Primer on New Version of CompTox Chemicals Dashboard

Antony John Williams
williams.antony@epa.gov
Center for Computational Toxicology and Exposure, US-EPA, RTP, NC

Communities of Practice, February 23rd 2022

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

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

**Computational Toxicology**

*Volume 12, November 2019, 100096*

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

Christopher M. Gruke, Antony J. Williams, Intirany Thilanadarajah, Ann M. Richard

Available at: https://doi.org/10.1016/j.cctox.2019.100096

**Journal of Cheminformatics**

Published: 28 November 2017

The CompTox Chemistry Dashboard: a community data resource for environmental chemistry


Available at: https://dx.doi.org/10.1039/c7qi00162e

**Environment International**

Volume 154, September 2021, 106568

Sourcing data on chemical properties and hazard data from the US-EPA CompTox Chemicals Dashboard: A practical guide for human risk assessment

Antony J. Williams, Jason C. Lambert, Kris Thayer, Jean-Lou C.M. Dorne
Rearchitecting the Dashboard

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

• The Dashboard was a successful “proof-of-concept” tool that required rearchitecting for ongoing support and performance.
The Vision for the Rebuild

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

• GUIDANCE: “1-to-1 mapping of new dashboard to old dashboard”

• EXPECTATIONS:
  • Improve search performance
  • Improve documentation and help manual
  • Unify navigation especially in regard to tabular data handling
  • Make future development easier with a new foundation
New Design - >906,000 chemicals
• On-click informational icon gives latest important information
Data updates across the board

- Data have been updated across the application
  - Chemistry data and lists
  - ToxVal
  - Exposure
  - Bioactivity

- Data versioning is under:
  [https://comptox.epa.gov/dashboard/about](https://comptox.epa.gov/dashboard/about)
New Table Capabilities
http://ccte-ccd.epa.gov/dashboard/chemical-lists/40CFR355

40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities

Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey

Start typing to search.

Identifier substring search

List Details

Description: Extremely Hazardous Substance List and Threshold Planning Quantities; Emergency Planning and Release Notification Requirements; Final Rule. (52 FR 13378) This FR notice contains the EHS list of chemicals as published in 1987. This list has been revised over time and should not be used for current compliance purposes. The current EHS list can be found at 40 CFR 355.

Number of Chemicals: 354

Showing 353 of 353 chemicals

Structure | DTXSID | Preferred Name | CASN | GC Level | %ToxCast Active | %ToxCast Active | %CPDF | Sources | PubChem | PubMed | Monom. Mass | Mol. Formula
--- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | ---

DHNC | DTXSID5021881 | Dihydronicotinamide | 107-15-3 | 2 | 9/398 | 2% | 218 | 170 | 272 | 745 | 60.068748 | C12H6N2
Find all chemicals with N2 in formula

<table>
<thead>
<tr>
<th>CASRN</th>
<th>QC Level</th>
<th>#ToxCast Active</th>
<th>%ToxCast Active</th>
<th>#CPTat</th>
<th>#Sources</th>
<th>#PubChem</th>
<th>#PubMed</th>
<th>Mono. Mass</th>
<th>Mol. Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>107-15-3</td>
<td>2</td>
<td>9/398</td>
<td>2%</td>
<td>218</td>
<td>170</td>
<td>272</td>
<td>745</td>
<td>60.068748</td>
<td>C2H8N2</td>
</tr>
<tr>
<td>77-81-6</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>54</td>
<td>36</td>
<td>271</td>
<td>162.055815</td>
<td>C5H11N2O2P</td>
</tr>
<tr>
<td>26419-73-8</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>38</td>
<td>15</td>
<td>-</td>
<td>234.049670</td>
<td>C8H14N2O2...</td>
</tr>
</tbody>
</table>
Find “aniline” in the names
Substructure searching is better

<table>
<thead>
<tr>
<th>Structure</th>
<th>DTXSID</th>
<th>Preferred Name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DTXSID9024512</td>
<td>3-(Trifluoromethyl)aniline</td>
</tr>
<tr>
<td></td>
<td>DTXSID5043847</td>
<td>2,4,6-Trimethylaniline</td>
</tr>
<tr>
<td></td>
<td>DTXSID8020090</td>
<td>Aniline</td>
</tr>
</tbody>
</table>
Capabilities to switch columns on/off

- Table functionality for selecting columns for view is built in
Capabilities to Filter the Table

- Filtering on the tables is possible
- It is not available on every column on every table
Bioactivity Refresh

- Collapse of previously segregated EDSP21 and ToxCast/Tox21
- Concentration Response Data views have been rebuilt
- New display of ToxCast Summary view and Conc. Response curves
Bioactivity Refresh

- Concentration Response fully reworked
More flexible interface for data viewing
Based on Plotly

- All the advantages of Plotly visualization now built into bioactivity
Direct URL access – useful for referencing in publications

Multiple plots available via URL

Analytical QC Data updated

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

• Tox21 browser data now searchable based on DTXSID

• New grade added “Unknown/Inconclusive”
External Links Updated and Expanded

• Link decay is an ongoing issue and URLs need updating

• New resources added by request (if feasible)
### Abstract Sifter Refresh

#### Literature - PubMed Abstract Sifter

**Abstract Sifter Instructions**

1. **Select PubMed starting point query**
   - [Hazard](#)

2. **Optional, edit the query before retrieving**
   - "(80-05-7 OR 'Bisphenol A') AND (NOEL OR NOEL OR LOEL OR LOEL OR RLD OR 'reference dose' OR 'reference concentration' OR 'adverse effect level' [tab] OR 'cancer slope factor'[tab])"

3. **Click Retrieve Articles to begin download**
   - **RETRIEVE ARTICLES**
   - 143 of 143 articles loaded

4. **Optional, export articles**
   - ![Export button](#)

---

**To find articles quickly, enter terms to sift abstracts.**

<table>
<thead>
<tr>
<th>estrogen</th>
<th>Bisphenol</th>
<th>T</th>
<th>P</th>
<th>PubMed</th>
<th>Y</th>
<th>Title</th>
<th>Authors</th>
<th>Journal</th>
<th>Rev</th>
<th>DOI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>9</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

*Relevance scores are based on the number of terms shared with the query.*

*Authors:*
- Edelfax, Xi; Puish, Kerkoura, Assafy; Egerdly;... Reproductive toxicology (Elmsford, N.Y.)
- Pella-Corona, V. Agehr; Varga, Aybrey;... Reproductive toxicology (Elmsford, N.Y.)
- Chiacchio; Milić; Jličić;... Environmental sciences and pollution research in... (Netherlands)
- Fuku, Campo; Affarman; Bi-Shakhr, Barcade... The science of the total environment
- Vom Saal, Vandenberg... Endocrinology
Abstract Sifter as a TOOL

Literature - PubMed Abstract Sifter

1. Select PubMed starting point query

Choose Query Term

2. Optionally, edit the query before retrieving.

3. Click Retrieve Articles to begin download.

4. Optionally, export articles

Some examples are: Fauci AS (au)
covid-19 spike protein OECD AND qua
migraine AND 2020(dp)
fluorid or 120068-37-3

---

To find articles quickly, enter terms to sift abstracts.

- Judson
- Estrogen
- [CLEAR TERMS]

<table>
<thead>
<tr>
<th>Judson</th>
<th>Estrogen</th>
<th>Title</th>
<th>Authors</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Exposure and Toxicity Characterization of C...</td>
<td>Fantas-Cho, Ayala, Judson, Huang, Jiang...</td>
<td>The International Journal of Life Cycle Assessment</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Using Chemical Structure Information to Da...</td>
<td>Pradips, Patel, Reitzer, Wambough, Vila...</td>
<td>Computational Toxicology</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Correction to: Notch regulates vascular coll...</td>
<td>Gross, Web, Peterlin, Durrstock, Judson, Raz...</td>
<td>Angiogenesis</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Structure-based QSAR Models to Predict R...</td>
<td>Pradips, Friedman, Judson</td>
<td>Computational Toxicology</td>
</tr>
</tbody>
</table>
### Batch Search Refresh

#### Batch Search

**Select Input Type(s)**
- Substance Identifiers
  - Chemical Name
  - CASRN
  - InChIKey
- DSSTox Substance ID
- DSSTox Compound ID
- InChIKey Skeleton
- MS-Ready Formula(e)
- Exact Formula(e)
- Monoisotopic Mass

**Enter Identifiers to Search**

(To enter one identifier per line; processing time increases with number of inputs.)

- DTXSID9020374
- DTXSID9020827
- DTXSID2022678
- DTXSID4023381
- DTXSID9044164
- DTXSID7032004
- DTXSID4022361
- DTXSID8021771

**Options:**
- Display All Chemicals
- Choose Export Options

**45% loaded**

45 Chemicals Found from 110 Input(s)

<table>
<thead>
<tr>
<th>Structure</th>
<th>DTXSID</th>
<th>Preferred Name</th>
<th>CASRN</th>
<th>Mono. Mass</th>
<th>Mol. Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DTXSID2022678</td>
<td>Bicarbonate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DTXSID0290621</td>
<td>Biuret</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Batch Search – List Filtering

Customize Export Results

Your file will be exported in Microsoft Excel Format (.xlsx)

Select all columns available

Chemical Identifiers
- DTXSID
- Chemical Name
- CASRN
- INCHIKey
- EINECS, ELINNE

Structures
- Mol File
- SMILES
- IUPAC String
- MS-Ready SMILES
- OSIRIS, Ready SMILES

Intrinsic and Predicted Properties
- Molecular Formula
- Average Mass
- Monoisotopic Mass
- TEST. Model Predictions
- OPERA Model Predictions

Metadata
- Citation Level Details
- Safety Data
- NAME_Predicted Exposure
- Data Sources
- Include ToxRef Data Availability
- Assay Hit Count
- Number of PubMed Articles
- PubChem Data Sources
- CPDat Product Occurrence Count
- IRIS
- FPRTV
- Wikipedia Article
- QC Notes
- Include links to ACTOR reports

Presence in Lists

- Title
  - Description

Enhanced Data Sheets
- MetaFlag Input File (Beta)
- ToxPrint single fingerprints
- Abstract Sifter Input File
- Synonyms and Identifiers
- Related Substance relationships
- ToxPrint fingerprints (ChemoType)
- Associated ToxCast Assays

Download Export file for the chemicals selected
List Filtering

- List filtering is very beneficial
- Now ~320 lists so new functionality quickly filters
- Select the lists to push flags into the export file
<table>
<thead>
<tr>
<th>DTXSID</th>
<th>PREFERRED_NAME</th>
<th>EPAPFASDW</th>
<th>EPAPFASDW537</th>
<th>EPAPFASDWTREAT</th>
<th>EPAPFASNONDW</th>
<th>EPAPFASVALDW</th>
<th>PFASTDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTXSID4059916</td>
<td>Perfluorobutanoic acid</td>
<td></td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID70191136</td>
<td>Perfluoro-3-methoxypropanoic acid</td>
<td></td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID6067331</td>
<td>6:2 Fluorotelomer sulfonic acid</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID6062599</td>
<td>Perfluoropentanoic acid</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID3031862</td>
<td>Perfluoroheptanoic acid</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID1037303</td>
<td>Perfluorooctanoic acid</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID00192353</td>
<td>8:2 Fluorotelomer sulfonic acid</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID60500450</td>
<td>Perfluoro(4-methoxybutanoic acid)</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID8031865</td>
<td>Perfluorooctanoic acid</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID30382063</td>
<td>Perfluoro-3,6-dioxaheptanoic acid</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID8031863</td>
<td>Perfluorononanoic acid</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID8062600</td>
<td>Perfluoropentanesulfonic acid</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID5030030</td>
<td>Perfluorobutanesulfonic acid</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID5062780</td>
<td>2-(N-Ethylperfluorooctanesulfonamido)acetic acid</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>DTXSID3031860</td>
<td>Perfluorodecanoic acid</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>
The new architecture brings significant performance enhancements especially for list loading and batch searching.

- List Loading: 5000 chemicals: 7 secs vs 21 secs
- Batch Search: 5000 chemicals: 2.5 secs vs 81 secs!
- Search limitation lifted from 5000 to 10,000 inputs
New Online Help

CCD Help Pages

Searching CCD

Simple search

For information about the CCD Batch Search tool, see Batch Search.

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

- Please note that Submit Comments is how we want feedback for CHEMICAL LEVEL detail. Select relevant text and submit.
- All comments will go directly into Jira for tracking and metrics.
Highlight any part of the screen and submit comments will capture the context, URL, date and time etc.
Comments are all viewable
https://comptox.epa.gov/dashboard/comments

<table>
<thead>
<tr>
<th>Chemical Name</th>
<th>Structure</th>
<th>DTXSID</th>
<th>Date Submitted</th>
<th>Comment</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E,E)-2,4-Hexadienoic acid, propyl ester</td>
<td></td>
<td>DTXSID101888111</td>
<td>January 21, 2022</td>
<td>You seem to have the name 2,4-Hexadienoic acid, propyl ester. (2E,4E)- 2,4-Hexadienoic acid, propyl ester. (2E,4E)- on file twice as a synonym. The name (E,E)-Propyl 2,4-hexadienoate is a bad one. steroid belongs with the acid end, not the ester/alcohol end?? I would remove it! You could add FEMA-4614 UNII-1579567H1</td>
<td>Resolved</td>
</tr>
<tr>
<td>Benzilic acid, 4-hydroxy-3-methoxy-ethyl ester</td>
<td></td>
<td>DTXSID2060670</td>
<td>January 21, 2022</td>
<td>Consider adding Ethyl vanilate as a common synonym. See for example: <a href="https://www.sigmaaldrich.com/US/en/product/aldrich/6499267">https://www.sigmaaldrich.com/US/en/product/aldrich/6499267</a></td>
<td>Resolved</td>
</tr>
<tr>
<td>2,4-Dioxo-2,4-dimethyloctane</td>
<td></td>
<td>DTXSID00904130</td>
<td>January 21, 2022</td>
<td>Should be &quot;ether&quot;, not &quot;etha&quot;</td>
<td>Resolved</td>
</tr>
<tr>
<td>2,4-Dihydroxy-5-11,13,15-triheptadecylpentadecene</td>
<td></td>
<td>DTXSID00804377</td>
<td>January 20, 2022</td>
<td><a href="https://comptox.epa.gov/dashboard/calculation-details?model_id=22&amp;search=804377">https://comptox.epa.gov/dashboard/calculation-details?model_id=22&amp;search=804377</a> This link is not loading the page</td>
<td>Resolved</td>
</tr>
</tbody>
</table>
Work in Progress

- Data updates in preparation – >1.2M chemicals release next

**CAS Common Chemistry™ expands collection of publicly available chemical information**

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

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

- Structure/substructure/similarity search module in development
Access the CCD-API

• We WANT you to use the underlying data

• API (application programming interface) servicing the Dashboard will go public this year

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

- A much-needed re-architecture has been successfully delivered
- This sets us up for more frequent data releases and is a foundation for new functionality
- Thank you for all of your feedback and support to this point
• Feedback and follow-up is welcomed! Your questions help

• Use the comments system to provide feedback

• Thanks to all data providers, the software development and infrastructure teams, our scientific staff and our management for support and guidance
Transition of Product Owners

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

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

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