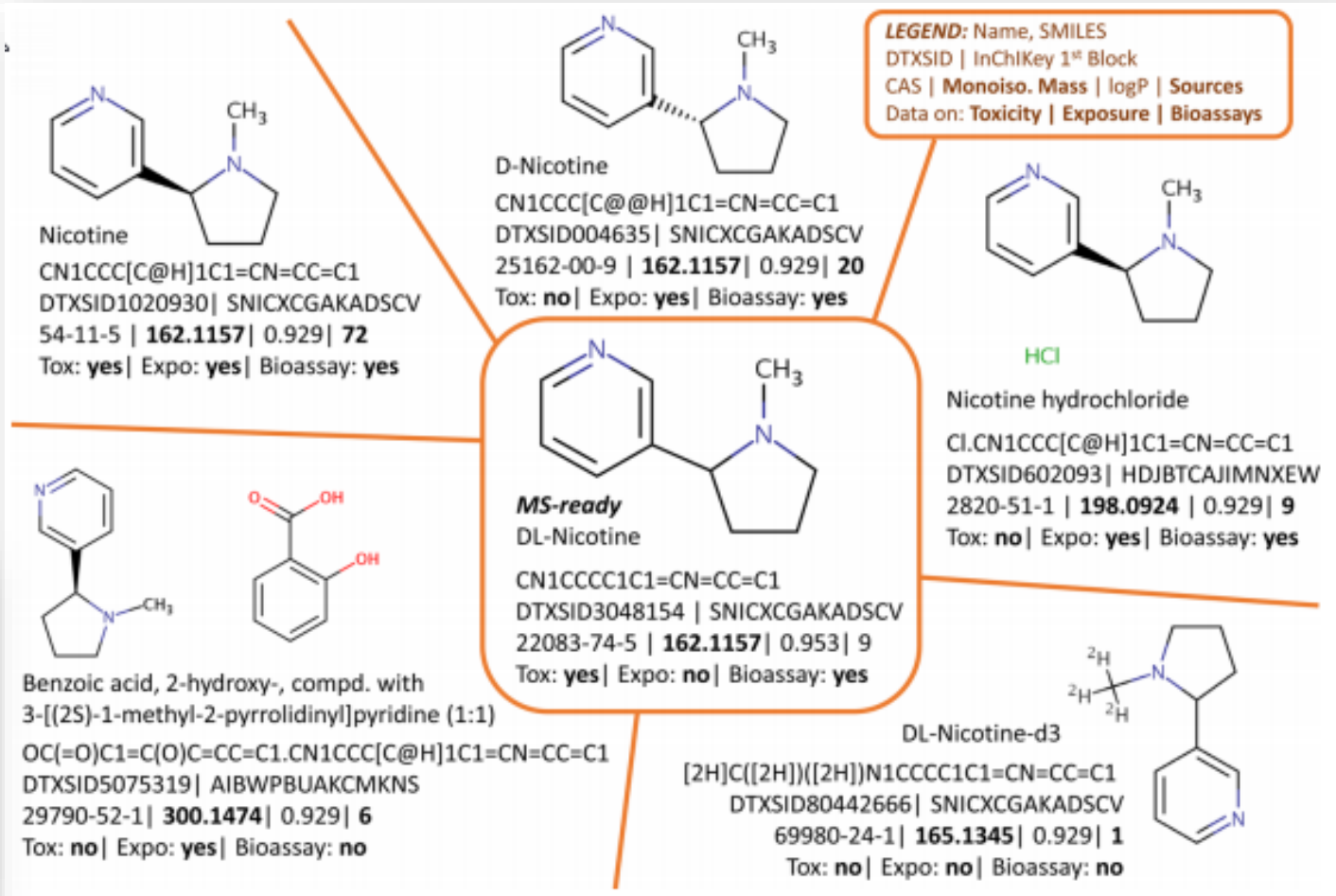
Searching using InChI

- Demo an internet search using InChIs – Cholesterol has the InChIKey: HVYWMOMLDIMFJA-DPAQBDIFSA-N
- Demo Atrazine – Linked Substances – Skeleton
- More about Linked Substances….

Linked Substances – more interesting

- We map chemicals together using cheminformatics approaches
- Use desalting, destereo, split multicomponents etc to map chemicals together





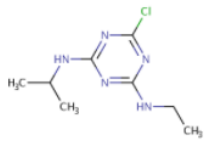
Atrazine Linked Substances

EPA United States Environmental Protection Agency

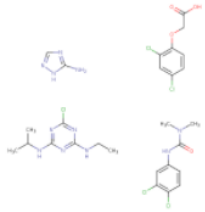
MS-Ready Mappings of Atrazine (Isotopes pre-filtered)

20 of 25 chemicals visible

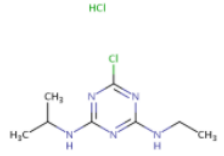
Select all | Download | Send to Batch Search | Default | DTXSID | CASRN | TOXCAST | Isotopes | Filter by Name or CASRN



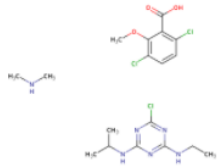
Atrazine
DTXSID:DTXSID9020112
CASRN:1912-24-9
TOXCAST:62/1024



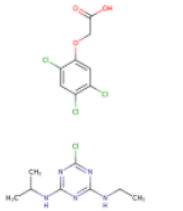
Anox M
DTXSID:DTXSID50156021
CASRN:128996-76-9
TOXCAST:-



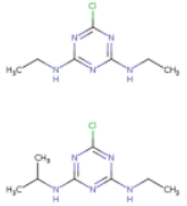
s-Triazine, 2-chloro-4-(ethylamino)-6-(is...
DTXSID:DTXSID30165459
CASRN:15386-47-7
TOXCAST:-



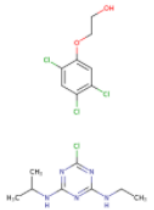
Marksman
DTXSID:DTXSID80166936
CASRN:160544-50-3
TOXCAST:-



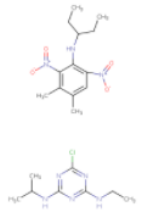
Acetic acid, (2,4,5-trichlorophenoxy)-, mi...
DTXSID:DTXSID70192527
CASRN:39283-62-0
TOXCAST:-



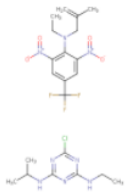
1,3,5-Triazine-2,4-diamine, 6-chloro-N,N...
DTXSID:DTXSID60192556
CASRN:39331-45-8
TOXCAST:-



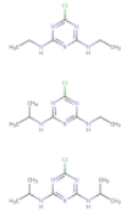
Buvinol
DTXSID:DTXSID10199555
CASRN:51602-05-2
TOXCAST:-



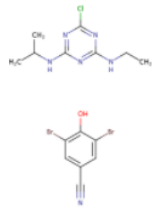
Atrazine mixture with pendimethalin
DTXSID:DTXSID10209527
CASRN:60704-01-0
TOXCAST:-



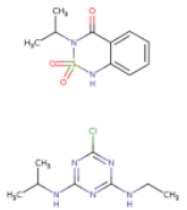
Maizor
DTXSID:DTXSID20215154
CASRN:64867-15-8
TOXCAST:-



Polytriazine
DTXSID:DTXSID00222508
CASRN:72172-70-4
TOXCAST:-



Benzonitrile, 3,5-dibromo-4-hydroxy-, m...
DTXSID:DTXSID20226063
CASRN:75084-56-9
TOXCAST:-



Bentazon / atrazine
DTXSID:DTXSID80226064
CASRN:75084-57-0
TOXCAST:-

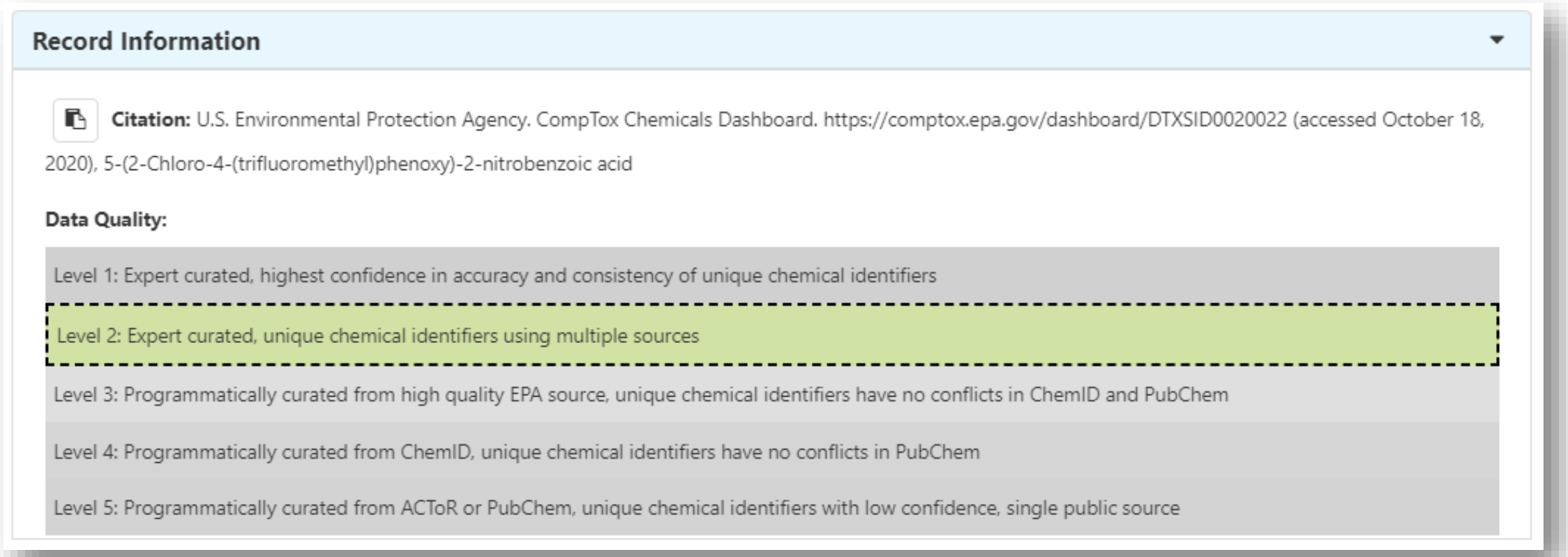
QUESTION 2

- How many salts of perfluorooctanesulfonic acid are there?

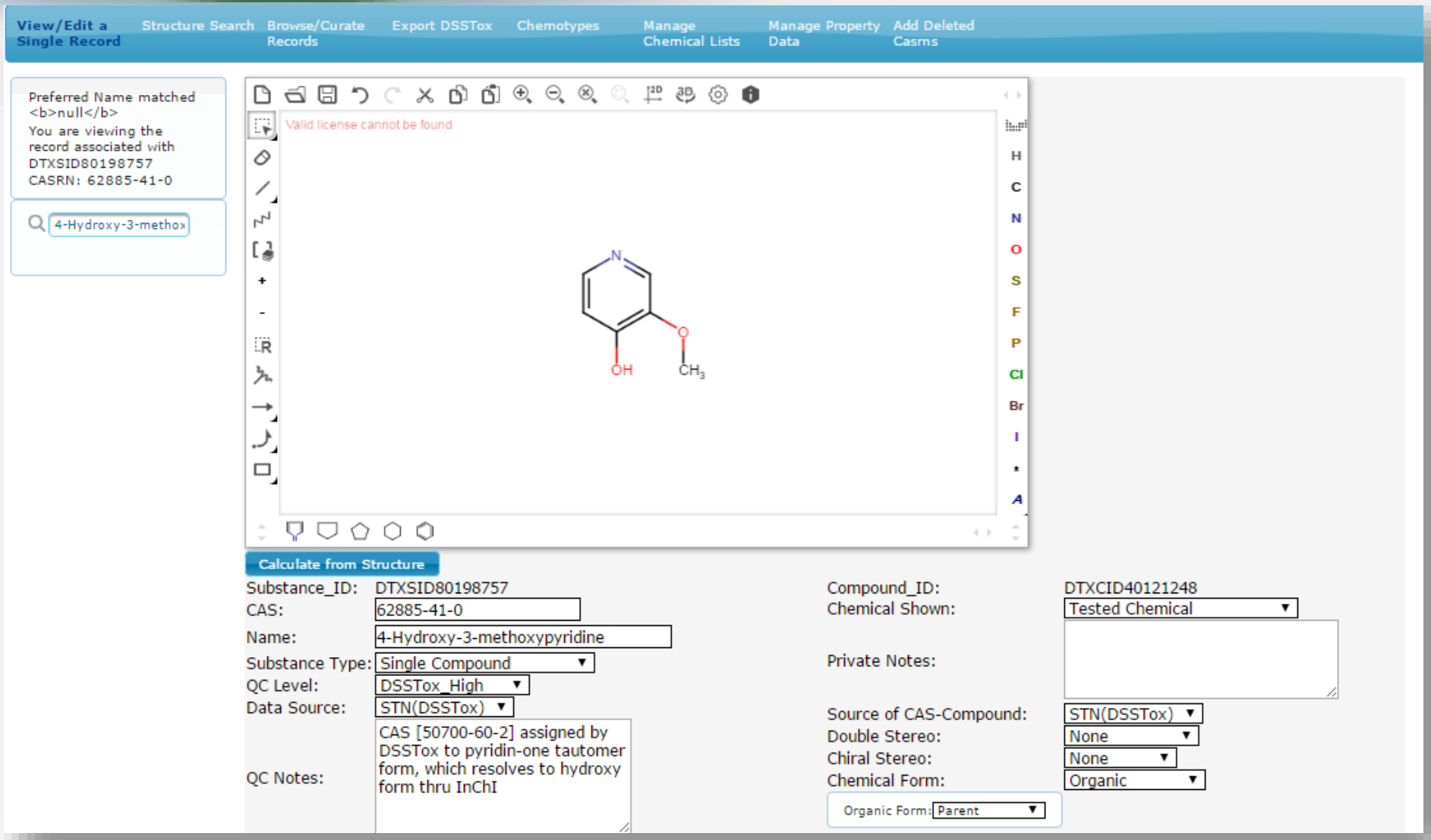
6 15 18 43

A little more about our data quality

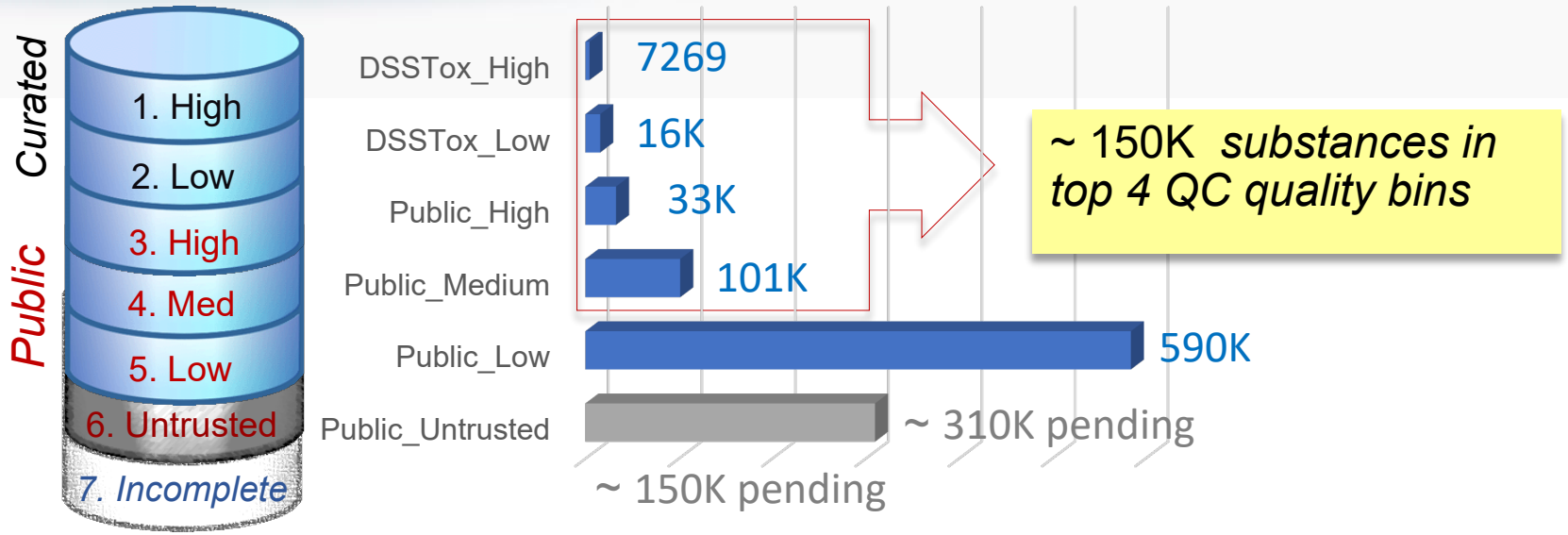
- Five full time curators register and curate data to elevate quality



Underneath the Dashboard

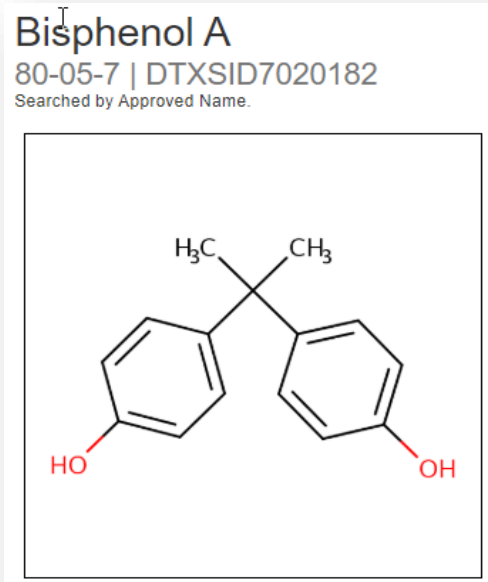


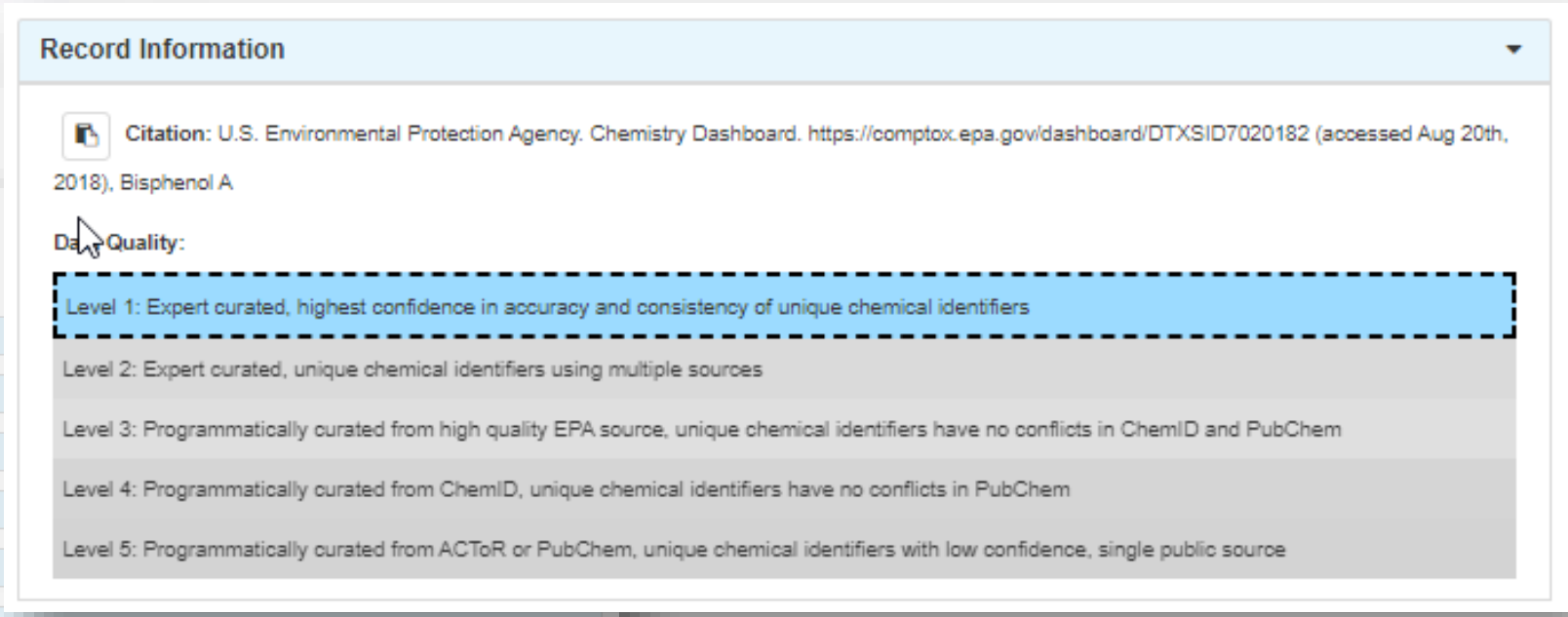
Distribution of curated data
Now at >910k substances



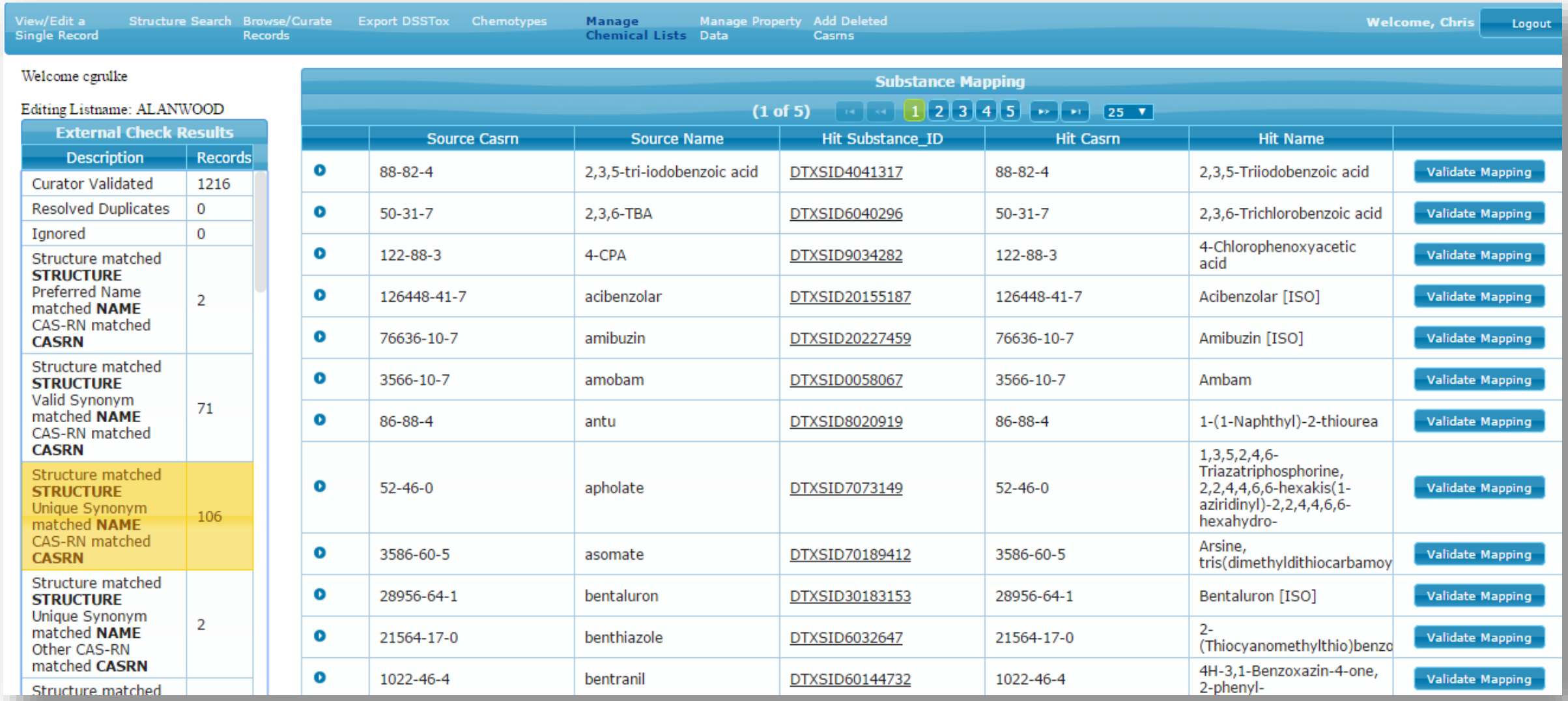
QC Levels	
DSSTox_High:	Hand curated and validated
DSSTox_Low:	Hand curated and confirmed using multiple public sources
Public_High:	Extracted from EPA SRS and confirmed to have no conflicts in ChemID and PubChem
Public_Medium:	Extracted from ChemID and confirmed to have no conflicts in PubChem
Public_Low:	Extracted from ACToR or PubChem
Public_Untrusted:	Postulated, but found to have conflicts in public sources

Record Information Quality Flags





ChemReg Curation
The ALANWOOD Pesticide Set



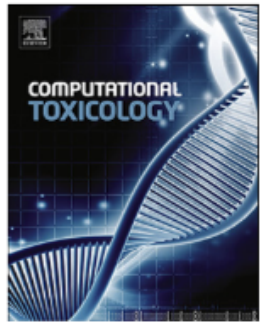
A little more about our data quality

Computational Toxicology 12 (2019) 100096

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com/locate/comtox



EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research

Christopher M. Grulke[a], Antony J. Williams[a], Inthirany Thillanadarajah[b], Ann M. Richard[a,*]

[a] National Center for Computational Toxicology, Office of Research & Development, US Environmental Protection Agency, Mail Drop D143-02, Research Triangle Park, NC 27711, USA

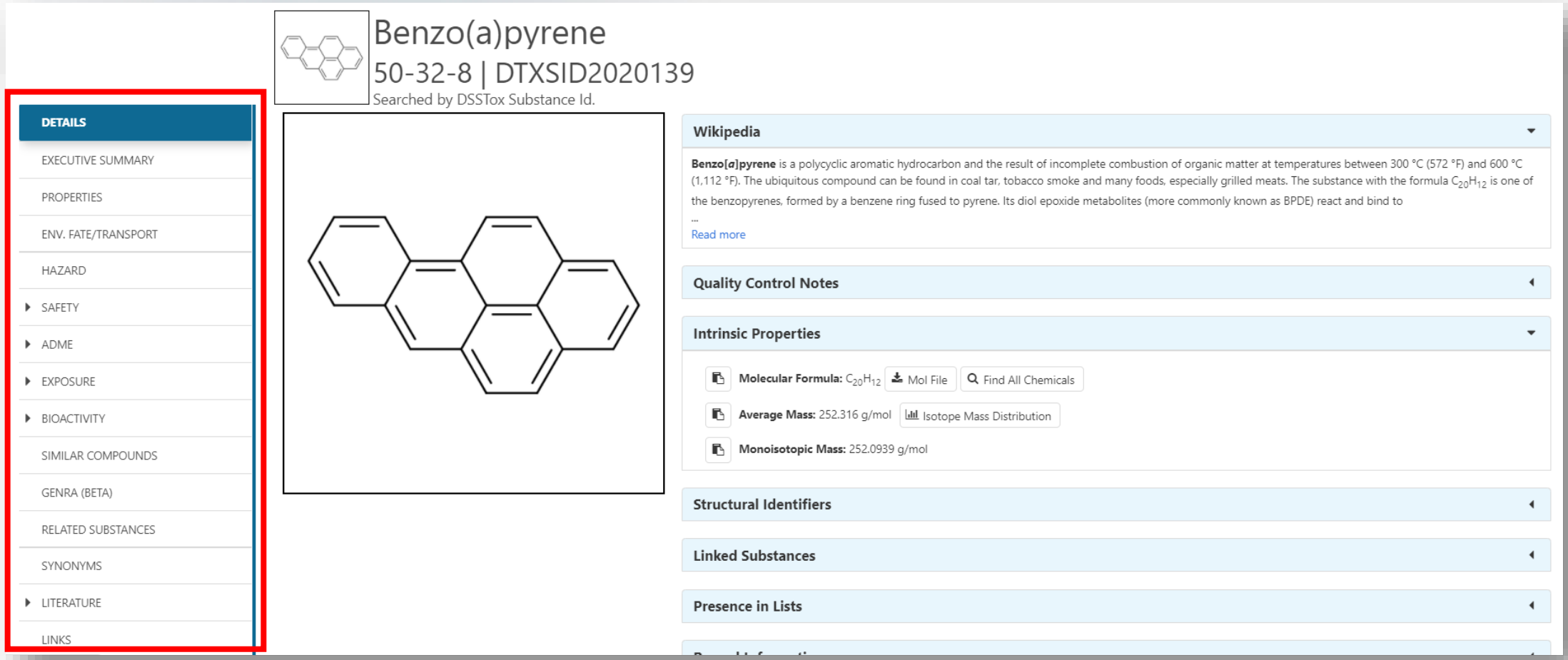
[b] Senior Environmental Employment Program, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

QUESTION 2

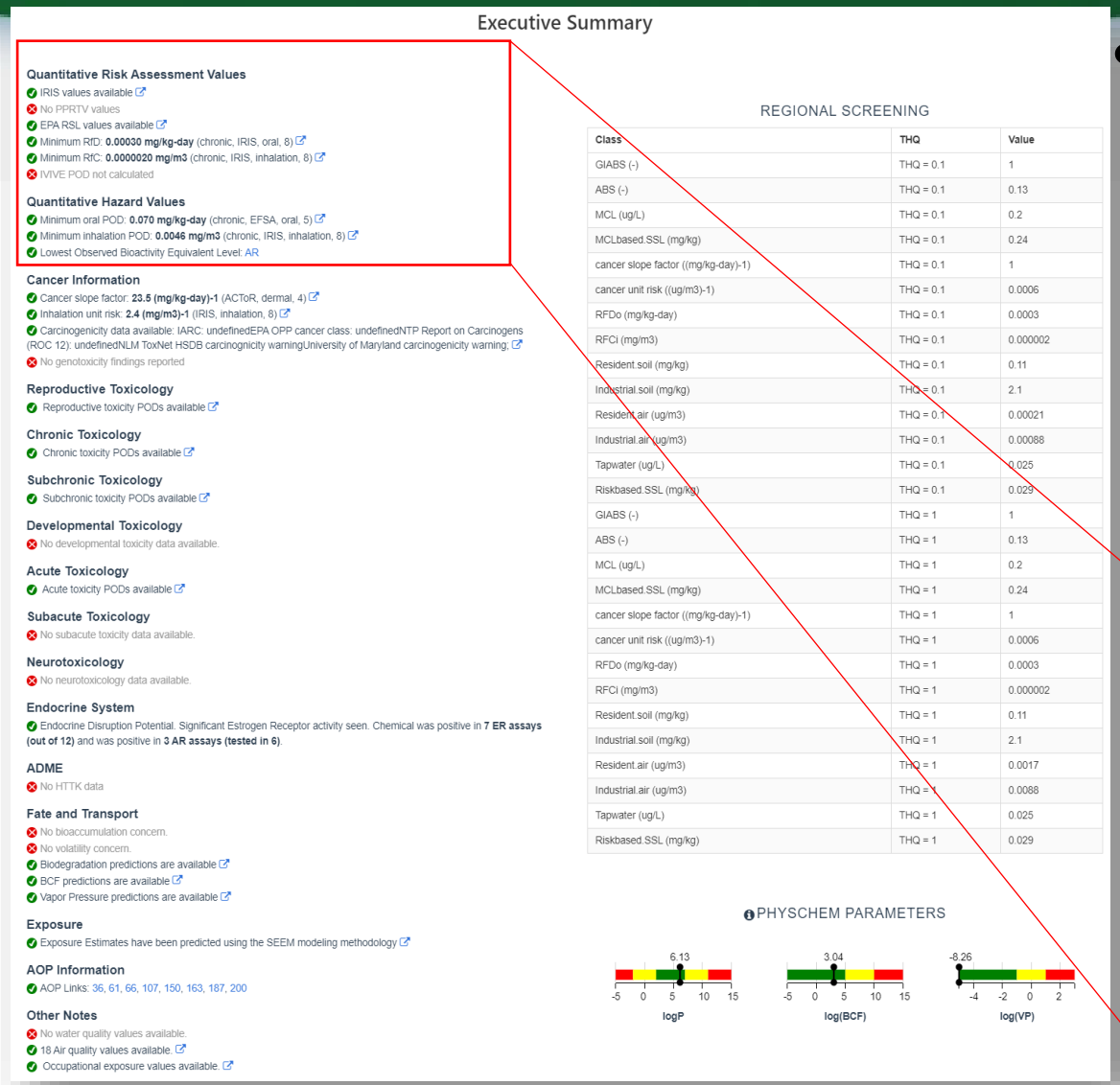
- How many salts of perfluorooctanesulfonic acid are there?

6 **15** 18 43

Navigating data via the Left Hand Tabs

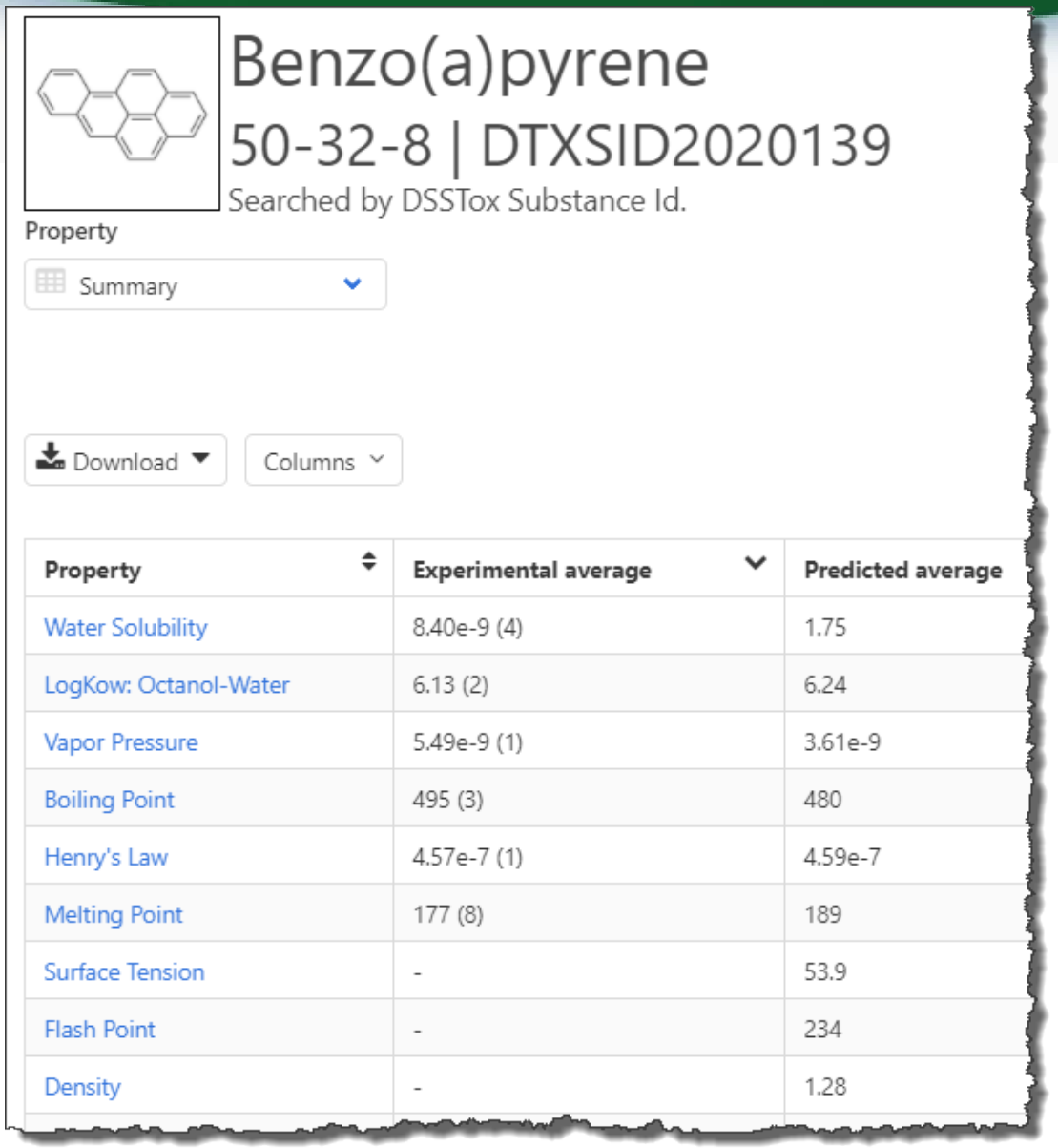


“Executive Summary”



- Overview of toxicity-related info
  - Quantitative values
  - Info re. toxicology subsets
  - Physchem. and Fate & Transport
  - Adverse Outcome Pathway links
  - *In vitro* bioactivity summary plot

Experimental and Predicted Data



Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

Property

Summary

Download Columns

Property	Experimental average	Predicted average
Water Solubility	8.40e-9 (4)	1.75
LogKow: Octanol-Water	6.13 (2)	6.24
Vapor Pressure	5.49e-9 (1)	3.61e-9
Boiling Point	495 (3)	480
Henry's Law	4.57e-7 (1)	4.59e-7
Melting Point	177 (8)	189
Surface Tension	-	53.9
Flash Point	-	234
Density	-	1.28

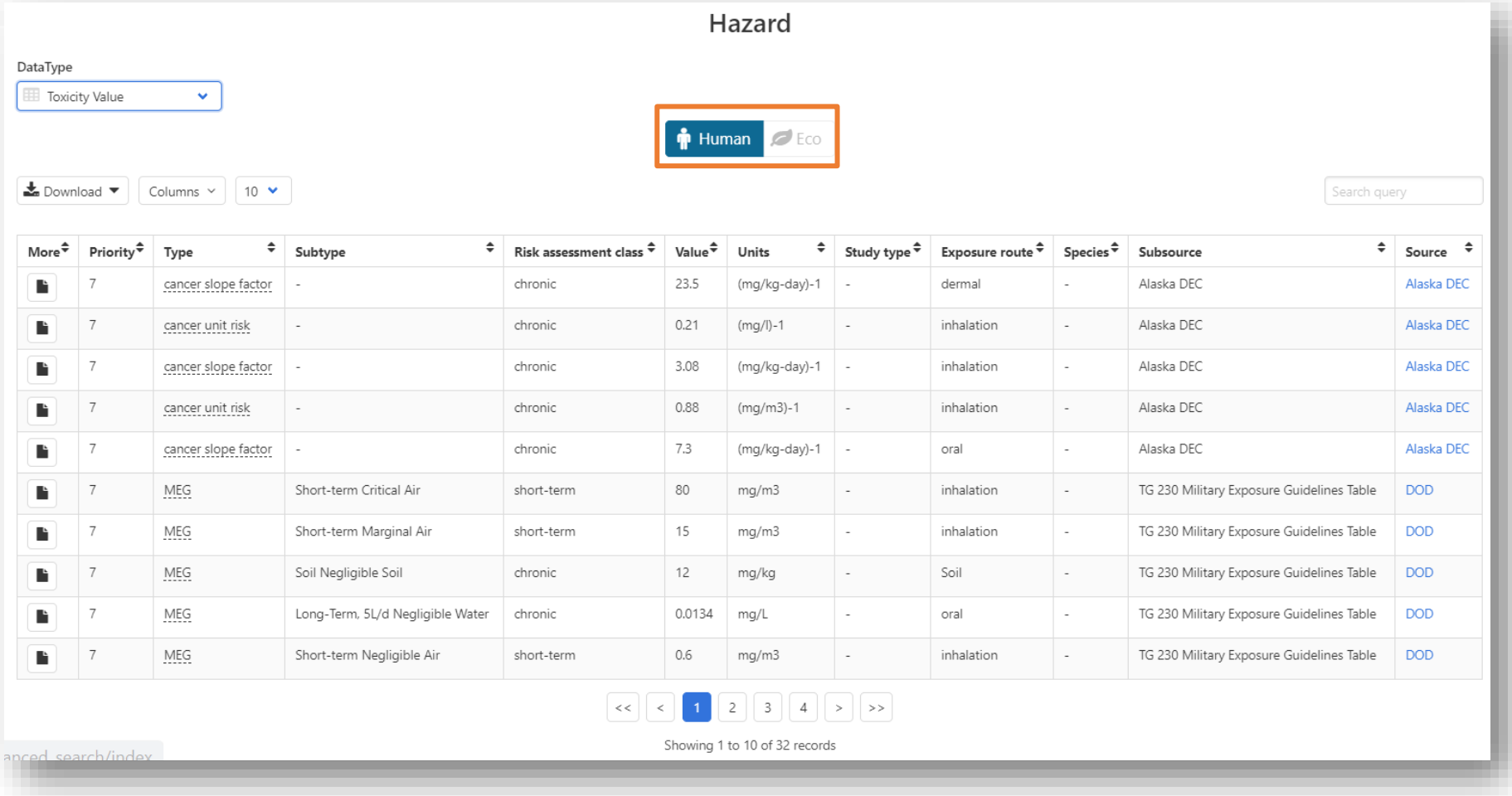
- Physchem and Fate & Transport experimental and predicted data
- Data can be downloaded as Excel, TSV and CSV files
- Predictions: multiple algorithms
  - EPI Suite: Estimation Program Interface
  - ACD/Labs (commercial)
  - TEST: **T**oxicity **E**stimation **S**oftware **T**ool
  - OPERA: **OPE**n structure–activity/ property **R**elationship **A**pp

Chemical Hazard Data

EPA United States Environmental Protection Agency

ToxVal Database

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



More	Priority	Type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	7	cancer slope factor	-	chronic	23.5	(mg/kg-day)-1	-	dermal	-	Alaska DEC	Alaska DEC
	7	cancer unit risk	-	chronic	0.21	(mg/l)-1	-	inhalation	-	Alaska DEC	Alaska DEC
	7	cancer slope factor	-	chronic	3.08	(mg/kg-day)-1	-	inhalation	-	Alaska DEC	Alaska DEC
	7	cancer unit risk	-	chronic	0.88	(mg/m3)-1	-	inhalation	-	Alaska DEC	Alaska DEC
	7	cancer slope factor	-	chronic	7.3	(mg/kg-day)-1	-	oral	-	Alaska DEC	Alaska DEC
	7	MEG	Short-term Critical Air	short-term	80	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Short-term Marginal Air	short-term	15	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Soil Negligible Soil	chronic	12	mg/kg	-	Soil	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Long-Term, 5L/d Negligible Water	chronic	0.0134	mg/L	-	oral	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Short-term Negligible Air	short-term	0.6	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD

Let’s talk about Export Formats

- Anywhere you see a table you can export
  - CSV is great for integration with other applications (plus read into Excel)
  - Excel file is generally the best for “viewing” as it can have multiple worksheets, color flagging of cells and offers all
- If you have cheminformatics tools SDF files are the best – view structures directly as concatenated “molfiles”
- Some examples… download file(s)

Advantages of Open Data

- By doing data exchange between databases you get to use other peoples work
- Linkages between systems (more later)
- Calling “APIs and Web Services” on the fly….

- SHOW EXAMPLES

Safety Data

GHS Data

Print Page

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

CID 2336

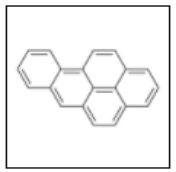
Benzo[a]pyrene

GHS Classification

Showing 6 of 6

Pictogram(s)	 Irritant, Health Hazard, Environmental Hazard
Signal	Danger
GHS Hazard Statements	H317: May cause an allergic skin reaction [Warning Sensitization, Skin] H340: May cause genetic defects [Danger Germ cell mutagenicity] 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 Codes	P201, P202, P261, P272, P273, P280, P281, P302+P352, P308+P313, P321, P333+P313, P363, P391, P405, and P501 (The corresponding statement to each P-code can be found at the GHS Classification page.)

Identifiers Support Searches in other systems



Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

Synonyms

Download 25 Search query

Synonym	Quality
Benzo(a)pyrene	**Valid**
Benzo[pqr]tetraphene	**Valid**
Benzo[a]pyrene	**Valid**
50-32-8 Active CAS-RN	**Valid**
BaP	**Valid**
Benzo[a]pyrene	*Good*
3,4-Benz[a]pyrene	*Good*
3,4-Benzopyrene	*Good*
3,4-Benzpyrene	*Good*
6,7-Benzopyrene	*Good*
BENZ(A)PYREN	*Good*
Benz(a)pyrene	*Good*
Benz[a]pyrene	*Good*

QUESTION 3

- How many CAS Registry Numbers does Atrazine have?

3 5 7 12

More About CASRNs

- CASRNs are very useful, and still limited
- Not every chemical has a STRUCTURE…substances vs structures
- “Chemical Abstracts Service” – numbers don’t exist until they abstracted and indexed
- Not every chemical on the dashboard necessarily has a CASRN – how would you find those that didn’t??? Hint: Search NOCAS_
- There are ~6000 chemicals without CASRN on the dashboard
- A chemical can also have many deleted CASRNs
- Where do you look for all identifiers for a chemical???

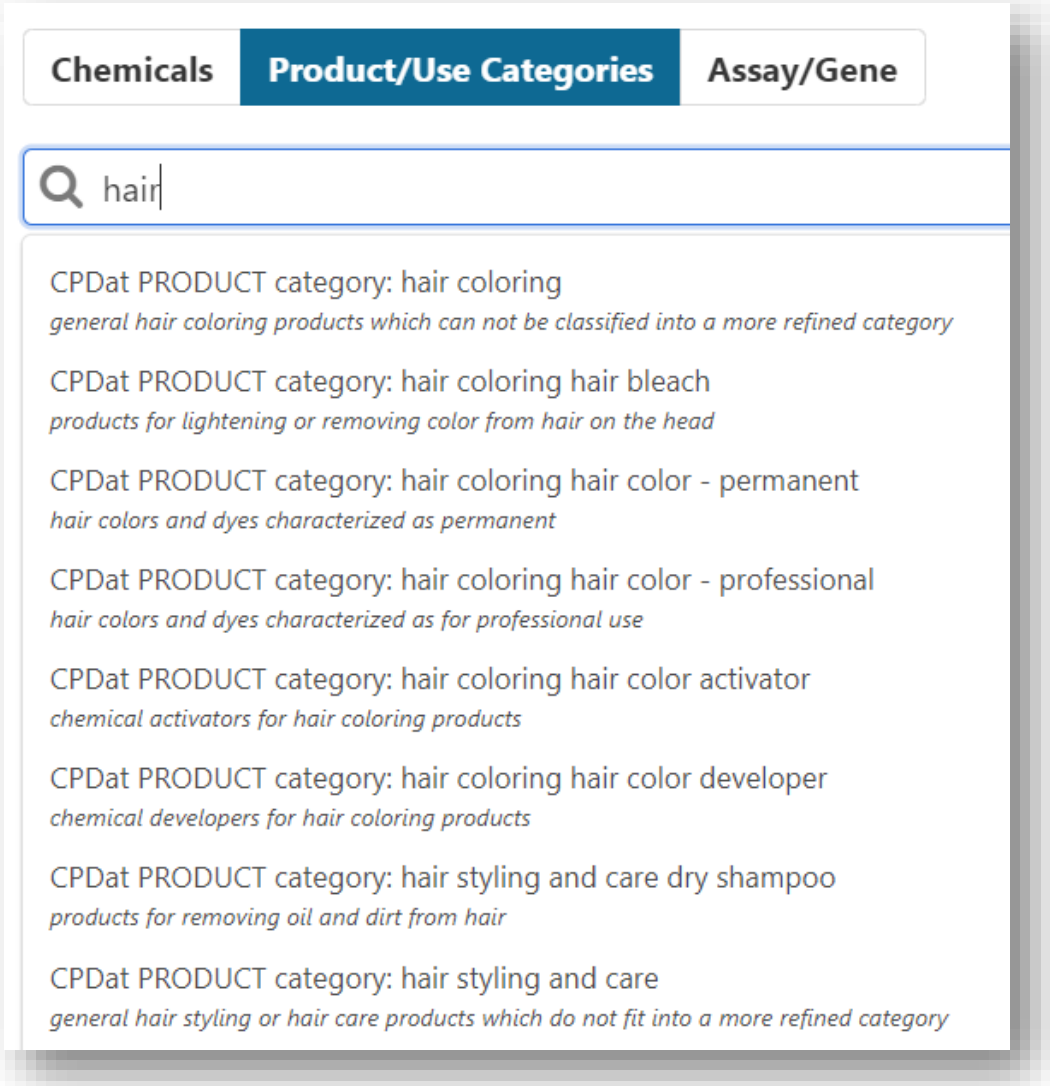
QUESTION 3

- How many CAS Registry Numbers does Atrazine have?

3 5 7 12

Products Searching

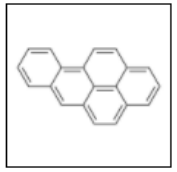
What chemicals are in hair care products?



Let’s Talk Exposure

- Types of Exposure Data on the Dashboard
  - Consumer product categories and uses
  - Products containing the chemical
  - Predicted exposure levels from modeling (more in next session)

Sources of Exposure to Chemicals



Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

Chemical Weight Fractions

Download | Columns | 10 | Search query

Product Name	Product Use Category	Minimum Weight Fraction	Maximum Weight Fraction	Data Type	Source
m-525-1-5x pah mixtures 0.5 mg/ml for method 525	Not Yet Categorized:			MSDS	SIRI
mm6125 surface conditioner	Not Yet Categorized:			Health Product Declaration	Health Product Declaration Collaborative
monolithic membrane 6125 (mm6125) / monolithic membrane	Not Yet Categorized:			Health Product Declaration	Health Product Declaration Collaborative
organic potablewatr pw 32_component h:reg semi-volatile 690	Not Yet Categorized:	0.00	1.00e-3	MSDS	SIRI
polynuclear aromatic hydrocarbon mixture_ ep84627	Not Yet Categorized:			MSDS	SIRI
prestone(r) power steering fluid	engine maintenance: auto fluids and additives			MSDS	CPCPdb
r-12 shield tite wet surface coating	Not Yet Categorized:	0.00	0.500	MSDS	SIRI
sea tar 1010_ 0028	Not Yet Categorized:			MSDS	SIRI
supelpreme-hc kit pah mix_ 48909	Not Yet Categorized:			MSDS	SIRI
supelpreme-hc pah mix 1ml_ 48905	Not Yet Categorized:			MSDS	SIRI

<< < 1 2 3 > >>

Showing 11 to 20 of 21 records

QUESTION 4

- How many consumer products are listed as containing chlorothalonil?

45 67 138 203

About exposure data

- Not every chemical has associated categories and uses
- MODELING data depends on QSAR predictions (structures)
- Data gathering continues unabated and has interesting sources

QUESTION 4

- How many consumer products are listed as containing chlorothalonil?

45 67 **138** 203

While we are discussing QSAR modeling

- What do you trust more? Experimental or predicted data?
- Do you trust individual models or consensus models
- What if there are no experimental data, how good are predictions?

- This will be covered a *little* in my next presentation and in way more detail in a later session

QUESTION 5

- What is the LogKow for fluconazole?

~0.2 ~0.3 ~0.5 ~3.4

How do we gather data for models

- We are continuously gathering data…where from?
  - How do we validate?
  - What can we check?

- Projects underway at present that may be of interest
  - Water Solubility dataset
  - Mass Spec Amenability
  - Eye and Skin Sensitization and Irritation
  - PFAS chemicals

QUESTION 5

- What is the LogKow for fluconazole?

~0.2 ~0.3 **~0.5** ~3.4

QUESTION 6

- What is the brand name for Ketoconazole? (There may be more than one)

Glyphosate Bisphenol A Cialis Nizoral

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

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

- What can WE do to help you navigate the internet?

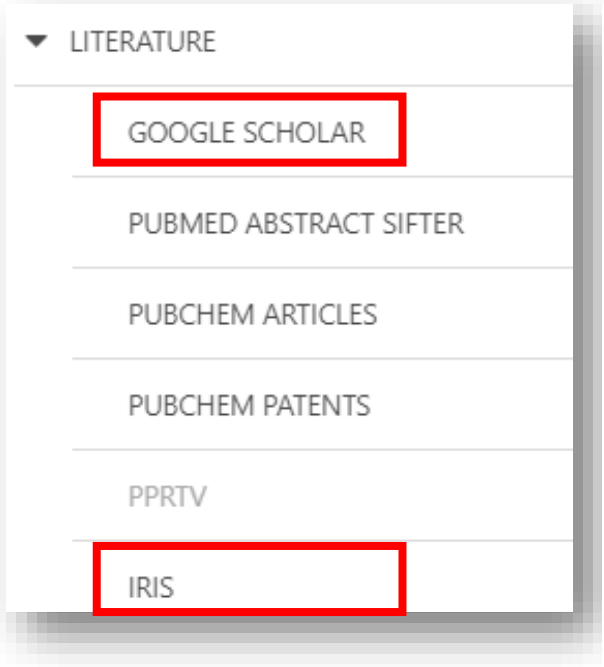
QUESTION 6

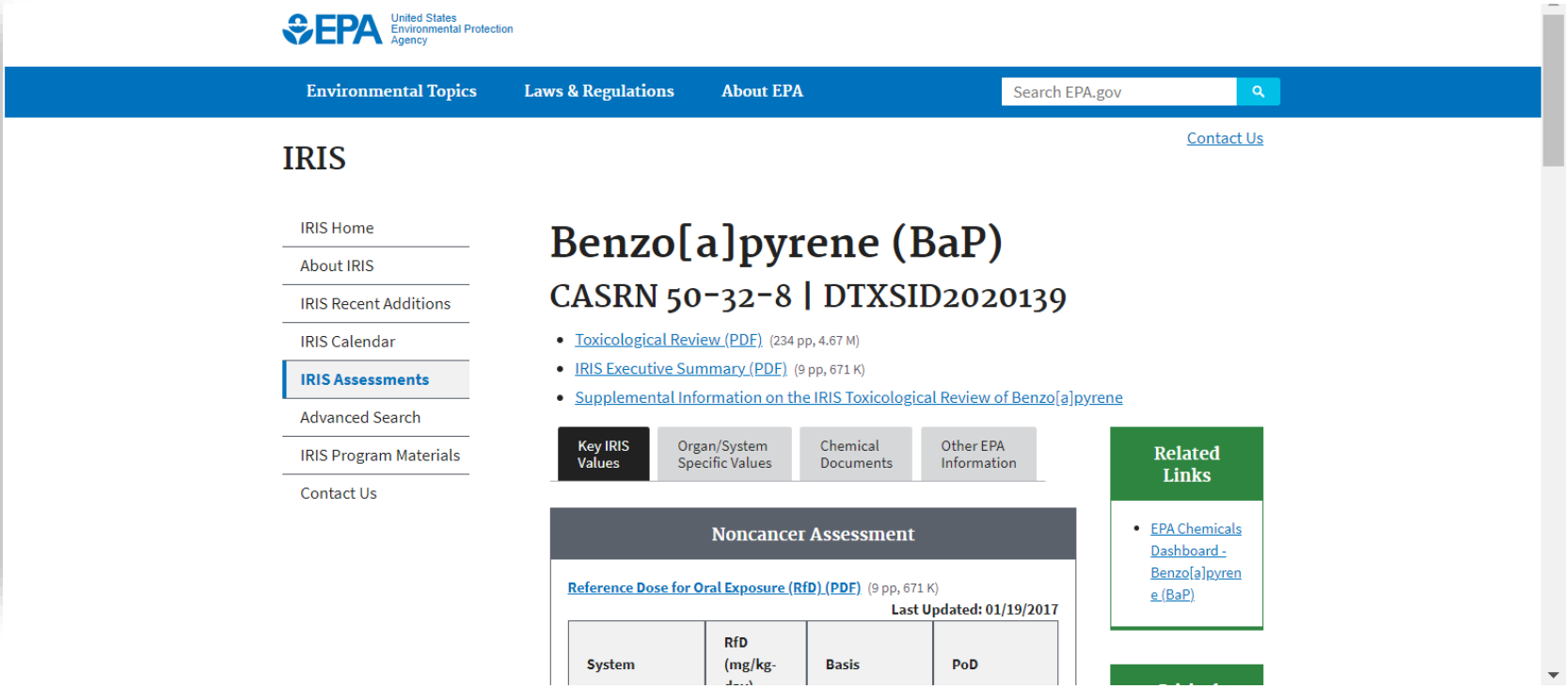
- What is the brand name for Ketoconazole? (There may be more than one)

Glyphosate Bisphenol A Cialis Nizoral

Identifiers are used in the app

- Identifiers are used to feed and link into “Literature”





Literature Searching

Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

Abstract Sifter

1) Select PubMed starting point query then 2) click on Retrieve.

Hazard

Retrieve Articles

- Select a Query Term
- Hazard
- Fate and Transport
- Metabolism/PK/PD
- Chemical Properties
- Exposure
- Mixtures
- Male Reproduction
- Androgen Disruption
- Female Reproduction
- GeneTox
- Cancer
- Clinical Trials
- Embryo and embryonic development
- Child (infant through adolescent)
- Dust and Exposure
- Food and Exposure
- Water and Exposure
- Algae
- Disaster / Emergency

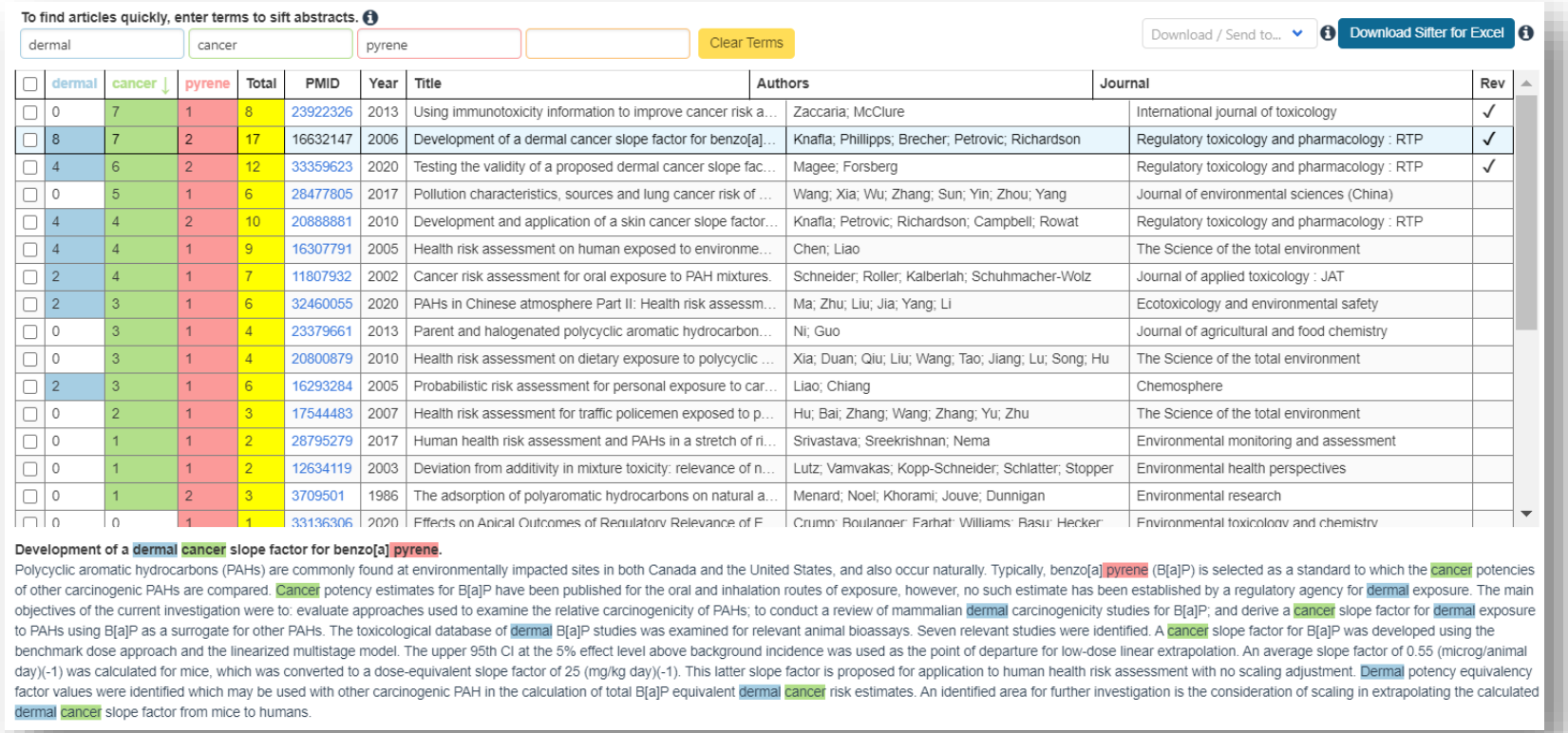
Optionally, edit the query before retrieving.

("50-32-8" OR "Benzo(a)pyrene") AND (NOAEL OR NOEL OR LOEL OR Rfd OR "reference dose" OR "reference concentration" OR "adverse effect level"[tiab] OR "cancer slope factor"[tiab])

- Real-time retrieval of data from PubMed ~30 million abstracts and growing)
- Choose from set of pre-defined queries
- Adjust and fine tune queries based on interests

Literature Searching

- "Sifting" of results using multiple terms
- Frequency counting terms
- Color highlighting of terms
- Download list to Excel
- Send list to PubMed for downloading ref. file
- Direct link via PubMed ID



To find articles quickly, enter terms to sift abstracts.

dermal | cancer | pyrene | Clear Terms | Download / Send to... | Download Sifter for Excel

	dermal	cancer ↓	pyrene	Total	PMID	Year	Title	Authors	Journal	Rev
	0	7	1	8	23922326	2013	Using immunotoxicity information to improve cancer risk a…	Zaccaria; McClure	International journal of toxicology	✓
	8	7	2	17	16632147	2006	Development of a dermal cancer slope factor for benzo[a]…	Knafla; Phillipps; Brecher; Petrovic; Richardson	Regulatory toxicology and pharmacology : RTP	✓
	4	6	2	12	33359623	2020	Testing the validity of a proposed dermal cancer slope fac…	Magee; Forsberg	Regulatory toxicology and pharmacology : RTP	✓
	0	5	1	6	28477805	2017	Pollution characteristics, sources and lung cancer risk of …	Wang; Xia; Wu; Zhang; Sun; Yin; Zhou; Yang	Journal of environmental sciences (China)	
	4	4	2	10	20888881	2010	Development and application of a skin cancer slope factor…	Knafla; Petrovic; Richardson; Campbell; Rowat	Regulatory toxicology and pharmacology : RTP	
	4	4	1	9	16307791	2005	Health risk assessment on human exposed to environme…	Chen; Liao	The Science of the total environment	
	2	4	1	7	11807932	2002	Cancer risk assessment for oral exposure to PAH mixtures.	Schneider; Roller; Kalberlah; Schuhmacher-Wolz	Journal of applied toxicology : JAT	
	2	3	1	6	32460055	2020	PAHs in Chinese atmosphere Part II: Health risk assessm…	Ma; Zhu; Liu; Jia; Yang; Li	Ecotoxicology and environmental safety	
	0	3	1	4	23379661	2013	Parent and halogenated polycyclic aromatic hydrocarbon…	Ni; Guo	Journal of agricultural and food chemistry	
	0	3	1	4	20800879	2010	Health risk assessment on dietary exposure to polycyclic …	Xia; Duan; Qiu; Liu; Wang; Tao; Jiang; Lu; Song; Hu	The Science of the total environment	
	2	3	1	6	16293284	2005	Probabilistic risk assessment for personal exposure to car…	Liao; Chiang	Chemosphere	
	0	2	1	3	17544483	2007	Health risk assessment for traffic policemen exposed to p…	Hu; Bai; Zhang; Wang; Zhang; Yu; Zhu	The Science of the total environment	
	0	1	1	2	28795279	2017	Human health risk assessment and PAHs in a stretch of ri…	Srivastava; Sreekrishnan; Nema	Environmental monitoring and assessment	
	0	1	1	2	12634119	2003	Deviation from additivity in mixture toxicity: relevance of n…	Lutz; Vamvakas; Kopp-Schneider; Schlatter; Stopper	Environmental health perspectives	
	0	1	2	3	3709501	1986	The adsorption of polyaromatic hydrocarbons on natural a…	Menard; Noel; Khorami; Jouve; Dunnigan	Environmental research	
	0	0	1	1	33136306	2020	Effects on Apical Outcomes of Regulatory Relevance of E…	Crump; Boulanger; Farhat; Williams; Basu; Hecker	Environmental toxicology and chemistry	

Development of a dermal cancer slope factor for benzo[a] pyrene.

Polycyclic aromatic hydrocarbons (PAHs) are commonly found at environmentally impacted sites in both Canada and the United States, and also occur naturally. Typically, benzo[a] pyrene (B[a]P) is selected as a standard to which the cancer potencies of other carcinogenic PAHs are compared. Cancer potency estimates for B[a]P have been published for the oral and inhalation routes of exposure, however, no such estimate has been established by a regulatory agency for dermal exposure. The main objectives of the current investigation were to: evaluate approaches used to examine the relative carcinogenicity of PAHs; to conduct a review of mammalian dermal carcinogenicity studies for B[a]P; and derive a cancer slope factor for dermal exposure to PAHs using B[a]P as a surrogate for other PAHs. The toxicological database of dermal B[a]P studies was examined for relevant animal bioassays. Seven relevant studies were identified. A cancer slope factor for B[a]P was developed using the benchmark dose approach and the linearized multistage model. The upper 95th CI at the 5% effect level above background incidence was used as the point of departure for low-dose linear extrapolation. An average slope factor of 0.55 (microg/animal day)(-1) was calculated for mice, which was converted to a dose-equivalent slope factor of 25 (mg/kg day)(-1). This latter slope factor is proposed for application to human health risk assessment with no scaling adjustment. Dermal potency equivalency factor values were identified which may be used with other carcinogenic PAH in the calculation of total B[a]P equivalent dermal cancer risk estimates. An identified area for further investigation is the consideration of scaling in extrapolating the calculated dermal cancer slope factor from mice to humans.

External Links – Also use Identifiers
Names, CASRN, PubChem IDs, InChIs…

Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

General

- EPA Substance Registry Service
- PubChem
- Chemspider
- CPCat
- DrugBank
- Wikipedia
- MSDS Lookup
- ChEMBL
- ToxPlanet
- ACS Reagent Chemicals
- Wolfram Alpha
- ECHA Infocard
- ChemAgora
- Consumer Product Information Database
- ChEBI
- NIST Chemistry Webbook
- WEBWISER
- PubChem Safety Sheet
- Consumer Product Information Database
- PubChem: Chemical Vendors

Toxicology

- ACToR
- DrugPortal
- CCRIS
- ChemView
- CTD
- eChemPortal
- Gene-Tox
- HSDB
- ACToR PDF Report
- CREST
- National Air Toxics Assessment
- ECOTOX
- ChemView
- Chemical Checker
- BindingDB
- CalEPA OEHHA
- NIOSH IDLH Values
- LactMed
- ECOTOX

Publications

- Toxline
- PPRTVWEB
- PubMed
- IRIS Assessments
- EPA HERO
- NIOSH Skin Notation Profiles
- NIOSH Pocket Guide
- RSC Publications
- BioCaddie DataMed
- Springer Materials
- Bielefeld Academic Search Engine
- CORE Literature Search
- Google Books (Text Search)
- Google Patents (Text search)
- Google Scholar (Text search)
- Google Patents (Structure search)
- Google Books (Structure Search)
- Google Scholar (Structure search)
- Federal Register

Analytical

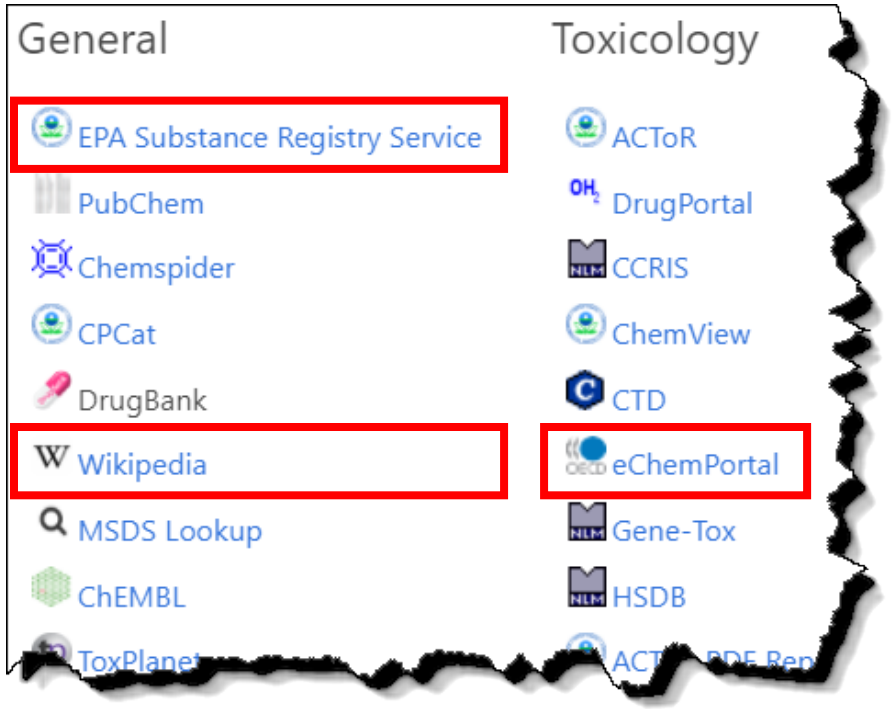
- RSC Analytical Abstracts
- Tox21 Analytical Data
- MONA: MassBank North America
- mzCloud
- NIST IR Spectrum
- NIST MS Spectrum
- MassBank
- NIST Antoine Constants
- IR Spectra on PubChem
- NIST Kovats Index values
- Protein DataBank
- National Environmental Methods Index

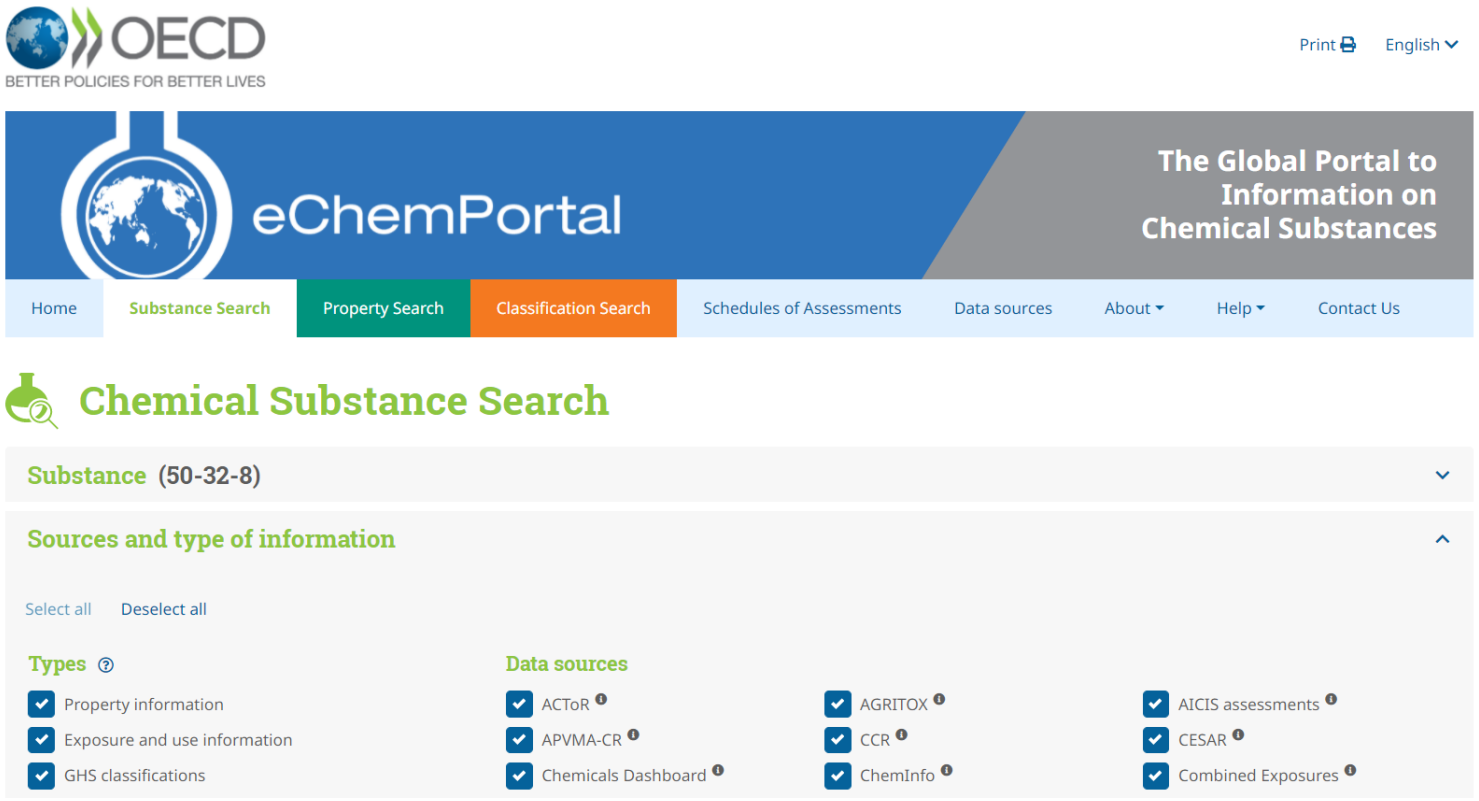
Prediction

- 2D NMR HSQC/HMBC Prediction
- Carbon-13 NMR Prediction
- Proton NMR Prediction
- ChemRTP Predictor
- LSERD

External Links

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





QUESTION 7

- Searches can be done for CLASSES of Chemicals too
- How many **explicit** chemicals are included in the class of chemicals “polychlorinated biphenyls”

302 211 209 104

Some example classes where we map data

- Classes of chemicals – what do we have?
  - PAHs, PCBs, PBDEs etc…
  - Strings and Substrings
  - InChI “first block” for skeleton

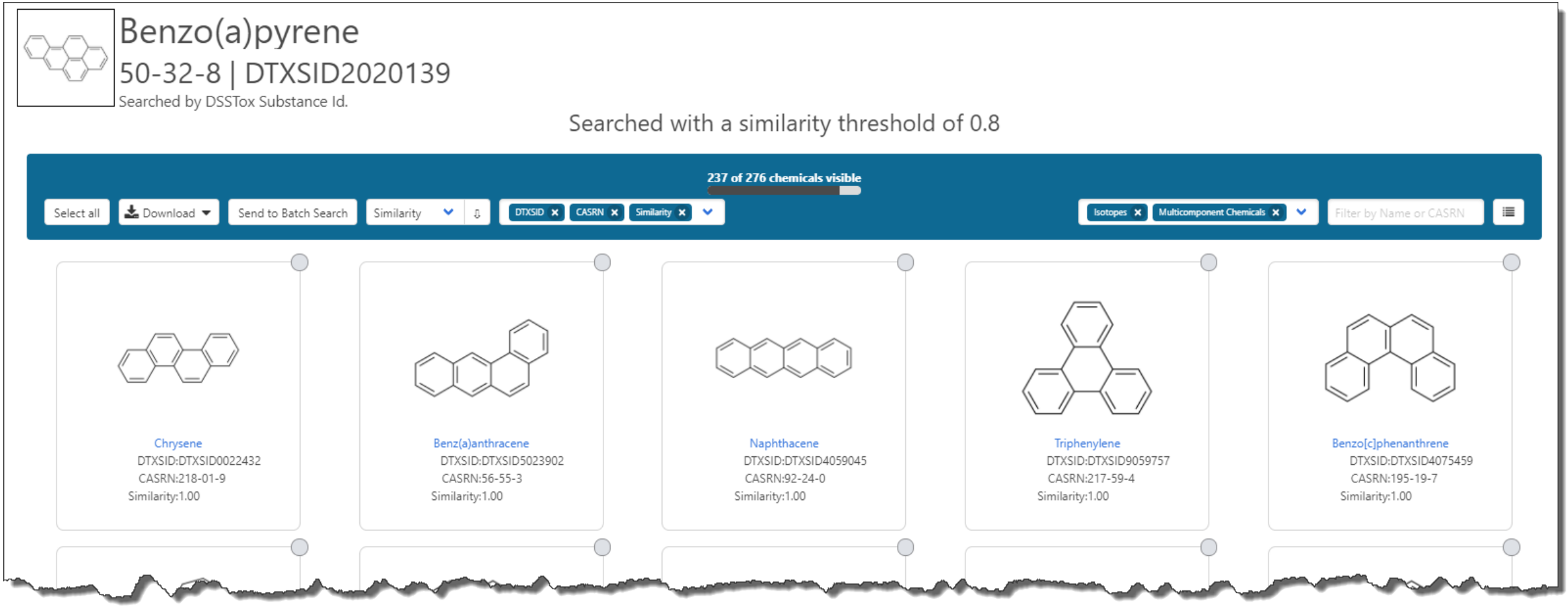
- Other “classes”
  - Similar chemicals
  - Related Substances
  - Lists of chemicals

QUESTION 7

- Searches can be done for CLASSES of Chemicals too
- How many **explicit** chemicals are included in the class of chemicals “polychlorinated biphenyls”

302 211 209 104

Similarity Searching

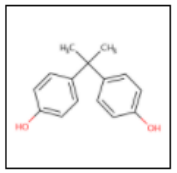


QUESTION 8

- How many chemicals are “similar” in structure to ketoconazole with a threshold of >0.8 similarity (based on Tanimoto score).

80 800 63 8

Related Substances



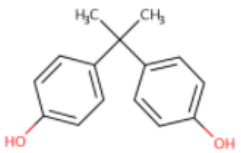
Bisphenol A

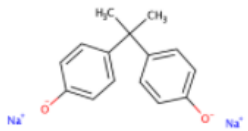
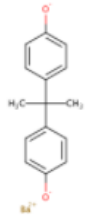
80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

9 chemicals

Select all | Download | Send to Batch Search | Relationship | DTXSID × | CASRN × | TOXCAST × | Hide chemicals that are: | Filter by Name or CASRN

Searched Chemical	Polymer	Polymer	Polymer	Predecessor: Component
	2 related chemical structures with this substance	2 related chemical structures with this substance	3 related chemical structures with this substance	4 related chemical structures with this substance
Bisphenol A DTXSID:DTXSID7020182 CASRN:80-05-7 TOXCAST:217/1152	Formaldehyde, polymer with bisphenol A DTXSID:DTXSID3049627 CASRN:25085-75-0 TOXCAST:-	Bisphenol A/ Epichlorohydrin resin DTXSID:DTXSID0050479 CASRN:25068-38-6 TOXCAST:-	Bisphenol A-epichlorohydrin-polyformal... DTXSID:DTXSID9050480 CASRN:28906-96-9 TOXCAST:-	Fatty acids, C18-unsatd., dimers, polyme... DTXSID:DTXSID40105886 CASRN:106906-26-7 TOXCAST:-

Predecessor: Component	Predecessor: Component	Salt Form	Salt Form
3 related chemical structures with this substance	2 related chemical structures with this substance		
...rohydrin-2,2',6,6'-tet...	Phenoxy Resin	Disodium 4,4'-isopropylidenediphenolate	Barium(2+) 4,4'-isopropylidenebispheno...

...anced_search/index

QUESTION 8

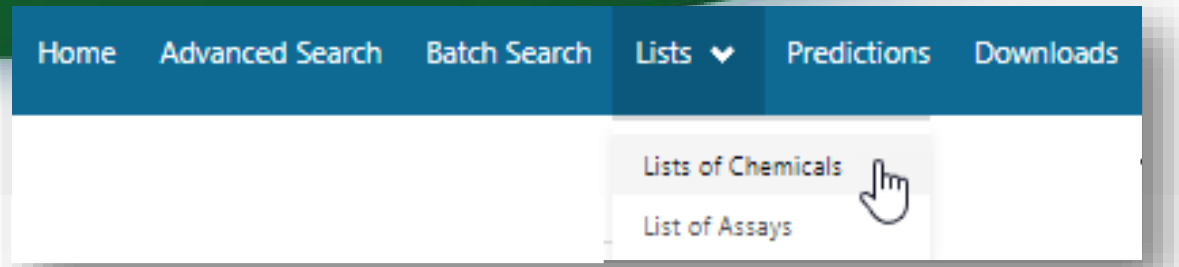
- How many chemicals are “similar” in structure to ketoconazole with a threshold of >0.8 similarity (based on Tanimoto score).

80 800 63 8

EPA
United States
Environmental Protection
Agency

Chemical Lists and Categories

Example: AEGLs list



AEGLS: Acute Exposure Guideline Levels

Search AEGLVALUES Chemicals

Identifier substring search

List Details

Description: Acute Exposure Guideline Level (AEGLs) values are intended to protect most individuals in the general population, including those that might be particularly susceptible to the harmful effects of the chemicals. Acute exposure guideline levels (AEGLs) describe the human health effects from once-in-a-lifetime, or rare, exposure to airborne chemicals. Used by emergency responders when dealing with chemical spills or other catastrophic exposures, AEGLs are set through a collaborative effort of the public and private sectors worldwide.
Number of Chemicals: 174

Select all | Download | Send to Batch Search | Name | CASRN | DTXSID | 174 chemicals | Hide chemicals that are: | Filter by Name or CASRN

(2-Chlorobenzylidene)propanedinitrile
CASRN:2698-41-1
DTXSID:DTXSID9020297

(E)-1,2-Dichloroethylene
CASRN:156-60-5
DTXSID:DTXSID7024031

(E)-Crotonaldehyde
CASRN:123-73-9
DTXSID:DTXSID6020351

(Z)-1,2-Dichloroethylene
CASRN:156-59-2
DTXSID:DTXSID2024030

PFAS lists of Chemicals

EPA United States Environmental Protection Agency

Select List

Download | Columns | PFAS | Copy Filtered Lists URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
EPAPFAS75S1	PFAS|EPA: List of 75 Test Samples (Set 1)	2018-06-29	74	PFAS list corresponds to 75 samples (Set 1) submitted for initial testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFAS75S2	PFAS|EPA: List of 75 Test Samples (Set 2)	2019-02-21	75	PFAS list corresponds to a second set of 75 samples (Set 2) submitted for testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFASCAT	PFAS|EPA Structure-based Categories	2018-06-29	64	List of registered DSSTox "category substances" representing PFAS categories created using ChemAxon's Markush structure-based query representations.
EPAPFASINSOL	PFAS|EPA: Chemical Inventory Insoluble in DMSO	2018-06-29	43	PFAS chemicals included in EPA's expanded ToxCast chemical inventory found to be insoluble in DMSO above 5mM.
EPAPFASINV	PFAS|EPA: ToxCast Chemical Inventory	2018-06-29	430	PFAS chemicals included in EPA's expanded ToxCast chemical inventory and available for testing.
EPAPFASRL	PFAS|EPA: Cross-Agency Research List	2017-11-16	199	EPAPFASRL is a manually curated listing of mainly straight-chain and branched PFAS (Per- & Poly-fluorinated alkyl substances) compiled from various internal, literature and public sources by EPA researchers and program office representatives.
PFASKEMI	PFAS: List from the Swedish Chemicals Agency (KEMI) Report	2017-02-09	2416	Perfluorinated substances from a Swedish Chemicals Agency (KEMI) Report on the occurrence and use of highly fluorinated substances.
PFASMASTER	PFAS Master List of PFAS Substances	2018-07-26	5061	PFASMASTER is a consolidated list of PFAS substances spanning and bounded by the below lists of current interest to researchers and regulators worldwide.
PFASOECD	PFAS: Listed in OECD Global Database	2018-05-16	4729	OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances, (PFASs) listing more than 4700 new PFAS
PFASTRIER	PFAS Community-Compiled List (Trier et al., 2015)	2017-07-16	597	PFASTRIER community-compiled public listing of PFAS (Trier et al, 2015)

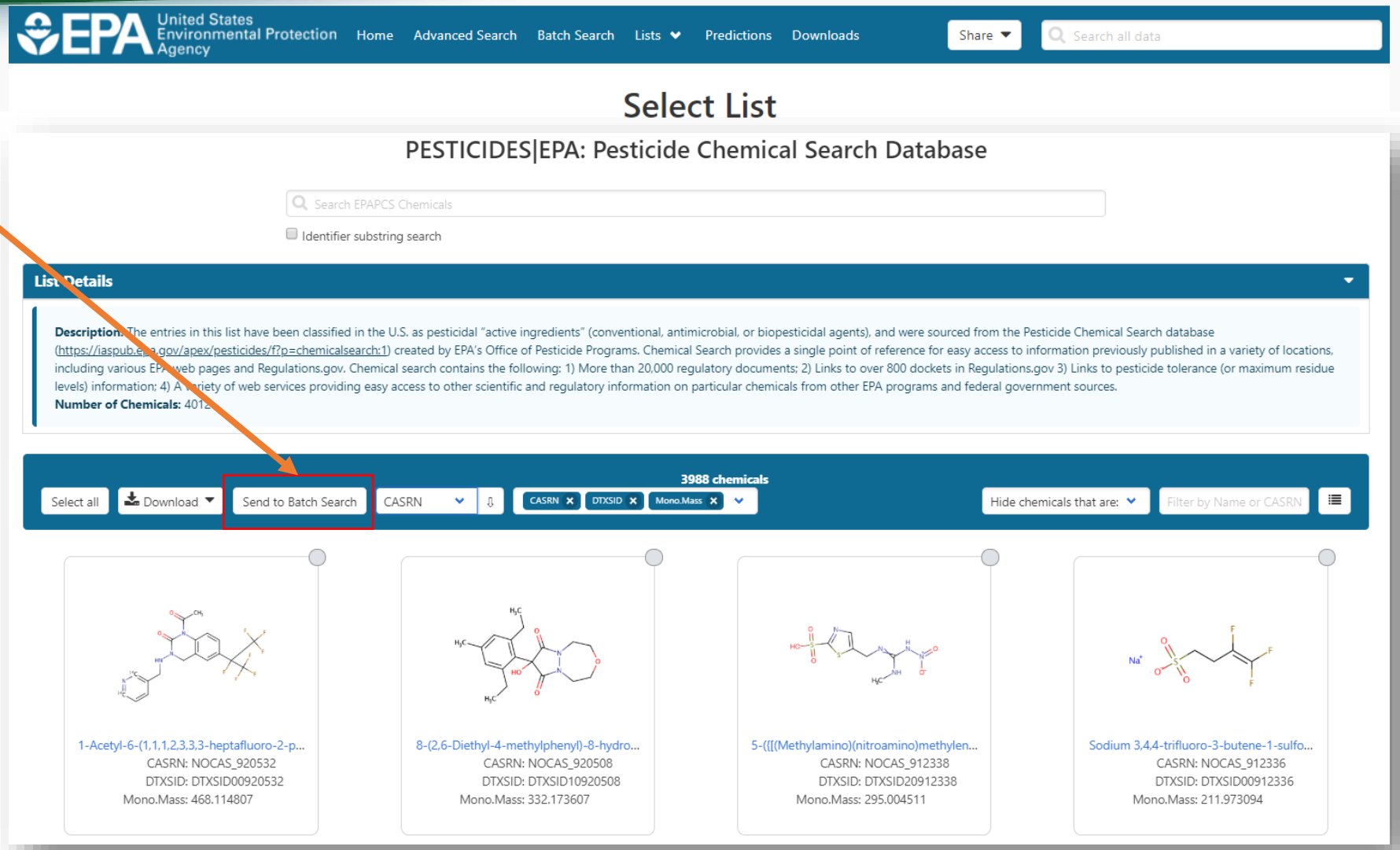
QUESTION 9

- How many chemicals are in the EPA Toxics Release Inventory based on the associated list available on the dashboard?

666 677 766 777

Curated List of Pesticides

- Find list of interest
- Select list and send to batch



Batch Searching

- Singleton searches are great but…
- …we generally want data on LOTS of chemicals!

- Typical questions
  - What are the structures for a set of chemical names? Set of CASRNs?
  - Can I get chemical lists in Excel files? As a list of SMILES strings? Can I get an SDF file?
  - Can I include predicted properties in the download file? OPERA? TEST?
  - Are “these chemicals” screened in Toxcast?
  - I’m a mass spectrometrist and need masses and formulae for a list of chemicals

QUESTION 9

- How many chemicals are in the EPA Toxics Release Inventory based on the associated list available on the dashboard?

666 677 766 777

Access data *en masse* for thousands of chemicals….

EPA United States Environmental Protection Agency

EPA United States Environmental Protection Agency | Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Share | Search all data

Batch Search

Step 1 — Step 2 — Step 3 — Step 4 — Step 5

Step Three: Select Download Data or Display Chemicals

Please enter one identifier per line

Select Input Type(s)

- [x] Identifiers
  - [x] Chemical Name
  - [x] CASRN
  - [ ] InChIKey
  - [ ] DSSTox Substance ID
- [ ] DSSTox Compound ID
- [ ] InChIKey Skeleton
- [ ] MS-Ready Formula(e)
- [ ] Exact Formula(e)
- [ ] Monoisotopic Mass

Enter Identifiers to Search (One per line. Searches should be limited to <5000 identifiers.)

```
Sodium azide
O,O-Dimethyl S-[(4-oxo-1,2,3-benzotriazin-3(4H)-yl)methyl] phosphorodithioate
(Trichloromethyl)benzene
100-44-7
(Chloromethyl)benzene
98-07-7
111-44-4
1-Chloro-2-(2-chloroethoxy)ethane
Chloro(chloromethoxy)methane
2-(Butan-2-yl)-4,6-dinitrophenol
```

Display All Chemicals | Download Chemical Data

Select Output Format and Content

United States Environmental Protection Agency

Step Four: Select Data Output Format and Choose Data Fields to Download

Select Output Format:

Excel

Download

Customize Results

- [ ] Select All
- [ ] Select All in Lists

Chemical Identifiers

- [x] DTXSID
- [x] Chemical Name
- [ ] DTXCID
- [ ] CAS-RN
- [ ] InChIKey
- [ ] IUPAC Name

Structures

- [ ] Mol File
- [ ] SMILES
- [ ] InChI String
- [ ] MS-Ready SMILES
- [ ] QSAR-Ready SMILES

Intrinsic And Predicted Properties

- [ ] Molecular Formula
- [ ] Average Mass
- [ ] Monoisotopic Mass
- [ ] TEST Model Predictions
- [ ] OPERA Model Predictions

Metadata

- [ ] Curation Level Details
- [ ] NHANES/Predicted Exposure
- [ ] Data Sources
- [ ] Include ToxVal Data Availability
- [ ] Assay Hit Count
- [ ] Number of PubMed Articles
- [ ] PubChem Data Sources
- [ ] CPDat Product Occurrence Count
- [ ] IRIS
- [ ] PPRTV
- [ ] Wikipedia Article
- [ ] QC Notes
- [ ] Include links to ACToR reports - SLOW! (BETA)

Presence in Lists:

- [ ] 40CFR116.4 Designation of Hazardous Substances (Above Ground Storage Tanks)
- [ ] 40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities
- [ ] AEGLS: Acute Exposure Guideline Levels
- [ ] ANDROGEN: Androgen Receptor Chemicals
- [ ] ARTICLE; Bench-Mark Dose Human Health Assessment List (Wignall et al., 2014)
- [ ] ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP)
- [ ] ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (COMPARA)
- [ ] ATSDR Toxicological Profiles
- [ ] ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances
- [ ] ATSDR: Toxic Substances Portal Chemical List
- [ ] California Office of Environmental Health Hazard Assessment
- [ ] Canadian Domestic Substances List 2019
- [ ] CATEGORY: Amino acids
- [ ] CATEGORY: Color Index dyes
- [ ] CATEGORY: Flame Retardants

Batch Search CASRNs

EPA United States Environmental Protection Agency

Batch Search

Step 1 — Step 2 — Step 3 — Step 4 — Step 5

Step Four: Select Data Output Format and Choose Data Fields to Download

Please enter one identifier per line

Chemical Identifiers
- [x] DTXSID
- [x] Chemical Name
- [ ] DTXCID
- [x] CAS-RN
- [x] InChIKey
- [x] IUPAC Name

Structures
- [ ] Mol File
- [ ] SMILES
- [ ] InChI String
- [x] MS-Ready SMILES
- [ ] QSAR-Ready SMILES

Intrinsic And Predicted Properties
- [x] Molecular Formula
- [ ] Average Mass
- [ ] Monoisotopic Mass
- [ ] TEST Model Predictions
- [x] OPERA Model Predictions

Metadata
- [ ] Curation Level Details
- [ ] NHANES/Predicted Exposure
- [ ] Data Sources
- [x] Include ToxVal Data Availability
- [x] Assay Hit Count
- [x] Number of PubMed Articles

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	INPUT	FOUND_BY	DTXSID	PREFERRE	CASRN	INCHIKEY	IUPAC_NAI	MOLECULA	TOXVAL_D	TOXCAST_	TOXCAST_	MS_READY	NUMBER_C	IRIS_LINK	PPRTV_LIN	ATMOSPHE	BIOCONCE
2	534-52-1	CAS-RN	DTXSID10	2-Methyl-4,6	534-52-1	ZXVONLUN	2-Methyl-4,6	C7H6N2O5	Y	27.56	261/947	CC1=C(O)C	179	-	Y	1.571E-12	2.88959
3	115-21-9	CAS-RN	DTXSID70	Ethyl silicon	115-21-9	ZOYFEXPF	Trichloro(eth	C2H5Cl3Si	Y	-	-	-	-	-	-	8.289E-12	13.9658
4	111-44-4	CAS-RN	DTXSID90	Bis(2-chloro	111-44-4	ZNSMNVML	1-Chloro-2-(	C4H8Cl2O	Y	1.12	10/891	ClCCOCCCl	12	Y	-	2.647E-12	9.99608
5	2763-96-4	CAS-RN	DTXSID50	Muscimol	2763-96-4	ZJQHPWUV	5-(Aminome	C4H6N2O2	Y	-	-	NCC1=CC((	4308	-	-	1.179E-10	5.05695
6	1464-53-5	CAS-RN	DTXSID00	2,2'-Bioxiran	1464-53-5	ZFIVKAOQE	2,2'-Bioxiran	C4H6O2	Y	-	-	C1OC1C1C	363	-	-	4.383E-12	1.27107
7	22224-92-6	CAS-RN	DTXSID30	Fenamiphos	22224-92-	ZCJPOPBZI	Ethyl 3-meth	C13H22NO3	Y	10.8	105/972	CCOP(=O)(	58	Y	-	1.66E-11	2.3394
8	359-06-8	CAS-RN	DTXSID40	Fluoroacetyl	359-06-8	ZBHDTYQJ	Fluoroacetyl	C2H2ClFO	Y	-	-	FCC(Cl)=O	-	-	-	3.513E-13	4.49379
9	5344-82-1	CAS-RN	DTXSID40	1-(o-Chlorop	5344-82-1	YZUKKTCD	N-(2-Chloro	C7H7ClN2S	Y	-	-	NC(=S)NC1	-	-	-	2.482E-11	9.95206
10	7446-18-6	CAS-RN	DTXSID10	Thallium (I) s	7446-18-6	YTQVHRVI	Dithallium(1+	O4STl2	Y	-	-	-	34	Y	Y	-	-
11	62207-76-5	CAS-RN	DTXSID40	Bis(3-fluoros	62207-76-	YRZXYIHD	-	C16H14CoF	Y	-	-	-	-	-	-	-	-
12	66-81-9	CAS-RN	DTXSID60	Cycloheximi	66-81-9	YPHMISFO	4-{(2R)-2-[(1	C15H23NO4	Y	32.27	294/911	CC1CC(C)C	18709	-	-	1.756E-11	2.81761
13	106-96-7	CAS-RN	DTXSID30	Propargyl br	106-96-7	YORCIIVHU	3-Bromopro	C3H3Br	Y	-	-	BrCC#C	-	-	-	1.069E-11	10.4968
14	315-18-4	CAS-RN	DTXSID70	Mexacarbat	315-18-4	YNEVBPNZ	4-(Dimethyla	C12H18N2O	Y	5.11	12/235	CNC(=O)OC	27	-	-	1.447E-11	26.2914
15	110-00-9	CAS-RN	DTXSID60	Furan	110-00-9	YLQBMQCU	Furan	C4H4O	Y	0.0	0/235	O1C=CC=C	919	Y	-	4.019E-11	5.01648
16	3037-72-7	CAS-RN	DTXSID20	4-(diethoxyn	3037-72-7	YHFFINXFN	4-[Diethoxy(	C9H23NO2S	Y	-	-	-	-	-	-	5.95E-12	6.71292
17	75-44-5	CAS-RN	DTXSID00	Phosgene	75-44-5	YGYAWVDI	Carbonyl dic	CCl2O	Y	-	-	ClC(Cl)=O	489	Y	-	9.994E-16	13.0711
18	2032-65-7	CAS-RN	DTXSID30	Methiocarb	2032-65-7	YFBPRJGD	3,5-Dimethy	C11H15NO2	Y	18.14	88/485	CNC(=O)OC	65	-	-	1.446E-11	34.1692
19	2778-04-3	CAS-RN	DTXSID20	Endothion	2778-04-3	YCAGGFXS	S-[(5-Metho	C9H13O6PS	Y	-	-	COC1=COC	-	-	-	4.355E-11	1.18341
20	12108-13-3	CAS-RN	DTXSID90	(Methylcyclc	12108-13-	YASXMYPV	Tricarbonyl(	C9H7MnO3	Y	2.56	11/430	-	68	-	-	-	-
21	7803-51-2	CAS-RN	DTXSID20	Phosphine	7803-51-2	XYFCBTPG	Phosphane	H3P	Y	-	-	-	928	Y	-	-	-
22	107-18-6	CAS-RN	DTXSID80	Allyl alcohol	107-18-6	XXROGKLT	Prop-2-en-1	C3H6O	Y	3.99	17/426	OCC=C	627	Y	Y	2.592E-11	4.03901
23	108-05-4	CAS-RN	DTXSID30	Vinyl acetate	108-05-4	XTXRWKRV	Ethenyl acet	C4H6O2	Y	1.7	4/235	CC(=O)OC=	206	Y	-	2.5E-11	5.52157
24	19624-22-7	CAS-RN	DTXSID10	Pentaborane	19624-22-	XPIBKKWN	-	B5H9	Y	-	-	-	-	-	-	-	-
25	75-74-1	CAS-RN	DTXSID00	Tetramethyl	75-74-1	XOOGZRUI	Tetramethyl	C4H12Pb	Y	-	-	C[Pb](C)(C)	24	-	-	-	-

Worksheet1

Send to batch and select….

Intrinsic And Predicted Properties

- ☐ Molecular Formula
- ☐ Average Mass
- ☐ Monoisotopic Mass
- ☐ TEST Model Predictions
- ☐ OPERA Model Predictions

Metadata

- ☐ Curation Level Details
- ☐ NHANES/Predicted Exposure
- ☐ Data Sources
- ☐ Include ToxVal Data Availability
- ☐ Assay Hit Count
- ☐ Number of PubMed Articles
- ☐ PubChem Data Sources
- ☐ CPDat Product Occurrence Count
- ☐ IRIS
- ☐ PPRTV

- A few seconds to assemble
  - ToxCast data - #actives/#assays and % active
  - # articles in PubMed
  - Links to IRIS or PPRTV reports
  - TEST or OPERA predictions
  - Exposure data: predictions and CPDat

DTXSID	PREFERRED_NAME	EXPOCAST_MEI	EXPOCAST	NHANES	TOXVAL_DATA	TOXCAST_%_ACT	TOXCAST_	#PUBMED_	PUBCHEM_	CPDAT_COUI	IRIS_LINK	PPRTV_LINK
DTXSID2021105	Pentachloronitrobenzene	1.14e-07	Y	Y	Y	11.8	99/839	69	96	164	Y	-
DTXSID4022527	Propylparaben	1.4e-05	Y	Y	Y	13.77	99/719	201	121	1476	-	-
DTXSID4024064	Dinex	8.29e-08	Y	-	Y	42.13	99/235	-	35	5	Y	-
DTXSID0032493	Triadimenol	1.73e-08	Y	-	Y	10.54	98/930	163	74	83	-	-
DTXSID4032667	Esfenvalerate	1.7e-06	Y	-	Y	11.45	98/856	483	45	198	-	-
DTXSID6020561	Endrin	1.29e-07	Y	-	Y	14.02	98/699	284	16	98	Y	Y
DTXSID6025355	Glutaraldehyde	2.03e-05	Y	-	Y	14.35	98/683	6515	139	1144	-	-
DTXSID8032417	Isofenphos	1.87e-08	Y	-	Y	16.28	98/602	30	42	60	-	-
DTXSID6032352	Chlorpyrifos-methyl	1.07e-07	Y	Y	Y	11.27	97/861	72	50	116	-	-
DTXSID8020620	Fenthion	8.99e-08	Y	Y	Y	11.56	97/839	354	100	99	-	-
DTXSID2020189	FD&C Blue No. 1	0.000178	Y	-	Y	13.72	97/707	174	49	672	-	-
DTXSID7044843	Erythrosin B	6.3e-07	Y	-	-	24.25	97/400	14843	51	7	-	-
DTXSID5041778	Chloropropylate	1.05e-07	Y	-	Y	40.93	97/237	-	36	12	-	-
DTXSID5023900	Benomyl	1.11e-07	Y	-	Y	11.23	96/855	476	91	105	Y	-
DTXSID9020247	Carbaryl	5.61e-08	Y	Y	Y	11.51	96/834	1135	117	245	Y	-
DTXSID8024109	Flutolanil	1.63e-08	Y	-	Y	11.4	95/833	6	59	80	-	-
DTXSID1023998	Cypermethrin	1.62e-06	Y	Y	Y	10.78	94/872	1148	148	246	-	-
DTXSID2024242	Paclobutrazol	9.19e-08	Y	-	Y	11.11	94/846	139	-	40	Y	-
DTXSID1020807	2-Mercaptobenzothiazole	4.7e-05	Y	-	Y	12.82	94/733	111	181	86	-	Y

QUESTION 10

- How many chemicals are in the Tox21 Screening library, based on the **associated dashboard list**

4987 6587 8947 7632

- *What is the Tox21 Screening Library?? How is it valuable?* ***NEXT TIME***

In the next Session we discuss NAMS

EPA United States Environmental Protection Agency

Wednesday, Aug. 18, 2021 | 1:00–3:00 p.m. (Central US Time)

"NEW APPROACH METHODS"—WHAT IS THAT?

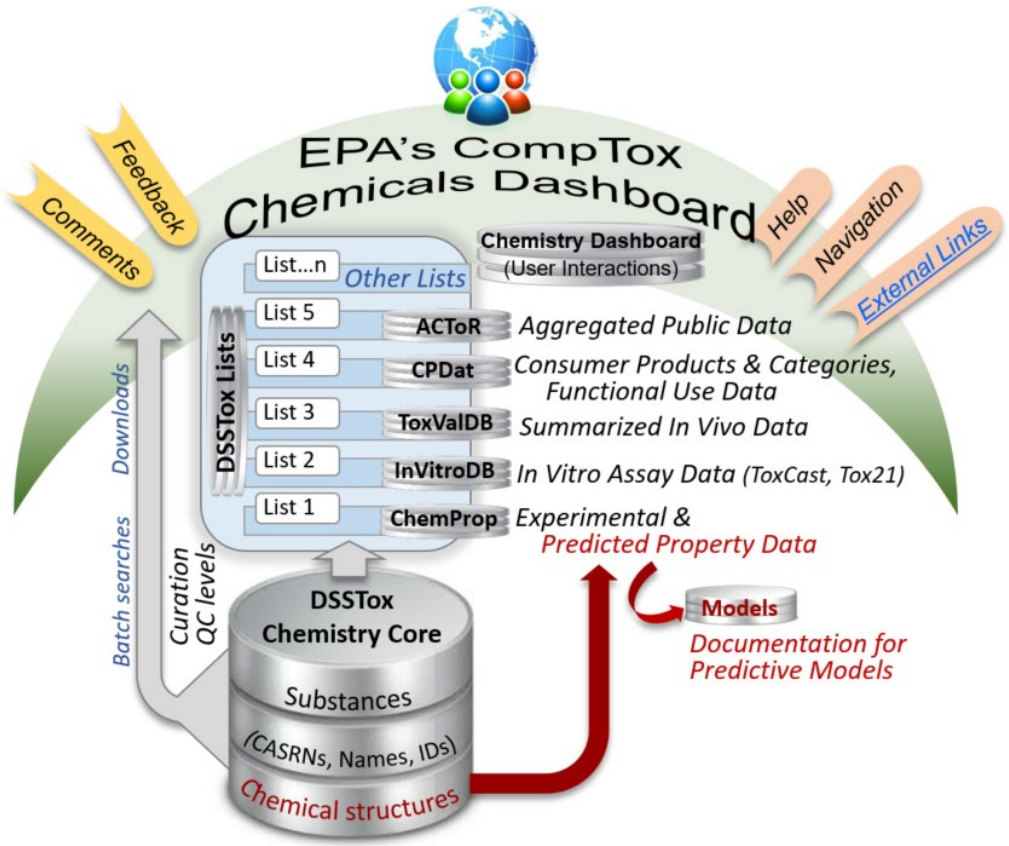
- An introduction to New Approach Methods
  - What's a NAM?
    - *In silico* – QSAR and read-across
    - *In vitro* assays
    - *In vitro* toxicokinetics
    - Computer modeling
- Short introduction to QSAR model data in the Dashboard
  - TEST predictions
  - OPERA predictions
  - Calculation reports
  - Realtime prediction
- An introduction to ToxCast and Tox21
- An overview of assay endpoints and biology

QUESTION 10

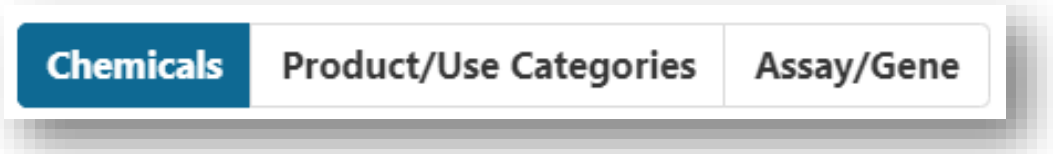
- How many chemicals are in the Tox21 Screening library, based on the associated dashboard list

4987 6587 8947 7632

Summary and Conclusion



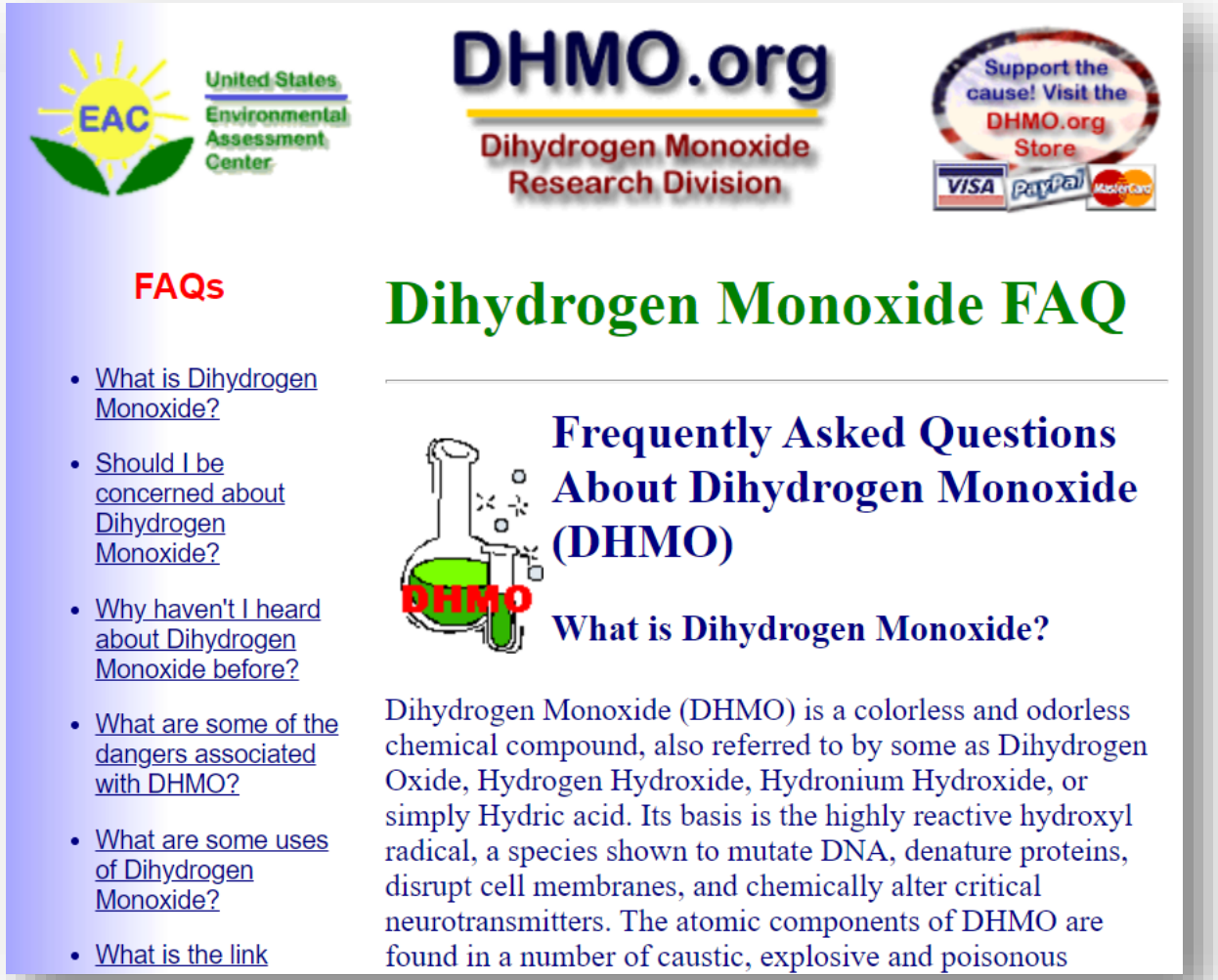
- CompTox Chemicals Dashboard - a central hub for environmental data
  - ~875k chemical substances
  - Integrating property data, hazard data, exposure data, *in vitro* bioactivity data
  - Interrogation of bioactivity data -
  - Multiple types of searches



  - Batch search for thousands of chemicals
  - Real-time property and toxicity predictions
  - Downloadable files – CSV, TSV and Excel

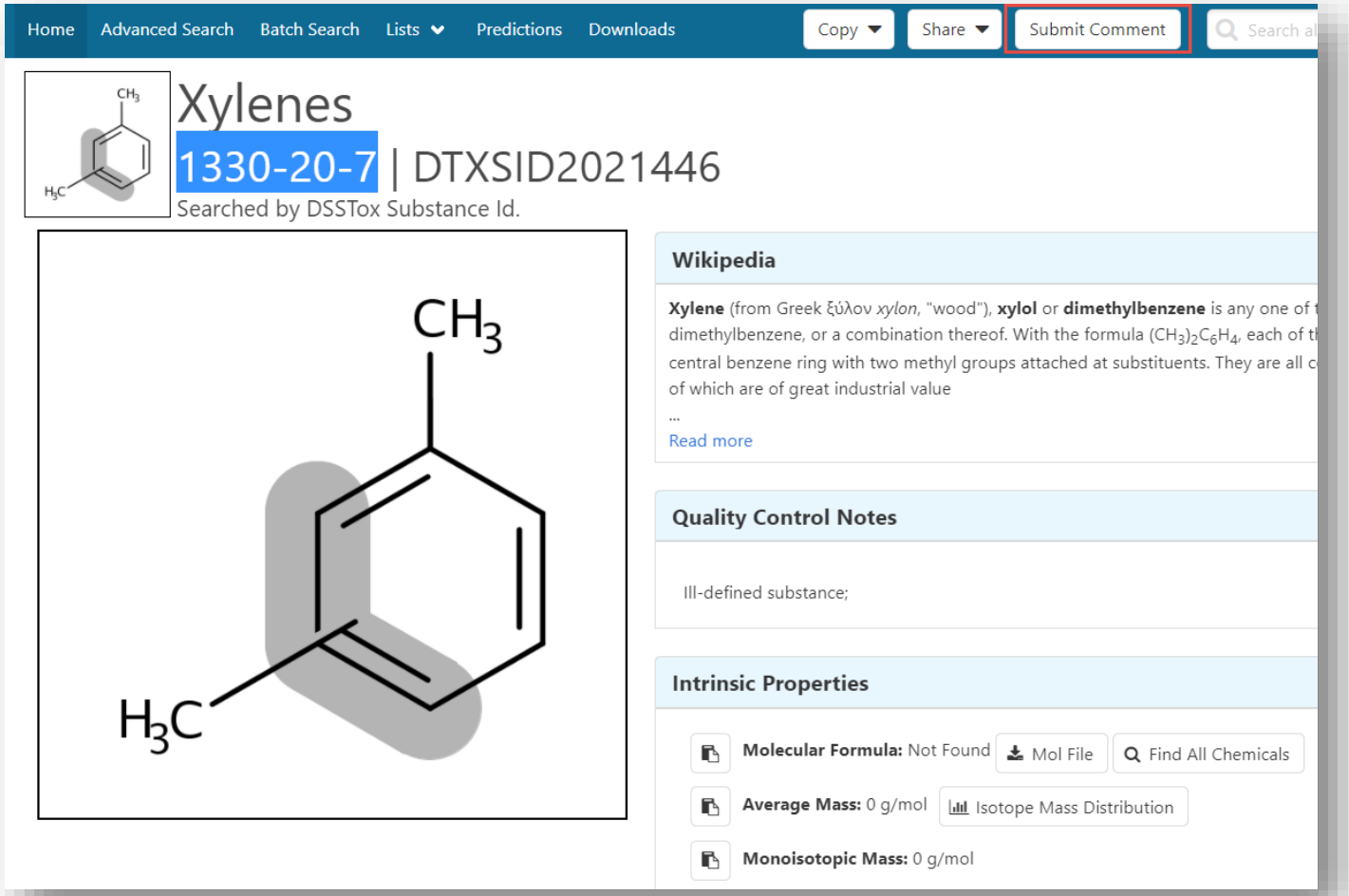
Summary and Conclusions

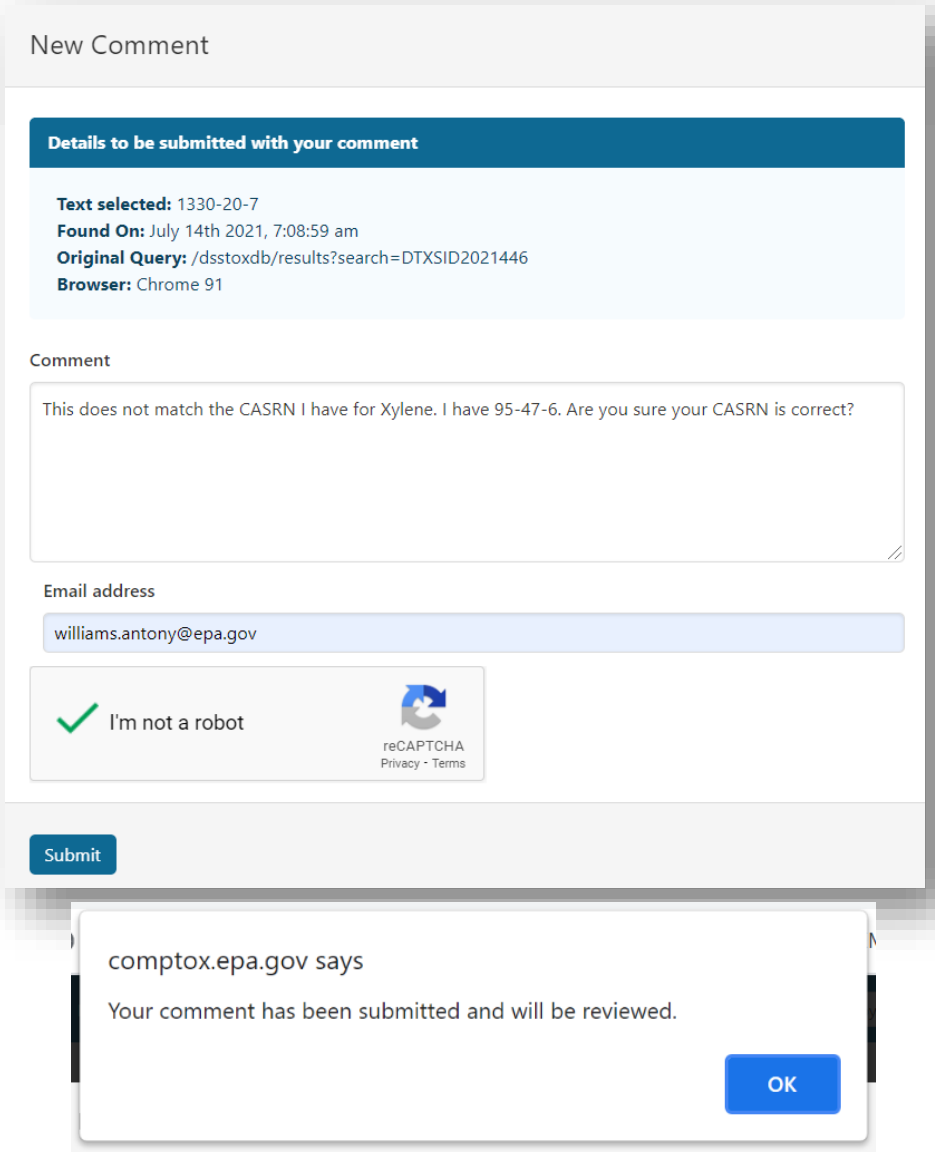
- The Dashboard is one of *many* applications on the internet from which you can source data
- It is not difficult to do a Google search on get some form of answer
- Always question the data, and sometimes even "facts"
- We work hard to qualify, curate and validate data but obtain some from public sources



If you find an error, or want to comment…
Select text and “Submit Comment”

EPA United States Environmental Protection Agency



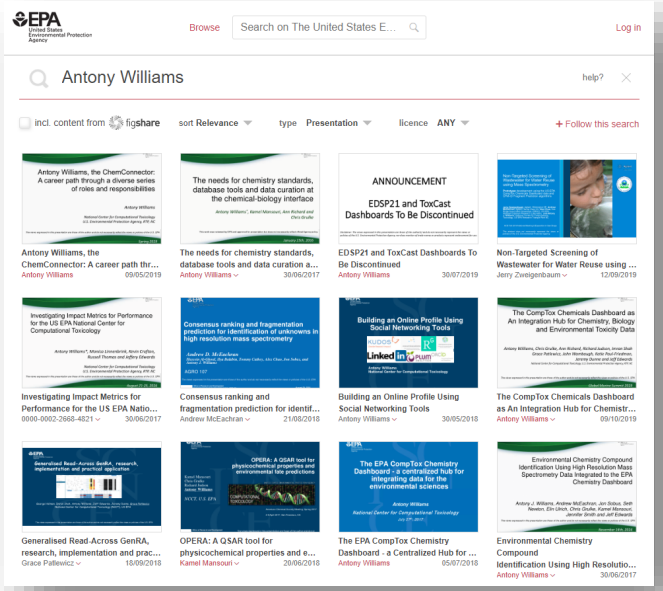


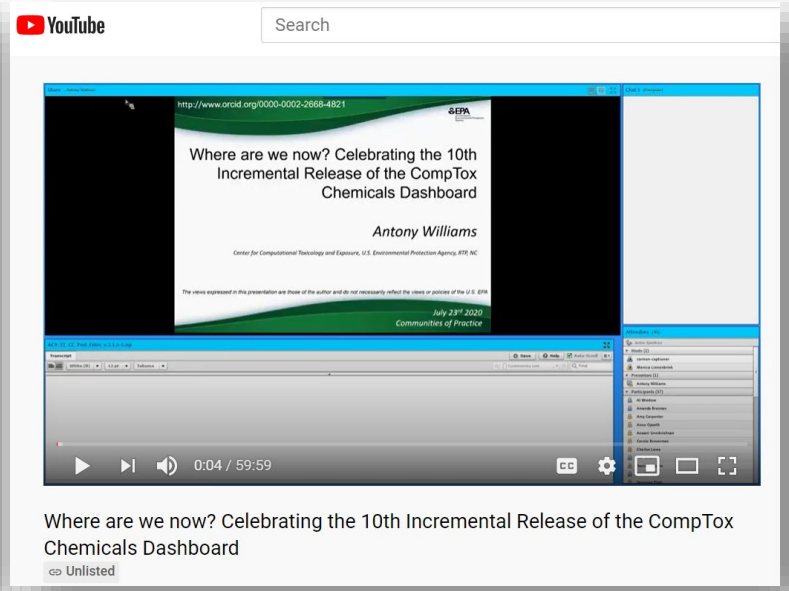
References

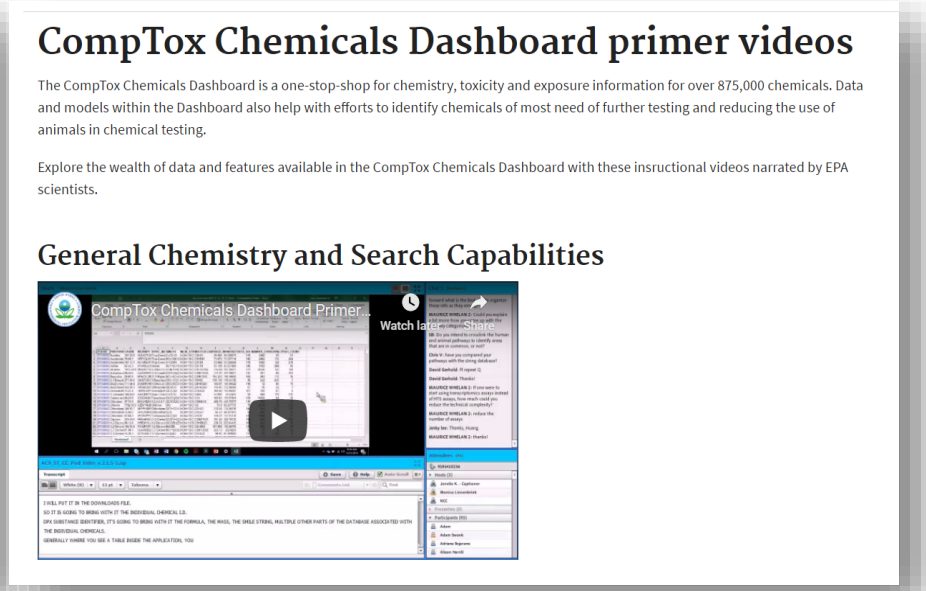
- The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, *J. Cheminformatics*, **9**, 61 (2017)
- EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research, *Comp. Tox.* **12**, 100096 (2019)
- OPERA models for predicting physicochemical properties and environmental fate endpoints, *J. Cheminformatics*, **10**, 10 (2018)
- Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model, *Environ. Sci. Technol.* **49**, 8804-8814 (2015)
- ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology, *Chem. Res. Toxicol.* **29**, 1225-51 (2016)
- Development and Validation of a Computational Model for Androgen Receptor Activity, *Chem. Res. Toxicol.* **30**, 946-964 (2017)
- CERAPP: Collaborative Estrogen Receptor Activity Prediction Project, *Environ. Health Perspect*. **124**, 1023 (2016)
- Abstract Sifter: a comprehensive front-end system to PubMed, *F1000,* **6**, 2164 (2017)

You want to know more…

- Lots of resources available
  - Presentations: https://tinyurl.com/w5hqs55
  - Communities of Practice Videos: https://rb.gy/qsbno1
  - Manual: https://rb.gy/4fgydc
  - Latest News: https://comptox.epa.gov/dashboard/news_info







Acknowledgments

- Contact: Williams.Antony@epa.gov
- Feedback and follow-up is welcomed! Your questions help
- The dashboard is based on the efforts of many more team members than us. Many collaborators provide data also.

EPA's Center for Computational Toxicology and Exposure

BIG DATA IN ENVIRONMENTAL SCIENCE AND TOXICOLOGY

superfund.tamu.edu

TEXAS A&M UNIVERSITY
Superfund Research Center

Thank you for joining us !

Next Session:

Wednesday August 18, 2021 1:00 – 3:00 pm (Central US Time

"New Approach Methods – What is That?"

Antony Williams – US Environmental Protection Agency