

Literature Survey Guide

Draft 6

Contract: CPSC- CPSC-D-17-0001

Task Order: 61320621F1001

October 17, 2022

Prepared by:

Risk Science Center
Department of Environmental and Public Health Science
University of Cincinnati

Prepared for:

Charles Bevington, MPH
U.S. Consumer Product Safety Commission

Contact:

Lynne Haber (Lynne.Haber@uc.edu)

Document Authors/Contributors:

Mark Bradley
Annette Christianson
Sarah Fitzpatrick
Lynne Haber
Jacqueline Patterson
Sanjana Yerubandi
Ashley Williams (ICF)
Nicole Vetter (ICF)

Contents

List of abbreviations	7
1. Introduction, Scope, and Document Roadmap.....	9
1.1 Introduction and Version History	9
1.2 Scope of Literature Surveys and Searches for the OFR project	9
1.3 Roadmap of this Guide	10
2. Setup: Identifying chemicals and their synonyms, and finding and downloading key data .	11
2.1 Identify the OFRs in class and analogs.....	11
2.2 Uniquely identify each OFR and analog, using CAS RN and SMILES, and download key data	11
2.2.1 How to do a bulk download from the EPA CompTox Chemical Dashboard.....	13
2.2.2 How to identify duplicate SMILES using Excel	16
2.3 Identify breadth of available toxicity data	16
2.3.1 Generic Approach.....	17
2.3.2 Application for identifying availability of toxicity data.....	18
2.3.3 Flagging chemicals lacking data.....	19
3. Searching for Database Data	20
3.1 Comptox Dashboard and ToxVal Dashboard Hazard Data	20
3.2 OECD QSAR Toolbox	21
3.3 ToxCast and Tox21 (WebICE).....	26
3.4 Using PubChem to Identify Available Toxicity Data.....	31
3.4.1 Using PubChem Identifier Exchange Service	32
3.4.2 Using Upload ID List function in PubChem and Downloading Data	34
3.4.3 How to Download PubChem Bioassay Information in Bulk	40
3.4.4 The Comparative Toxicogenomic Database (CTD) as an alternative source to chemical-gene data from PubChem.....	40
3.5 Danish QSAR Database.....	41
3.6 Additional Grey Literature Sources Treated as Databases	42
3.7 MMDB (Multi-Media Monitoring Database).....	42
4. Develop database data format as needed for toxicity data and extract relevant data from databases.	42

4. 1 Approach for Database Logic	42
4. 2 Assumptions and Considerations	44
5. Searching for Peer-Review and Grey Literature (pdf Data).....	48
5.1 Targeted Searching for Completed Assessments and Literature Reviews in Bibliographic Databases.....	48
5.1.1 Literature type and study type terms	48
5.1.2 Class-specific search terms.....	49
5.1.3 Chemical-specific search terms	51
5.1.4 Documenting Search Meta Data.....	53
5.1.5 Saving the search and uploading to Endnote.....	53
5.2 Peer-reviewed (and grey) literature deduplication, screening, and full-text retrieval ..	55
5.2.1 How to remove duplicates and screened-out references in EndNote, while preserving a record of individual searches and their results.....	57
5.2.2 Deduplicating peer-reviewed against previous classes and grey literature against previous classes and against peer-reviewed literature.....	59
5.2.3 How to export an EndNote library as a .ris file, for import into DistillerSR	63
5.2.4 How to retrieve PDFs for references that pass DistillerSR Title/Abstract Screening ..	63
5.3 QA Review of PDF Sources	68
5.3.1 QA checks during PDF retrieval after DistillerSR title/abstract screening and prior to full text screening.	68
5.3.2 Additional QA steps after PDF retrieval	69
5.3.3 Keyword highlighting of PDFs.....	70
5.3.4 PRISMA diagrams for each class.....	70
5.4 Grey literature searching.....	71
5.4.0 How to obtain ATSDR Toxicological Profiles, and EPA IRIS and PPRTV documents	73
5.4.1 How to obtain EPA TSCA Data from EPA ChemView	74
5.4.2 How to get detailed HPVIS information	78
5.4.3 How to search for EPA documents via the National Service Center for Environmental Publications (NSCEP)	83
5.4.4 How to search for documents from Health Canada.....	84
5.4.5 How to search for ECHA assessments and links to dossiers.....	85

5.4.6	How to download ECHA REACH study results (dossier info) via IUCLID	86
5.4.7	How to search for NTP publications	94
5.4.8	How to search for documents on INCHEM	94
5.4.9	How to search for assessments on OECD Existing Chemicals Database	96
5.4.10	How to search for reviews on Cochrane.....	98
6.	Tagging.....	99
6.1	Tagging and aggregating database data	99
6.2	Tagging for peer-reviewed and grey literature	99
6.2.1	Data cleaning and preliminary visualization using Power Query and Pivot Tables in Excel	102
6.2.2	Selection of studies for 3B tagging.....	104
7.	Evidence Mapping.....	105

Table of Figures

Figure 1.	CompTox Chemicals Dashboard home screen.....	13
Figure 2.	Example batch search input on CompTox Dashboard	14
Figure 3.	Example of downloading SMILES and QSAR-Ready SMILES data from the CompTox Dashboard	15
Figure 4.	Example of input file format for OECD QSAR Toolbox search	21
Figure 5.	QSAR Toolbox List Input screen	22
Figure 6.	QSAR Toolbox screen after data is inputted.....	22
Figure 7.	QSAR Toolbox Data Tab	23
Figure 8.	Example of selecting databases for the search	24
Figure 9.	Selecting endpoints for the search.....	24
Figure 10.	Example of downloading data from the QSAR Toolbox	25
Figure 11.	Partial example of spreadsheet of data downloaded from OECD QSAR Toolbox.....	26
Figure 12.	Example of no chemicals returning data from OECD QSAR Toolbox	26
Figure 13.	Example of user-supplied CAS numbers	27
Figure 14.	Example of cHTS assays selected, with all subcategories included	28
Figure 15.	Example of search results page on WebICE	28
Figure 16.	Data formats for downloading results on WebICE	29
Figure 17.	Example of downloaded data in “Summary Data” format from WebICE	30
Figure 18.	Example of downloaded data in “Wide Format Data” from WebICE	30
Figure 19.	Example of downloaded data in “Wide Format Data” from WebICE with number only tab.....	30

Figure 20. Example of downloaded data in “Long Format Data” from WebICE	31
Figure 21. Example of PubChem Identifier Exchange Service input screen.....	33
Figure 22. Example of PubChem homepage, showing the upload ID list.....	34
Figure 23. PubChem Uploading an ID list screen	35
Figure 24. Example of the preview of CIDs after uploading an ID list.....	36
Figure 25. Example chemical profile.....	38
Figure 26. Downloading chemical data from PubChem.....	38
Figure 27. Example of downloaded bioassay data.....	39
Figure 28. Sample of PubChem Analytic Laboratory Methods saved as a PDF	39
Figure 29. EndNote Preferences for PDF Auto Renaming.....	66
Figure 30. ChemView homepage	75
Figure 31. Example of inputting a chemical in ChemView.....	76
Figure 32. Example of search results for a ChemView batch search	76
Figure 33. Example of search results being downloaded from ChemView	77
Figure 34. Example of exported results from EPA ChemView.....	77
Figure 35. Example of EPA Assessments window to save as a PDF document	78
Figure 36. Example of entering a CAS number to search HPVIS.....	78
Figure 37. How to access the "Entire Dataset" in HPVIS	79
Figure 38. Example of "Entire Dataset" page in HPVIS	79
Figure 39. Example of acute toxicity section of "Entire Dataset" page in HPVIS	80
Figure 40. Navigating to HPVIS section on Chem/View	81
Figure 41. Navigating to HPVIS website from ChemView	82
Figure 42. HPVIS data type counts.....	83
Figure 43. Example of searching the Canadian government websites for a certain chemical.....	84
Figure 44. Example of searching the ECHA website for a certain chemical	85
Figure 45. Example of search results when finding ECHA documents	86
Figure 46. Example of downloading IUCLID REACH Study Results	87
Figure 47. Download page for IUCLID 6 Desktop Installer	88
Figure 48. Example of searching for a CAS number in the downloaded Excel spreadsheet	89
Figure 49. Searching for a UUID in the folder of IUCLID dossiers	89
Figure 50. Example of dossier file loaded into IUCLID program.....	90
Figure 51. Example of main screen when dossier is opened	91
Figure 52. Creating a component PDF of the dossier information.....	91
Figure 53. PDF settings window.....	92
Figure 54. Deselecting irrelevant information and creating the PDF	93
Figure 55. NTP Publications home screen search bar	94
Figure 56. INCHEM home screen	95
Figure 57. Example of searching by CAS number on INCHEM	95
Figure 58. OECD Existing Chemicals Database home screen	96
Figure 59. Example of search input on OECD Existing Chemicals Database	97

Figure 60. Example of chemical detailed results on OECD Existing Chemicals Database	98
Figure 61. Example of Evidence Map	107

Table of Tables

Table 1. Constructing a formula to test data availability	Error! Bookmark not defined.
Table 2. Specifics for flagging availability of toxicity data.....	Error! Bookmark not defined.
Table 3. Example table for tab definitions in CTD downloads	41
Table 4. Names of EndNote libraries generated at each step of import, deduplication, and initial screen in EndNote	56
Table 5. Examples of references removed in initial screen of PHOP class peer-reviewed literature for purposes of relevance, with notes on reasons and DOI for reference	56
Table 6. Information definition and source for counts of references in PRISMA diagrams	70
Table 7. Example columns from a lookup table. Note the removal of a comma from the CAS number, and the replacement of a non-unique abbreviation with a unique one	103

List of abbreviations

AC10 = activity concentration at 10%; concentration that results in 10% activation of enzyme

AC50 = activity concentration at 50%; concentration that results in 50% activation of enzyme; half-maximal activity concentration

ACC = activity concentration at cut-off, a measure of the activity threshold for an assay response based on curve-fitting models

ATSDR = Agency for Toxic Substances and Disease Registry

BFR = Brominated Flame Retardant

CAS RN = Chemical Abstract Service Registration Number

CEBS = Chemical Effects in Biological Systems

cHTS = Curated high-throughput screening

CID = Chemical ID

CPSC = Consumer Product Safety Commission

CTD = Comparative Toxicogenomics Database

DA = Defined Approach

DNEL = Derived No-Effect Level

DOI = Digital Object Identifier

DSL = Domestic Substances List

DTXSID = DSSTox substance identifier

ECHA = European Chemicals Agency

EHC = Environmental Health Criteria

EPA = Environmental Protection Agency

GO = Gene Ontology

HERO = Health and Environmental Research Online

HH = Human Health

HPV = High Production Volume

HPVIS = High Production Volume Information System

IARC = International Agency for Research on Cancer

IATA = Integrated Approaches to Testing and Assessment

ICE = Integrated Chemical Environment

ICF = ICF International

ILL = Interlibrary loan

InChIKey = IUPAC International Chemical Identifier

IPCS = International Programme on Chemical Safety

IRIS = Integrated Risk Information System

IUCLID = International Uniform Chemical Information Database

JECFA = Joint FAO/WHO Expert Committee on Food Additives

JMPR = Joint FAO/WHO Meeting on Pesticide Residues

LC50 = Lethal Concentration 50%

LD50 = Lethal Dose 50%

LOAEL = Lowest Observed Adverse Effect Level
MMDB = Multi-Media Monitoring Database
MOA = Mode of (Toxic) Action
MRL = Minimal Risk Levels
NAM = New Approach Methodology
NAS = National Academy of Sciences
NBFR = Novel brominated flame retardant
NCBI = National Center for Biotechnology Information
NCCT = National Center for Computational Toxicology
NHANES = National Health and Nutrition Examination Survey
NTP = National Toxicology Program
OECD = Organisation for Economic Co-operation and Development
OEHHA = Office of Environmental Health Hazard Assessment
OFR = Organohalogen Flame Retardant
OPERA = OPEn structure-activity/property Relationship App
ORD = Office of Research and Development
PECO = Populations, Exposures, Comparisons, Outcomes
PHBAFs = Polyhalogenated Bisphenol Aliphatic Ethers
PHOPs = Polyhalogenated Organophosphates
PPRTV = Provisional Peer-Reviewed Toxicity Value
QA = Quality Assurance
QSAR = Quantitative Structure-Activity Relationships
REACH = Registration, Evaluation, Authorization, and Restriction of CHemicals
RfD = Reference Dose
RIS = Research Information Systems
SAR = Structure Activity Relationship
SIARS = SIDS Initial Assessment Reports
SIDS = Screening Information Data Set
SMILES = Simplified Molecular Input Line Entry System
SOP = Standard Operating Procedure
SOW = Statement of Work
TEST = Toxicity Estimation Software Tool
TRV = Toxicity reference value
TSCA = Toxic Substances Control Act
UC = University of Cincinnati
UUIDs = Universally Unique Identifier, used as IUCLID Dossier IDs
WHO = World Health Organization
WoS = Web of Science

1. Introduction, Scope, and Document Roadmap

1.1 Introduction and Version History

Version 1.0: Delivered to CPSC staff for comment on March 3, 2021; comments received March 12, 2021

Version 2.0: Delivered to CPSC staff for comment on May 11, 2021; comments received June 1, 2021

Version 3.0: Available to CPSC staff May – December, 2021, with ongoing updates

Version 4.0: Delivered to CPSC staff December 17, 2021

Version 5.0: Delivered to CPSC staff February 28, 2022, updated to reflect the entire process

Version 6.0 : Delivered to CPSC staff October 17, 2022, with final updates

This document is the fifth draft of the process guide documenting the approach for completing the literature survey for a class of Organohalogen Flame Retardants (OFRs), as part of Task Order 61320621F1001. The purpose of this guide is to be a detailed “how to” document that will guide the completion of future literature surveys and ensure consistency among the 14 OFR classes. It documents aspects such as databases to download and search terms for those databases; “how to” for batch downloads; search strategy for pdfs of reviews and assessments; approaches for tagging and organizing data; and approaches for developing evidence maps. The guide is a living document that details the approach to be undertaken for conducting literature surveys. It was updated frequently throughout the conduct of this task order, reflecting comments from CPSC staff review of the document, as well as the evolution and enhancements of the process as the work progressed. A separate guide has been developed for the screening and tagging of pdfs in Distiller® and post-Distiller® analyses.

1.2 Scope of Literature Surveys and Searches for the OFR project

Literature surveys are intended to gather readily available toxicity, exposure, and risk information. These surveys characterize which kinds and how much data is available for chemicals (and analogs) within a class. CPSC staff defined data sources for literature surveys as toxicity and chemistry databases; completed toxicity, exposure, or risk assessments; and completed literature reviews. Sources identified in the literature survey are screened to confirm utility and identify the type of data, but the actual data are not evaluated and extracted. The end result of a literature survey is a data evidence map that summarizes readily available information, determines whether treatment as a class is plausible or whether such treatment should be re-evaluated, and informs literature search efforts through development of PECO statements and conceptual models. Data gathered in literature surveys can be used to support class-based risk assessment. Literature surveys cover both OFR chemicals and analogs, and **are not** intended to identify all possible data.

In contrast, literature searches are intended to gather **all** (to the extent practical) toxicity, exposure, and risk information. Scientists screen literature searches to identify relevant data types for OFR chemicals in a class. After searching and screening, data evaluation, extraction, and integration are also part of literature searching efforts, but not part of the literature survey. CPSC staff define data sources for literature searches as peer reviewed literature from bibliographic databases. (It is assumed that all relevant data from other sources, including database data and grey literature, will be identified in the literature survey, although the actual extraction and evaluation of such data will be conducted as part of the assessment.) The end result of a literature search is a curated bibliography of citations and references that can be used to inform class-based risk assessment. Literature searches cover OFR chemicals, may also cover targeted analogs, and are designed to identify all possible data to the extent practical.

1.3 Roadmap of this Guide

This guide covers the following:

- 1) Introduction, scope and roadmap
- 2) Setup: Identifying chemicals and their synonyms, and finding and downloading key data
 - a) This section outlines how OFRs in the class and their analogs were originally identified, how to uniquely identify chemicals in various databases and download data, how to convert between the different identifiers used by different databases, and how to identify synonyms for each OFR and analog.
- 3) Searching for Database Data
 - a) This section outlines how to search for and download data from several databases, including EPA Comptox Dashboard, OECD QSAR Toolbox, ToxCast and Tox21 via ICE, the Multi-Media Monitoring Database (MMDB), and PubChem BioAssay. It also outlines how to identify and narrow down the list of analogs to only those analogs with data.
- 4) Developing database format as needed for toxicity data and extracting relevant data from databases
 - a) This section outlines how decisions were made regarding database formatting and combination, and extraction of data from databases. It describes how such formatting, combination, and extraction was done, once such decisions are made.
- 5) Searching for Peer-Review and Grey Literature (pdf Data)
 - a) This section outlines a targeted search strategy for completed assessments and literature reviews, including outlining search terms and syntax for PubMed (and Web of Science, where appropriate) targeting type of literature and study; class-specific search terms (e.g., targeting organophosphate flame retardants); and chemical-specific search terms. It also covers retrieval and pre-screening of peer-reviewed literature using EndNote, prior to moving to Distiller; and a search strategy for grey literature, noting sites and assessments to search for and how to search depending on the source.
- 6) Tagging
 - a) This section covers how data from databases and from pdf data sources (peer-reviewed and grey literature) are tagged. It also covers information on aggregating different databases, different levels of tags to be applied to pdf data sources, an outline of a pilot

- tagging process, and quality assurance procedures and standards for tagging.
- 7) Evidence Mapping
 - a) This section defines and gives examples of evidence mapping, outlines approaches to evidence mapping, and describes the final approach to evidence mapping.

2. Setup: Identifying chemicals and their synonyms, and finding and downloading key data

2.1 Identify the OFRs in class and analogs

NAS compiled an OFR data inventory from 7 sources and used these data to generate 14 sub-classes for 161 chemical substances. Next, EPA NCCT, in collaboration with CPSC staff, compiled an OFR inventory from 20 data sources and used these data to build out and refine the list of chemical substances included in the original NAS inventory, for the 14 sub-classes. Chemoinformatic modeling was used both (1) to confirm that the chemical belonged to the sub-class, and (2) along with data from other types of flame retardants (inorganic, polymeric, non-halogenated), to confirm that the chemical was a flame retardant. A flame retardant list is housed in the US EPA Comptox Chemistry Dashboard (referred to as the “Dashboard” for the rest of this guide). Cheminformatic modeling was used to generate the sub-classes and identify analog substances within high Tanimoto scores (>0.8) for these substances.

2.2 Uniquely identify each OFR and analog, using CAS RN and SMILES, and download key data

A critical part of the database searching described in later sections of this guide is uniquely identifying each chemical. Historically, the Chemical Abstract Service Registration Number (CAS RN or CAS number) has been commonly used to identify chemicals. Although the CAS number is commonly used and is often sufficient for identification, it has a number of limitations. In some cases, there is no one-to-one link between the CAS number and chemical structure, and a single CAS number may refer to multiple molecular structures (e.g., for complex substances or mixtures), while different oxidation states of the same metal may have different CAS numbers, and some structures may not have CAS numbers. In addition, CAS numbers are proprietary to the Chemical Abstract Service. Therefore, alternative approaches are used by some databases, such as the European Community number (EC number), to uniquely identify molecular structures. The DTXSID is used as a bridge for CAS ID, SMILES, and chemical name.

There is a separate market-use contract which will be providing an organohalogen chemistry database. PUBCHEM ID number can also be used as a unique identifier to resolve discrepancies between CAS and DTXSID.

A common approach for uniquely identifying chemicals is using SMILES (simplified molecular-input line entry system). This system has the advantage that a given SMILES string uniquely defines a chemical structure, and someone knowledgeable of the system can draw a molecule directly from a SMILES string. A variant of SMILES called “QSAR-ready SMILES” may also be useful for identifying chemicals. QSAR-Ready SMILES are derived from SMILES by removing information on salts, ionization, and stereochemistry; QSAR-Ready SMILES are used, for example in OPERA and other prediction-based models.

Once a flame retardant or analog has been uniquely identified, the batch download feature of the EPA dashboard was used to download chemical identifier information, OPERA physical-chemical property information, and other useful starting information on the availability of toxicity information. Class-specific OFR spreadsheets with this information for both OFRs and analogs were used as a starting point for literature surveys. Specifically, the initial download includes the following fields. (Left/right refers to the column in the dashboard batch download.):

1. DTXSID (left)
2. Chemical Name (left)
3. CASRN (left)
4. QSAR ready SMILES (left)
5. Average Mass (left)
6. TEST model predictions (left)
7. OPERA model predictions (left)
8. Data sources (left)
9. Include ToxVal Data Availability (left)
10. Number of PubMed Articles (left)
11. Number of PubChem Data Sources (left)
12. IRIS (left)
13. PPRTV (left)
14. Synonyms and identifiers¹ (left)
15. ToxValDB Details (left)
16. ATSDR Toxicological Profiles (right)
17. ATSDR: Minimal risk levels (right)
18. ATSDR: Toxic Substances Chemical Portal List (right)
19. California Office of Environmental Health hazard Assessment (right)
20. Canada Domestic Substances List (right)
21. EPA; Chemicals mapped to HERO (right)
22. EPA: High Production Volume List (right)
23. EPA: Provisional Advisory Levels (Inhalation) (right)
24. EPA: Provisional Advisory levels (oral) (right)
25. List: National Health and Nutrition Examination Survey (NHANES) data (right)
26. Tox21SL: Tox21 Screening Library (right)

¹ An alternative approach to identifying synonyms is via the left-hand navigation bar in an individual chemical CompTox file. This has the advantage of including information on synonym quality, and typically only “valid” or “good” synonyms would be used. However, this approach requires individual chemical downloading, and is not needed at the survey step. The synonyms downloaded as part of the batch download are a more limited list.

27. ToxCast: EPA ToxCast Screening Library (right)

2.2.1 How to do a bulk download from the EPA CompTox Chemical Dashboard

Note that the Dashboard has been updated since this section of the guide was written. The October 2021 update to the Dashboard includes a newly designed interface. While the new interface is similar, the precise location of specific items may have changed in addition to the new style.

1. Go to the following website: <https://comptox.epa.gov/dashboard>
2. Click on the tab labeled as “Batch Search” along the top of the page, as illustrated in Figure 1.

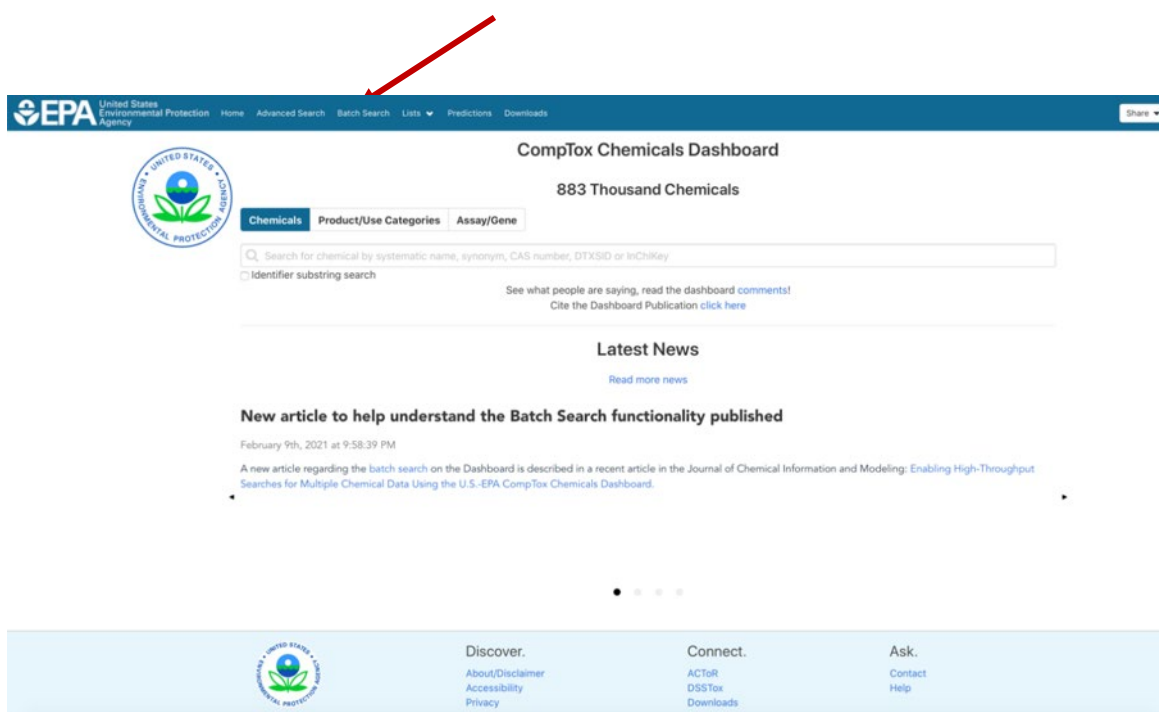


Figure 1. CompTox Chemicals Dashboard home screen

3. To input chemicals based on CAS numbers, select the box labeled as “CASRN” on the left. Then, copy and paste a list of CAS numbers into the box on the right of the screen. Each CAS number should be on a separate line, as illustrated in Figure 2.
 - a. Note that in some instances, the Dashboard cannot complete a search based on a CAS number. In other instances, the information initially provided for CAS number may be of the format NOCAS_####. In these instances, DTXSID should be used instead. CAS numbers and DTXSIDs can be searched simultaneously by checking both boxes under “Input Type(s).”
4. Click on the “Download Chemical Data” button below the list of input criteria, as illustrated in Figure 2.

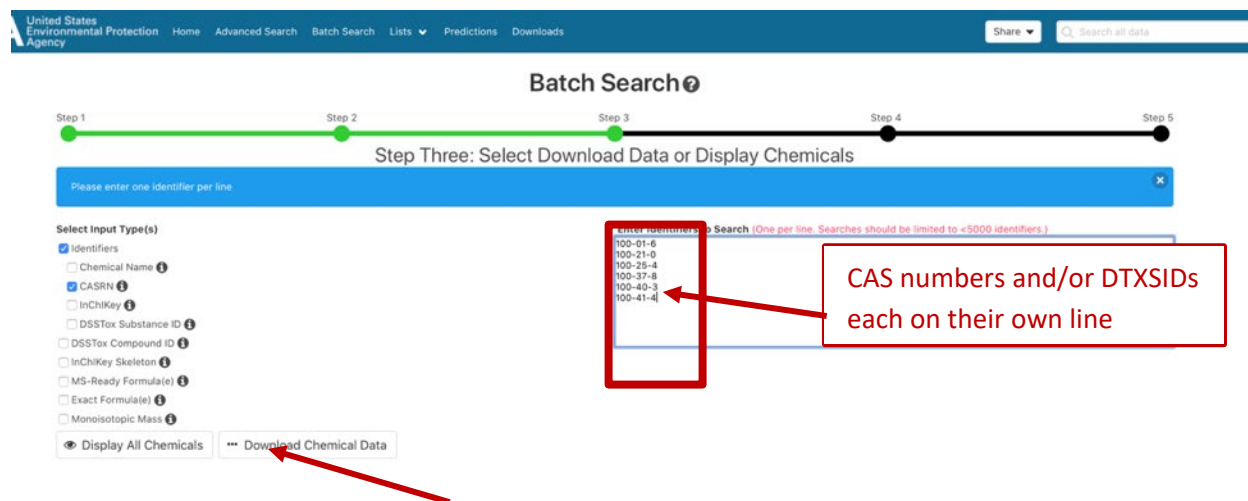


Figure 2. Example batch search input on CompTox Dashboard

5. Scroll down to select the desired output format. In order to download the required data as listed above in section 2.2, select
 - a. The corresponding 15 boxes in the left column
 - i. DTXSID,
 - ii. Chemical Name,
 - iii. CASRN,
 - iv. QSAR-ready SMILES,
 - v. Average Mass,
 - vi. TEST Model Predictions,
 - vii. OPERA model predictions,
 - viii. Data Sources,
 - ix. Include ToxVal Data Availability,
 - x. Number of PubMed Articles,
 - xi. PubChem Data Sources,
 - xii. IRIS,
 - xiii. PPRTV, and
 - xiv. Synonyms and identifiers
 - xv. ToxValDB Details.
 - b. And the corresponding 12 boxes in the right column under “Presence in Lists”
 - i. ATSDR Toxicological Profiles,
 - ii. ATSDR: Minimal risk levels,
 - iii. ATSDR: Toxic Substances Portal Chemical List,
 - iv. California Office of Environmental Health Hazard Assessment,
 - v. Canadian Domestic Substances List 2019,
 - vi. EPA; Chemicals mapped to HERO,
 - vii. EPA: High Production Volume List,
 - viii. EPA: Provisional Advisory Levels (Inhalation),
 - ix. EPA: Provisional Advisory Levels (Oral),

- x. List: National Health and Nutrition Examination Survey (NHANES) data,
 - xi. Tox21SL: Tox21 Screening Library, and
 - xii. ToxCast: EPA ToxCast Screening Library
- c. Note: the “Presence in Lists” section now includes options to search and filter lists by list title and by list description. This is helpful for finding the required lists. However, **be sure to remove any text from the search boxes prior to clicking download**. If searches/filters are still applied when downloading, only data for visible checkboxes will be downloaded, regardless of any other checkboxes that were selected previously.
- d. Figure 3 below illustrates a partial sample for illustration purposes only, selecting only downloads for SMILES and QSAR-ready SMILES

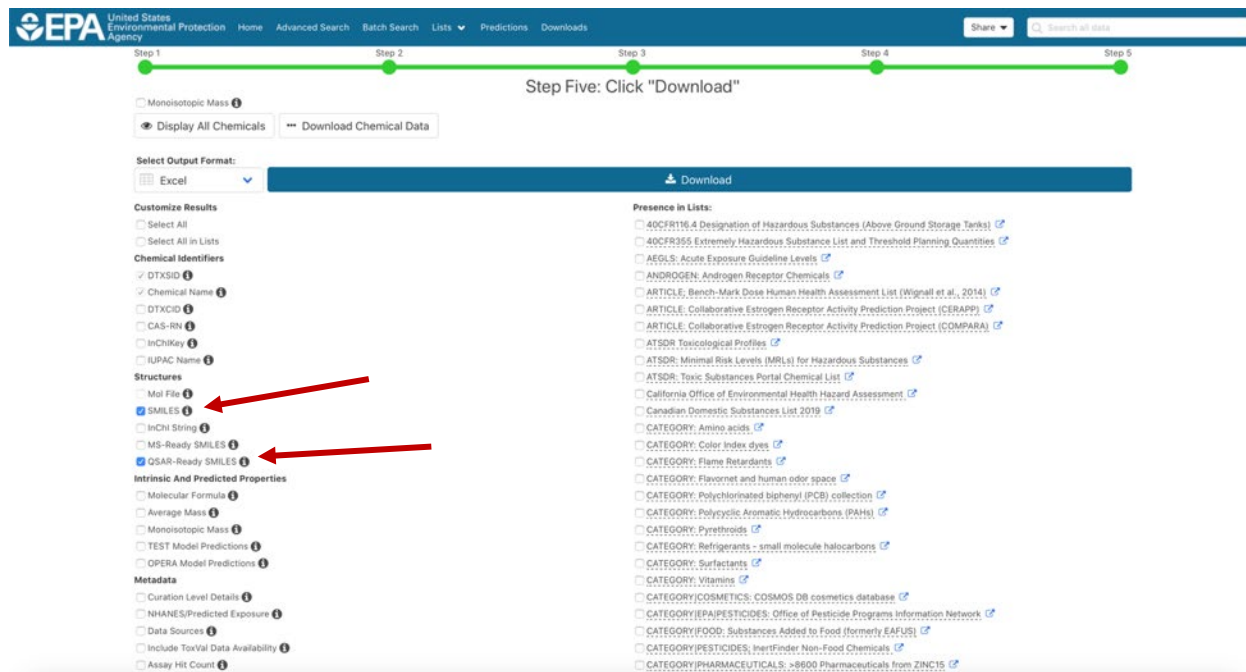


Figure 3. Example of downloading SMILES and QSAR-Ready SMILES data from the CompTox Dashboard

6. Click “Download” to download an excel spreadsheet with a table including all of the requested information, if available.
 - a. Notes for comparison before and after the October 2021 Dashboard update:
 - i. After the update, OPERA predictions include 2 additional columns (OPERA_PKAA_OPERA_PRED, and OPERA_PKAB_OPERA_PRED) which were not included previously.
 - ii. Some columns which existed before the update will appear in a different order in the downloaded data after the update.
7. Name the file as follows: CompTox Batch Search_ CLASS NAME_ DATE

2.2.2 How to identify duplicate SMILES using Excel

It is possible that chemicals with different CASRN will return identical SMILES from the EPA CompTox Dashboard. To identify identical SMILES, for each of 1) class members, and 2) analogs:

1. Open the file downloaded from CompTox.
2. Highlight the column containing SMILES (QSAR-Ready SMILES).
3. In the top menu, select Home > Conditional Formatting > Highlight Cell Rules > Duplicate Values. Duplicate SMILES will now be highlighted in red.
4. Determine which highlighted cells are duplicates of which other cells. Sorting the data by the SMILES column makes this comparison easier.
 1. **Note** that more than two chemicals may share an identical SMILES. E.g., in the PHOP class, a group of 5 chemicals (identified by 5 separate CASRNs) shared a single SMILES.
5. If any cells are found to be duplicates, use ChemIDplus to check 2D/3D structures – e.g., make sure they aren't isomers or something similar. Search by Registry Number under "Substance Identification" and compare.
 1. <https://chem.nlm.nih.gov/chemidplus/>
6. If they are not isomers, the less data-rich chemical may be able to be eliminated from later searching.

It is also possible that the CompTox Dashboard will not return a SMILES for some chemicals. In this case, SMILES may be obtained by searching ChemIDplus by CASRN. If a SMILES is then found, it should be rechecked against existing SMILES for duplicates, as outlined above.

2.3 Identify breadth of available toxicity data

As noted further in Section 5, some aspects of the survey depend on the amount of available toxicity information. It is possible to get a good initial sense of the available information based solely on information easily obtainable early in the survey process. This allows the scientist to focus efforts on the chemicals with the most information. For example, Section 5 notes that more synonyms are evaluated for data-rich chemicals when searching bibliographic databases. Conversely, where multiple sources show no available toxicity information for an analog, that analog will clearly not be useful as a source of toxicity information, and so it is not useful to go through the entire search process for that chemical. This section first provides generic information for noting data availability, and then describes how that generic approach is applied in this task. The result of this section will be a spreadsheet indicating for a list of chemicals which of the databases outlined in **Table 2. Specifics for flagging availability of toxicity data** have data for each chemical. A "1" will indicate the presence of data, and a "0" will indicate absence of data. Rows of the resulting spreadsheet which contain "0" for all databases in **Table 2. Specifics for flagging availability of toxicity data** will have their chemical names highlighted.

Note: using formulas across different Excel files requires working in a **local copy** of the file. Formulas across different Excel files will not work in the web browser interface for OneDrive/Excel. Files on OneDrive must be either 1) downloaded to work in a local copy, or 2)

accessed via OneDrive's integration with File Explorer.

2.3.1 Generic Approach

To create an Excel column with an indicator of whether a second spreadsheet has data for a CAS:

1. Open the spreadsheet where you are creating the column (spreadsheet A).
2. Open the spreadsheet you are summarizing (spreadsheet B). Leave this spreadsheet open until after step 7.

Complete the action(s) in the first column, then type the text from the second column. (Do not hit enter until you have completed the tasks shown in all rows in Table 1.)

Table 1. Constructing a formula to test data availability

Row	Action	Text to type
1	On spreadsheet A, click in the first non-header cell of the column you are creating	=IF(COUNTIFS(
2	Toggle to spreadsheet B; highlight the CASRN column; toggle to spreadsheet A	,
3	Click on the first cell of data in the CASRN column (should be in the same row as the cell you're working in for the new column; still in spreadsheet A)	,
4	Toggle to spreadsheet B; highlight the column which has the data; toggle to spreadsheet A	,"<>")>0,1,IF(COUNTIF(
5	Toggle to spreadsheet B; highlight the CASRN column; toggle to spreadsheet A	,
6	Click on the first cell of data in the CASRN column)>0,0,""))

3. Hit enter.
4. Hover over the bottom right corner of the cell which has been filled until you see a +, click, and drag to the bottom of the column to fill other cells.
5. Copy entire column.
6. Right click in heading row of same column and choose Paste Values. (At this point, you can close spreadsheet B.)
7. Your column now has 1's for CASRNs which have data in the column of interest, 0's for CASRNs which are on the spreadsheet but do not have data in that column, and blanks for CASRNs which are not on the second spreadsheet. We used this process to denote whether our QSAR downloads had endpoint path data available.

These steps can also be easily modified for slight differences in task. Following are two examples, along with the necessary changes to the process.

1. If you want to check for different criteria, you can replace "<>" in the row 4 of Table 1. For example, Table 2, row 3 refers to a column indicating whether the CompTox

spreadsheet contains toxicity data other than the results of QSAR models. To create this column, the entire process is the same except that the criterion becomes "<>QSAR Model".

2. If all data are in the same spreadsheet, the process is identical but without the need to toggle between the two spreadsheets.

An example of a completed formula checking for data other than the results of QSAR models in the CompTox download provided by EPA (row 4 in Table 2), for the Polyhalogenated Phenol Aliphatic Ethers class:

```
=IF(COUNTIFS('[Polyhalogenated Phenol Aliphatic Ethers and Analogs-with_HCD_output.xlsx]Hazard Records'!$B:$B,B4,'[Polyhalogenated Phenol Aliphatic Ethers and Analogs-with_HCD_output.xlsx]Hazard Records'!$F:$F,"<>QSAR Model")>0,1,IF(COUNTIF('[Polyhalogenated Phenol Aliphatic Ethers and Analogs-with_HCD_output.xlsx]Hazard Records'!$B:$B,B4)>0,0,""))
```

2.3.2 Application for identifying availability of toxicity data

The text in this section assumes that two downloads from the Comptox Dashboard are available – the initial bulk download described in Section 2.2, and the “backdoor” download provided by EPA and described in Section 3.1 (file name: OFRs_EPA Comptox_DASHBOARD....). Table 2 identifies the respective spreadsheets and/or columns for each column collating the availability of toxicity data. Spreadsheet A could be either the master spreadsheet created in the initial Comptox Dashboard bulk download, or a separate spreadsheet could be created with a subset of the data.

Table 2. Specifics for flagging availability of toxicity data

Row	Spreadsheet A column name	Spreadsheet B and column name
1	Toxicity data in PubMed articles?	CompTox Dashboard download, NUMBER_OF_PUBMED_ARTICLES (Note: this column in the Dashboard download includes dashes for chemicals without any PubMed articles. These dashes need to be deleted before determining data availability, since the dashes will be counted as data.)
2	Toxicity data in PubChem	CompTox Dashboard download, PUBCHEM_DATA_SOURCES (Note: this column in the Dashboard download includes dashes for chemicals without any PubMed articles. These dashes need to be deleted before determining data

		availability, since the dashes will be counted as data)
3	Bio Test or Chem Gene data in PubChem?	N/A. Populated manually based on whether relevant data downloaded from PubChem or CTD
4	Non-QSAR Toxicity Data on Comptox?	Comptox download provided by EPA, List Type For first two classes, these are the files: OFRs_EPA Comptox_DASHBOARD hazard-2021-02-20T00_49_49 and All OFR Analogs_Dashboard_hazard-2021-02-19T05_18_15 For the remaining 12 classes, this is in the “Hazard Records” tab of the files named with the format [CLASS NAME] and Analogs-with HCD output
5	HH Data on OECD toolbox (HH stands for Human Health)	OECD toolbox download, Endpoint Note there is a column called EndpointPath, and a column called Endpoint. We are using Endpoint, NOT EndpointPath.

The result of applying the steps in 2.3.1 for each of the new columns described in Table 2 will be a series of four columns, each indicating with 1’s and 0’s whether toxicity data are available for that chemical from a specified source.

2.3.3 Flagging chemicals lacking data

The spreadsheet columns developed in Section 2.3.2 can now be used to identify with highlighting those chemicals that do not have data in any of the sources listed in Table 2. A benefit of using conditional formatting as described here is that the highlighting will automatically update as data are changed.

1. Click on the first cell of data in the column you want to highlight. (Here, highlight the chemical name column.)
2. Click Conditional Formatting on the toolbar.
3. Choose Highlight Cells Rules.
4. Choose More Rules.
5. Choose Use a formula to determine which cells to format.
6. Under Edit the Rule Description, in the Format values where this formula is true box, type =SUM(G2,H2,J2,K2)=0, where the cells listed inside the sum function are the ones that should all be 0. (Here, the four cells correspond to the cells noted under “spreadsheet column name” in Table 2.)
7. Click format.

8. Click Fill.
9. Choose your color.
10. Click OK.
11. Click OK.
12. To apply the format to other cells in the column: First click Conditional Formatting again.
13. Choose Manage rules.
14. Select the rule you just created, and choose Edit rule.
15. Delete all dollar signs in the formula.
16. Click OK.
17. Click OK.
18. Right-click on the formatted cell and choose Copy.
19. Highlight the other cells in the column, right-click, and choose Paste Formatting.

Follow this approach for both the OFRs and OFR analog tabs. The result will be that the chemical name is highlighted for all chemicals where no toxicity data were located in any of the four sources. No further work is needed on the highlighted analogs. For the highlighted OFRs, it is not necessary to search PubMed for those OFRs shown as 0 PubMed hits. (See also Section 5.1.)

3. Searching for Database Data

The CPSC staff, along with the UC team, have identified several compilations of database data, as well as publications of additional data. CPSC staff has identified a comprehensive set of data sources for this task, including chemical, toxicity, and exposure databases. The UC team reviewed these databases to identify the areas of overlap, and the most effective approach to capture the landscape of available data while minimizing overlap, noting that many of the secondary or tertiary databases reference the same primary data sources. Based on this review, the final approach of database data to download and grey literature searching was developed. Note that, for all database downloads, it is important to also download and track associated meta data, including the date of the download, version of the database (if available), file names or format of file names, and the location of the downloaded files. Other meta-data that will be collected if available: any information on how the database was created, information on individual sources, (e.g., source of data, quality control/assurance procedures), and information about how often the database is updated. Meta-data will be tracked in a file titled “!Database tracking sheet-[CLASS NAME]+Analog.”

3.1 Comptox Dashboard and ToxVal Dashboard Hazard Data

Data (listed in Section 2.2) from the publicly accessible interface of the EPA CompTox Dashboard were downloaded according to the process outlined in Section 2.2.1, using CAS as an identifier.

ToxVal is now available from the EPA CompTox Dashboard, and can be downloaded according to the process outlined in Section 2.2.1. In the original literature survey, this information from the TOXVAL Version 5 databases has been provided by the EPA directly in a master spreadsheet for all of the OFRs on the CPSC list, including information on the type of hazard (e.g., developmental, eye irritation, etc.), CAS, chemical name, database source, and “original source” (which may be the name of another database source, a publication, or blank), list type (authoritative, screening, or QSAR model), and some categorical information on the quantitative result. [File names are of the format: OFRs_EPA Comptox_DASHBOARD hazard- DATE and All OFR Analogs_Dashboard_hazard- DATE.]

3.2 OECD QSAR Toolbox

Human health hazard information is to be gathered from the OECD QSAR Toolbox. For each class, version number and date of most recent update should be documented upon data download. As of June 22, 2021, the Toolbox was last updated 4/15/2020. Information on updates can be found here: <https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm>. The directions provided here and screenshots are based on Version 4.4 of the Toolbox. The Toolbox continues to be updated, with version 4.5 released in September, 2021.

Create an input file in the correct file format (Figure 4). Note the column header names in the example – variations on these names (e.g., CASRN instead of CAS) may not work. This input file should include the CAS number, chemical name, and SMILES – **these should be copied and pasted to avoid errors**. Save this file as a “Text (Tab delimited)” file. (Note: as of 2/24/2021, the Toolbox input directions imply that the file needs to have a .smi or .sdf format and extension. This is not correct.)

	A	B	C
1	CAS	Chemical	SMILES
2	78-43-3	Tris(2,3-dichloropropyl)phosphate	ClCC(Cl)COP(=O)(OCC(Cl)CCl)OCC(Cl)CCl
3	115-96-8	Tris(2-chloroethyl) phosphate	ClCCOP(=O)(OCCCl)OCCCl
4	115-98-0	Bis(2-chloroethyl) vinylphosphonate	ClCCOP(=O)(OCCCl)C=C
5	126-72-7	Tris(2,3-dibromopropyl) phosphate	BrCC(Br)COP(=O)(OCC(Br)CBr)OCC(Br)CBr
6	140-08-9	Tris(2-chloroethyl) phosphite	ClCCOP(OCCCl)OCCCl
7	4351-70-6	Phosphonic acid, P-[1-[[[2-chloroethoxy](2-chloroethyl)phosphinyl]oxy]ethyl]-, 1-[bis(2-chloroethoxy)phosphinyl]ethyl 2-chloroethyl ester	CC(OP(=O)(OCCCl)C(C)OP(=O)(CCCl)OCCCl)P(=O)(OCCCl)OCCCl
8	5412-25-9	Bis(2,3-dibromopropyl) hydrogen phosphate	OP(=O)(OCC(Br)CBr)OCC(Br)CBr
9	6145-73-9	Tris(2-chloropropyl) phosphate	CC(Cl)COP(=O)(OCC(C)Cl)OCC(C)Cl
10	6294-34-4	Bis(2-chloroethyl) 2-chloroethylphosphonate	ClCCOP(=O)(CCCl)OCCCl

Figure 4. Example of input file format for OECD QSAR Toolbox search

1. Open the QSAR Toolbox Program. Select “Input,” and under the “Chemical List” bar, select “List” (Figure 5). A dialogue box will open. Navigate to the .txt file with the list of chemicals for which you want to perform the batch search and double-click on it. (Note:

the .txt file needs to be closed in order to be uploaded to the Toolbox.) You will get a dialogue box stating the number of structures successfully imported.

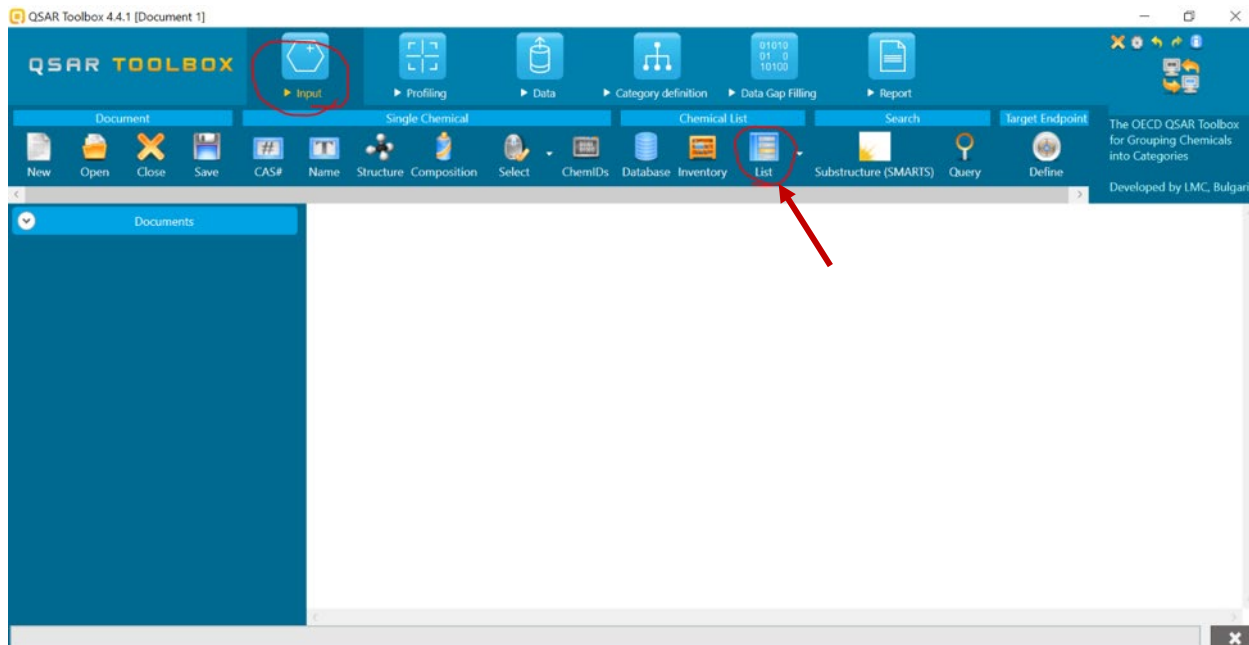


Figure 5. QSAR Toolbox List Input screen

2. After your file uploads, the resulting screen should look something like Figure 6 (after expanding the Structure info by clicking on the plus sign):

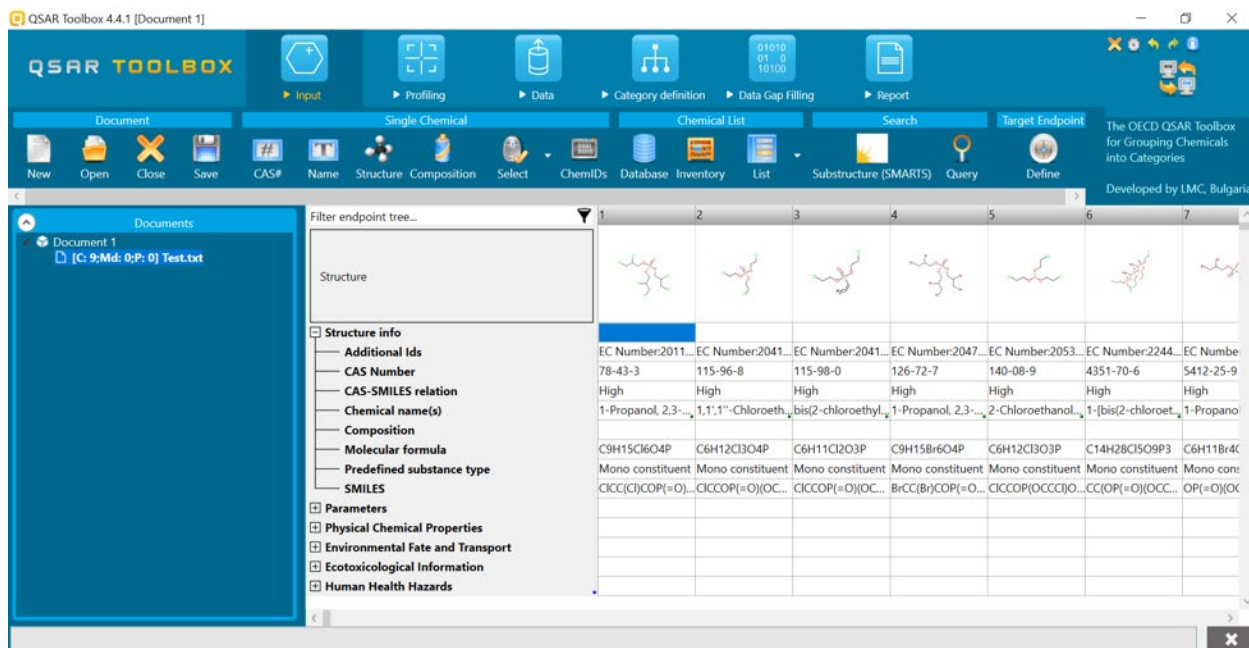


Figure 6. QSAR Toolbox screen after data is inputted

- Then, select the “Data” tab (Figure 7)

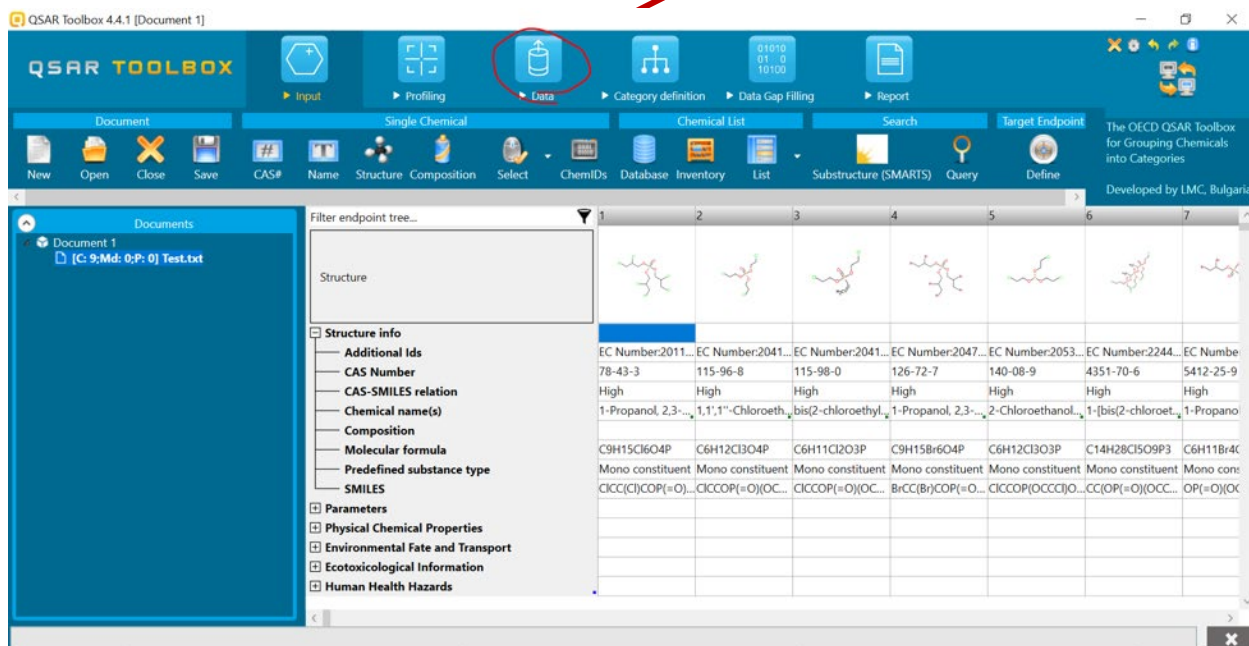


Figure 7. QSAR Toolbox Data Tab

- On the left side of the screen there is a column with “Documents,” “Databases,” and “Inventories” tabs. (You can minimize the documents and inventories tabs and expand databases.) Scroll down the list under the “Databases” tab until you see “Human Health Hazards.” Select the checkbox next to human health hazards; this will consequently select everything under it, as well as the same databases if they appear under other headings. Click “Gather” at the top row under the data section (Figure 8).

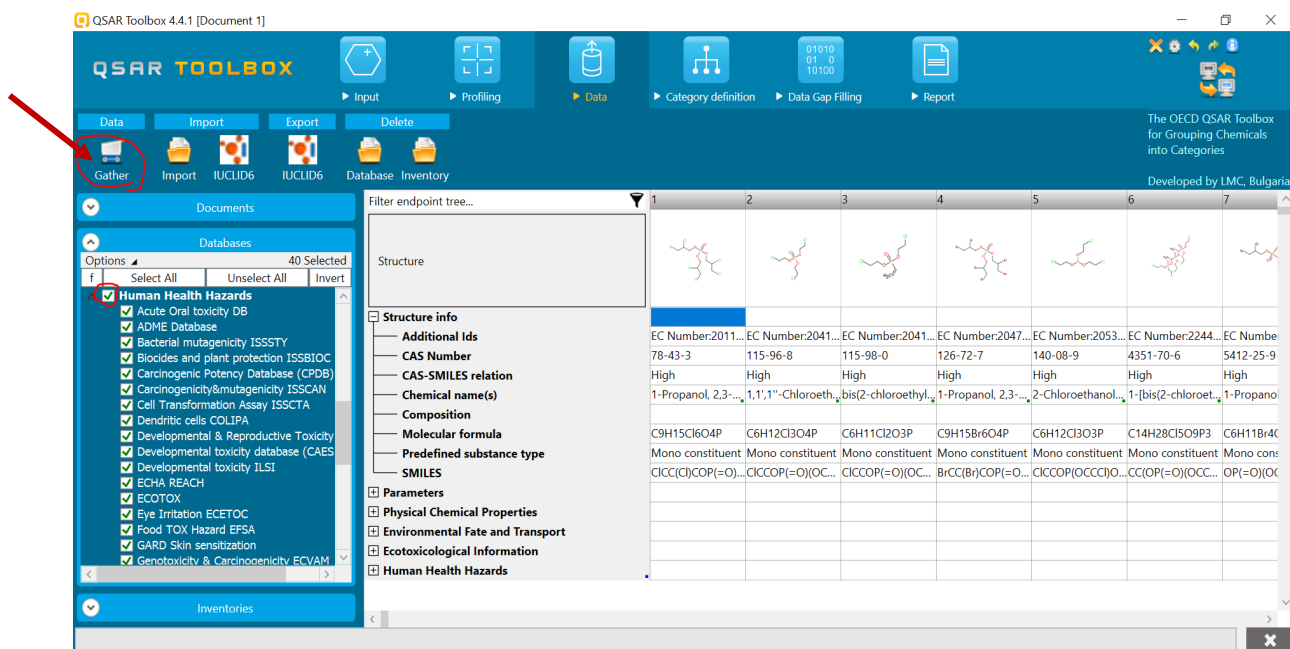


Figure 8. Example of selecting databases for the search

5. A pop-up message will appear. Choose “all endpoints” and click ok (Figure 9). The program will take several moments as it gathers data from the various databases.

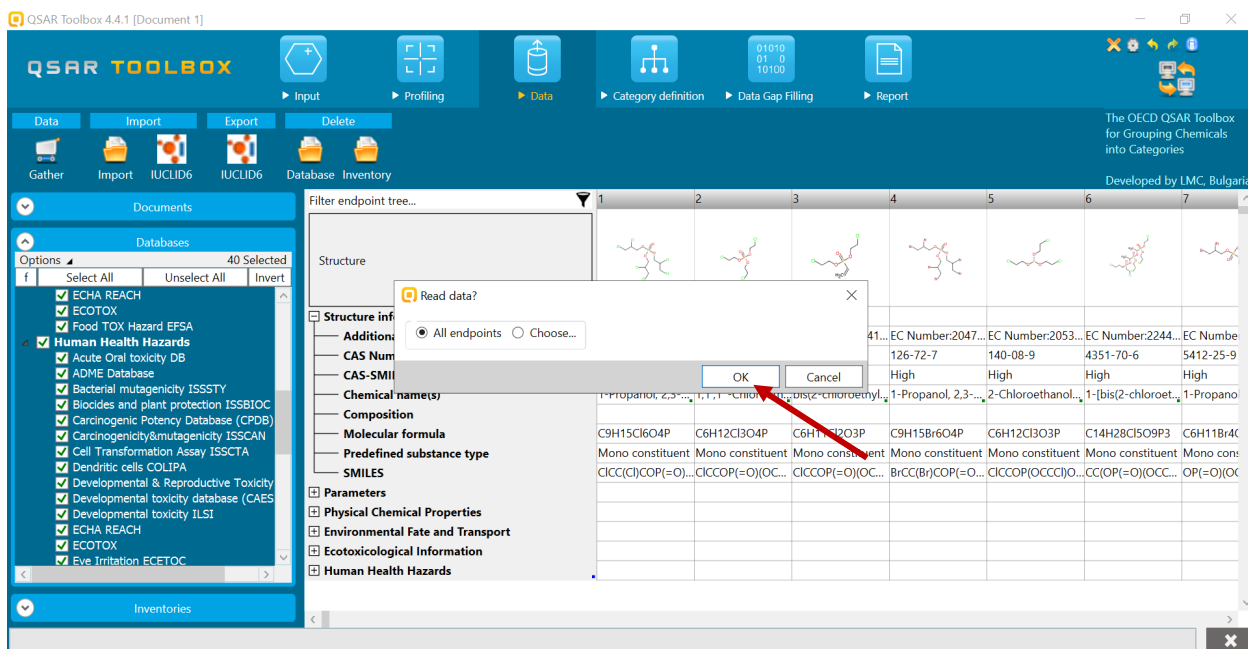


Figure 9. Selecting endpoints for the search

6. Once the data have been gathered, the rows for human health (and potentially other categories) will be populated with the number of records identified. A small green square in the lower right of a cell indicates that there are more records for that category for that chemical. Categories with a plus sign can be expanded to reveal additional levels of data.

7. To download the results, click on the “Report” tab and select “Data Matrix” under the “Exports” bar. A popup box will allow selection of different subsets of data. Click on “Human Health Hazards” and “Export detailed metadata,” and then click on “Export” (Figure 10). The file will be saved in your active folder.
 - a. **Note:** It is worth expanding the “Human Health Hazards” category in the “Matrix Export” window (Figure 10. *Example of downloading data from the QSAR Toolbox*) to make sure all sub-categories are selected. If human health category is already selected when this window pops up, sometimes not all sub-categories are selected.

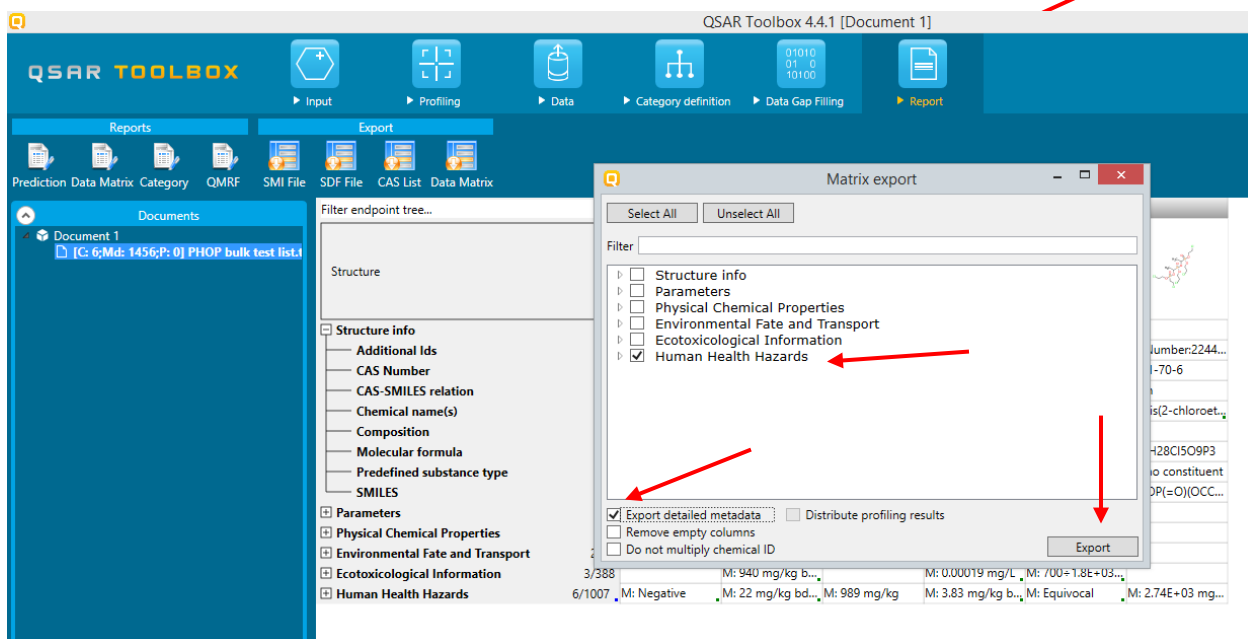


Figure 10. Example of downloading data from the QSAR Toolbox

8. The resulting file will be an excel spreadsheet with data from the QSAR Toolbox batch search. Document relevant meta data, including date of download and Toolbox version used. An example screenshot from one option for the downloading is shown below (Figure 11).

#	CAS Num	SMILES	Endpoint	Database	Year	Endpoint	Record ID	Type of m	Reference	Route of a	Test orgar	Assigned	Qualifier	Duration.I	Duration.I	Duration.I	Duration.I
1	1 115-96-8	CICCO	(= Human He Acute Ora		2012	LD50	4268	in Vivo	http://che	Oral (Gav	Rat	FALSE	No qualifi	24		h	Time
2	1 115-96-8	CICCO	(= Human He Carcinogeni	Potenc		Summary carcinog		in Vivo	TR 391; fir	Gavage	Mouse	FALSE					
3	1 115-96-8	CICCO	(= Human He Carcinoge		16.7.2011	Summary carcinogeni			NTP391	Gavage	Mouse	FALSE					
4	1 115-96-8	CICCO	(= Human He Carcinoge		16.7.2011	Summary carcinogeni			NTP391	Gavage	Mouse	FALSE					
5	1 115-96-8	CICCO	(= Human He Carcinogeni	Potenc		TD50		in Vivo	TR 391; fir	Gavage	Mouse	FALSE	No qualifier				
6	1 115-96-8	CICCO	(= Human He Carcinogeni	Potenc		Summary carcinog		in Vivo	TR 391; fir	Gavage	Rat	FALSE					
7	1 115-96-8	CICCO	(= Human He Carcinoge		16.7.2011	Summary carcinogeni			NTP391	Gavage	Rat	FALSE					
8	1 115-96-8	CICCO	(= Human He Carcinoge		16.7.2011	Summary carcinogeni			NTP391	Gavage	Rat	FALSE					
9	1 115-96-8	CICCO	(= Human He Carcinoge		16.7.2011	Summary carcinogeni			NTP391	Gavage	Rat	FALSE					
10	1 115-96-8	CICCO	(= Human He Carcinogeni	Potenc		TD50		in Vivo	TR 391; fir	Gavage	Rat	FALSE	No qualifier				
11	1 115-96-8	CICCO	(= Human He Carcinoge		16.7.2011	TD50			NTP391		Rat	FALSE	No qualifier				
12	1 115-96-8	CICCO	(= Human He Genotoxicity	OASIS		Gene mut	10177	in Vitro	Mutation Research	5	Salmonell	FALSE					
13	1 115-96-8	CICCO	(= Human He Bacteri	r	02.5.2011	Gene mut	9050	in Vitro	CCRIS database,Toxr		Salmonell	FALSE					
14	1 115-96-8	CICCO	(= Human He Genotoxicity	OASIS		Gene mut	10174	in Vitro			Salmonell	FALSE					
15	1 115-96-8	CICCO	(= Human He Bacteri	r	02.5.2011	Gene mut	9051	in Vitro	CCRIS database,Toxr		Salmonell	FALSE					
16	1 115-96-8	CICCO	(= Human He Genotoxicity	OASIS		Gene mut	10175	in Vitro			Salmonell	FALSE					

Figure 11. Partial example of spreadsheet of data downloaded from OECD QSAR Toolbox

- If no results are returned for an entire class of chemicals (or the analogs for that class), QSAR Toolbox will return a warning message. Capture a screenshot of that warning message for documentation (see Figure 12 for an example).

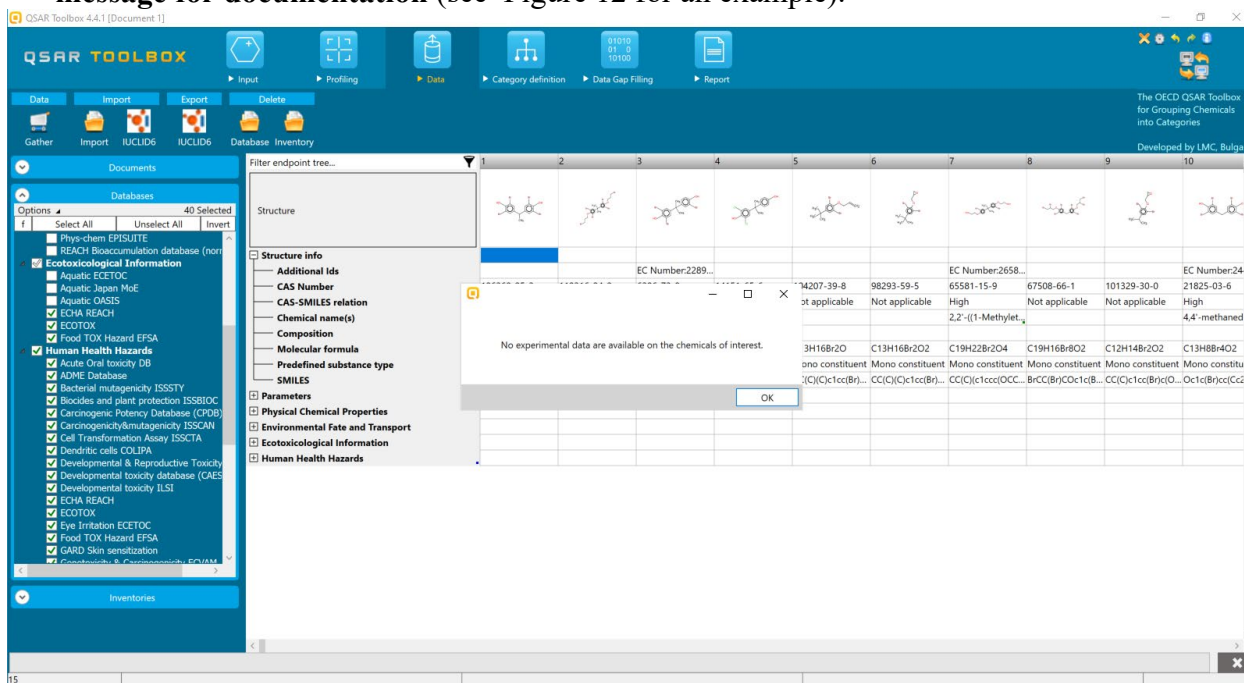


Figure 12. Example of no chemicals returning data from OECD QSAR Toolbox

- Name the file as follows: OECD QSAR CLASS NAME New Analogs_DATE.xlsx

3.3 ToxCast and Tox21 (ICE)

ToxCast (Toxicity Forecaster) is an EPA initiative in which approximately 1800 chemicals were screened in more than 700 high-throughput assays. Tox21 (Toxicology in the 21st Century) is a collaboration among EPA, NIH (including the NTP and the National Center for Advancing Translational Sciences) and the FDA. The Tox21 team has tested about 10,000 chemicals in about 70 high-throughput assays. The ToxCast and Tox21 data can be downloaded from several different platforms. We chose the NTP Integrated Chemical Environment (ICE, also referred to

as WebICE) because it allows for simultaneous bulk downloads of chemicals from both testing programs. Data for PHOPs was downloaded via WebICE Version 3.3. In June 2021, WebICE was updated to version 3.4, and data for PHBAFs were downloaded under this version. For each future class, version number and date of most recent update will be documented upon data download. Information on version updates can be found by following this link (<https://ice.ntp.niehs.nih.gov/Updates>) and clicking on the “Updates” tab along the left side. ICE implements additional curation of the curated high-throughput screening (cHTS) data beyond that done as part of the federal collaboration. More details of this curation process are provided at the following link: <https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION?section=cHTS>. Detailed

Note that in addition to data retrieved from ICE through the following process, ICE data were also downloaded as entire datasets grouped by endpoint. Such data are available as Excel spreadsheet downloads here: <https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION>

The same process is used for a single chemical search and for a batch search.

1. Navigate to <https://ice.ntp.niehs.nih.gov/Search>
2. In the Chemical Input section, chemicals can be searched by CAS number.
3. To input a list of chemicals, add one CAS number per line in the “User Chemical Identifiers” box. For example, this can be done by copying CAS numbers from a column in an Excel spreadsheet and pasting them in the “User CASRNs” box (Figure 13). **Note** that CAS numbers must be formatted with dashes for the CAS number search to return query results.

Figure 13. Example of user-supplied CAS numbers

4. In the Assay Input section, click “Select Assays” to select specific assays for download

- Select “cHTS”, which includes information from ToxCast. All subcategories should now also be checked. Click “Finished” (Figure 14)

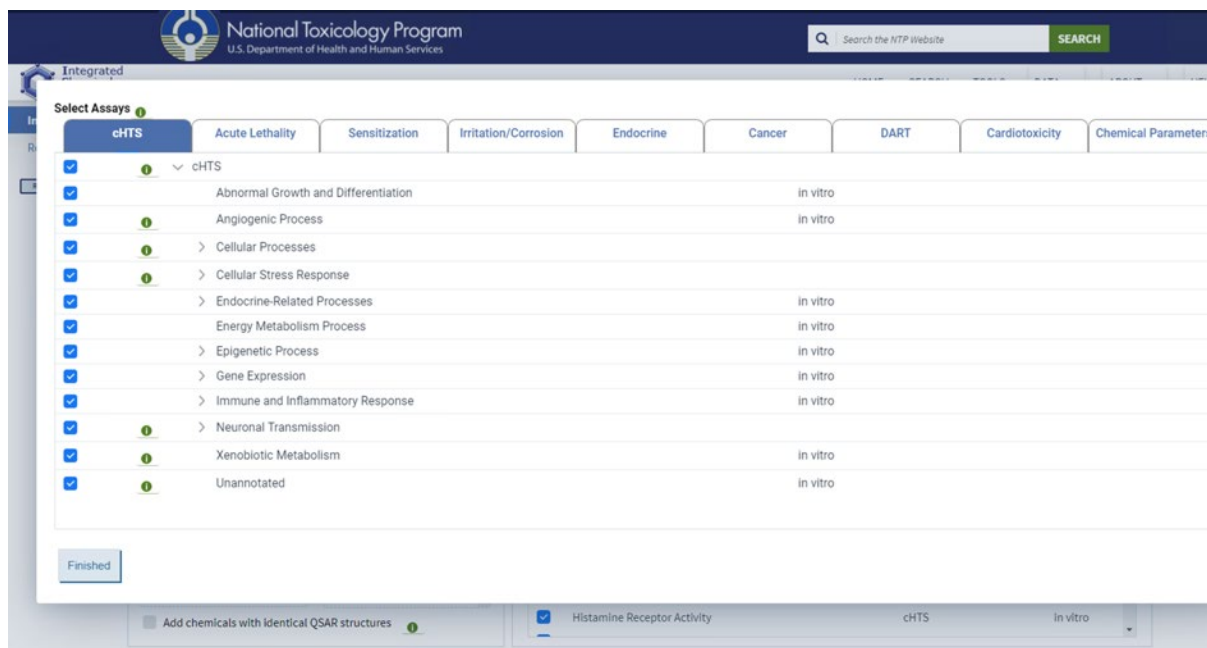


Figure 14. Example of cHTS assays selected, with all subcategories included

Click “Run”, and the results will load (Figure 15)

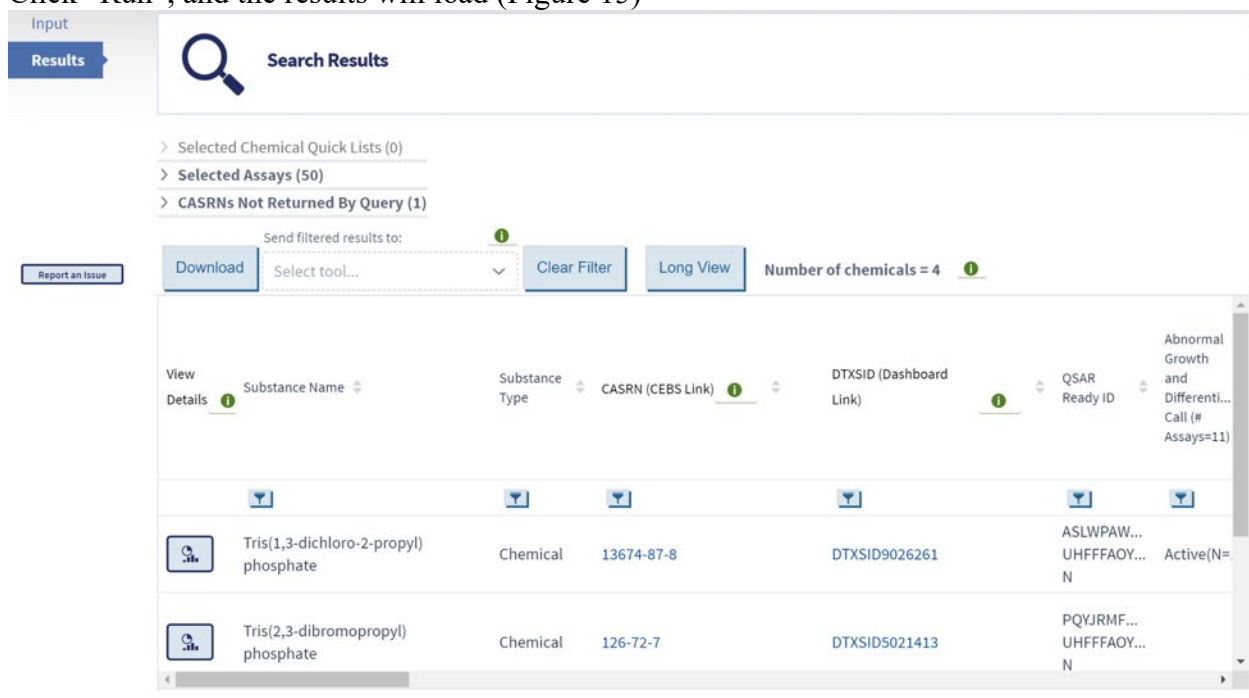


Figure 15. Example of search results page on WebICE

5. The data can be sorted using the arrows next to each column name. **Note** that the sorting applied to the displayed results will also be applied to any downloaded data.

The data can be downloaded as is by using the “Download” button (**Figure 15**). Download options include summary data, Wide Format data, and long format data. Each can be downloaded as either an Excel or Txt file. Figure 16 illustrates these choices. As noted in the figure, the “Wide Format,” cHTS assay data include AC50 values (half-maximal activity concentration) only for assays with an "Active" chemical. For assays with no active chemicals in the download, only the overall “call” is provided. The processing of screening data collected at multiple concentration is complex; therefore, the output contains several observations and statistical parameters, including the hit call determination, AC10 (activity concentration that results in a 10% response), AC50, and ACC (activity concentration at cut-off, a measure of the activity threshold for an assay response based on curve-fitting models).

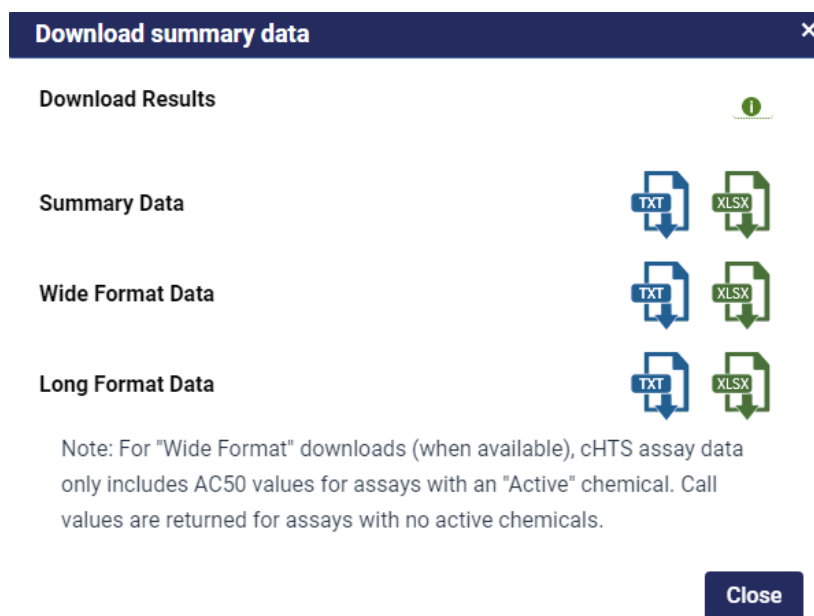


Figure 16. Data formats for downloading results on WebICE

6. Download each format as an Excel file.
 - Summary Data (Figure 17) – provides information for each chemical on a single row. If more than one data point is present for an endpoint, the number of data points, median, and range are provided. cHTS data are provided as group summaries.

	A	B	C	D	E	F	G	H	I	J
1	Substance	CASRN	DTXSID	Qsar Ready	Coagulation	Inflammation	Immune System	Process	Call	(# Assays=1)
2	Terephthal	100-21-0	DTXSID602	KKEYFWRC		Inactive(N=				
3	4-Nitroben	100-11-8	DTXSID704	VOLRSQPSJ		Inactive(N=				
4	4-Nitroben	100-14-1	DTXSID402	KGCNHWX		QC-Omit(N=				
5	1-Chloro-4	100-00-5	DTXSID502	CZGCEKJOL	Inactive(N=	Inactive(N=				
6	1,4-Diisopr	100-18-5	DTXSID202	SPPWGCYE	Active(N=5	Active(N=2				
7	N-Methyl-4	100-15-2	DTXSID702	XIFJZJPMH		Inactive(N=				
8	4-Nitrophe	100-02-7	DTXSID002	BTJIUGUIPI	Inactive(N=	Inactive(N=				
9	4-Nitroacet	100-19-6	DTXSID102	YQYGPCKTI		Inactive(N=				
10	4-Methoxy	100-06-1	DTXSID204	NTPLXRHD		Inactive(N=				
11	4-Dimethyl	100-10-7	DTXSID502	BGNGWHS		Inactive(N=				
12	4-Nitroanil	100-01-6	DTXSID802	TYMLOMAK	Inactive(N=	Inactive(N=	Inactive(N=1)			
13	1,4-Benzen	100-20-9	DTXSID702	LXEJRKJRKI		Inactive(N=				
14										
15										

Figure 17. Example of downloaded data in “Summary Data” format from WebICE

Wide Format Data (Figure 18 and Figure 19) – this download is similar to Summary Data view, with summary calls and statistics for AC50 and ACC values provided for each cHTS assay included in the query. This download option is only provided for queries that include cHTS assays.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	Substance	CASRN	DTXSID	Qsar Ready	Coagulation	Inflammation	Immune Sy	ATG_NF_KB_CIS_up	ATG_NRF2_ARE_CIS_up	ATG_STAT3_CIS_up	BSK_3C_Eselectin_down	BSK_3C_Ev	BSK_3C_HL	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6
2	Terephthal	100-21-0	DTXSID602	KKEYFWRC	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
3	4-Nitroben	100-11-8	DTXSID704	VOLRSQPSJ	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
4	4-Nitroben	100-14-1	DTXSID402	KGCNHWX	QC-Omit(N=	QC-Omit(N=	QC-Omit(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
5	1-Chloro-4	100-00-5	DTXSID502	CZGCEKJOL	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
6	1,4-Diisopr	100-18-5	DTXSID202	SPPWGCYE	Active(N=5	Active(N=2	Active(N=2	136.151	Inactive(N=1)	Inactive(N=1)														
7	N-Methyl-4	100-15-2	DTXSID702	XIFJZJPMH	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
8	4-Nitrophe	100-02-7	DTXSID002	BTJIUGUIPI	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
9	4-Nitroacet	100-19-6	DTXSID102	YQYGPCKTI	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
10	4-Methoxy	100-06-1	DTXSID204	NTPLXRHD	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
11	4-Dimethyl	100-10-7	DTXSID502	BGNGWHS	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
12	4-Nitroanil	100-01-6	DTXSID802	TYMLOMAK	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
13	1,4-Benzen	100-20-9	DTXSID702	LXEJRKJRKI	Inactive(N=	Inactive(N=	Inactive(N=	Inactive(N=1)	Inactive(N=1)	Inactive(N=1)														
14																								
15																								

Figure 18. Example of downloaded data in “Wide Format Data” from WebICE

Number only tab:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	Substance	CASRN	DTXSID	Qsar Ready	ATG_NF_KB	ATG_NRF2	ATG_STAT3	BSK_3C_Ev	BSK_3C_Ev	BSK_3C_HL	BSK_3C_HL	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6	BSK_3C_IL6
2	Terephthal	100-21-0	DTXSID602	KKEYFWRC																							
3	4-Nitroben	100-11-8	DTXSID704	VOLRSQPSJ																							
4	4-Nitroben	100-14-1	DTXSID402	KGCNHWX																							
5	1-Chloro-4	100-00-5	DTXSID502	CZGCEKJOL																							
6	1,4-Diisopr	100-18-5	DTXSID202	SPPWGCYE		136.151		40		40		40															
7	N-Methyl-4	100-15-2	DTXSID702	XIFJZJPMH																							
8	4-Nitrophe	100-02-7	DTXSID002	BTJIUGUIPI																							
9	4-Nitroacet	100-19-6	DTXSID102	YQYGPCKTI																							
10	4-Methoxy	100-06-1	DTXSID204	NTPLXRHD																							
11	4-Dimethyl	100-10-7	DTXSID502	BGNGWHS																							
12	4-Nitroanil	100-01-6	DTXSID802	TYMLOMAK																							
13	1,4-Benzen	100-20-9	DTXSID702	LXEJRKJRKI																							
14																											
15																											

Figure 19. Example of downloaded data in “Wide Format Data” from WebICE with number only tab

- Long Format Data (Figure 20) – returns each unique chemical-assay combination as a single row. This format is more useful for running additional analyses on

Search results, and the database logic described in Section 4 was applied to the Long Format Data. See below:

Record ID	Chemical Name	CASRN	DTXSID	QSAR Ready If Assay	Endpoint	Value	Unit	Species	Receptor Species	Route
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_4H_VC	AC50			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	AC50			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	AC50			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	AC50			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	AC50			40	uM			
R_117930; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA ATG_NRF2_	AC50			136.1514	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; AC50				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	AC50			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; AC50				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; AC50				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; AC50				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; AC50				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_hDFCC	AC50			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; AC50				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_LPS_E; Top of curv				0.2055	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_LPS_E; AC50				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_LPS_IL; Top of curv				0.1919	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_LPS_IL; AC50				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_3C_Es; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_hDFCC; Top of curv				0.5189	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_3C_HL; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_3C_IC; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; Top of curv				0.5305	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_3C_M; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; Top of curv				0.9523	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; Top of curv				0.3711	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; Top of curv				0.1379	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_3C_VC; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; Top of curv				0.4972	Log10 fold			
R_117930; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA ATG_NRF2_	Top of curv			1.8905	Log2 fold ir			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_4H_Eo; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; Top of curv				0.3698	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_CASM; Top of curv				0.2097	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_4H_M; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	Top of curv			0.2257	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	Top of curv			0.2562	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	Top of curv			0.1508	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_3C_Ti; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	Top of curv			0.2868	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_3C_uP; ACC				40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	ACC			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	Top of curv			0.3127	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	ACC			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_4H_uF; Top of curv				0.4351	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_BE3C_	ACC			40	uM			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_4H_VC; Top of curv				1.2063	Log10 fold			
R_118079; 1,4-Diisopropy	100-18-5	DTXSID202 SPPWGCYEYA BSK_4H_Ps; ACC				40	uM			

Figure 20. Example of downloaded data in “Long Format Data” from WebICE

- Rename each format from the default file name to the following shorter names:
 - ICE - ChemicalSearch_Summary data.xlsx
 - ICE - ChemicalSearch_Wide Format Data.xlsx
 - ICE - ChemicalSearch_Long Format Data.xlsx

3.4 Using PubChem to Identify Available Toxicity Data

PubChem can provide information on the availability of toxicity data beyond that in PubMed and WebICE. (See Sections 2 and 5 for additional information about PubMed; see above for additional information on WebICE). Obtaining this information in batch form requires first

obtaining the PubChem Chemical ID (CID), and then using the CID to search for available data. For each class, version number and date of most recent update should be documented upon data download. Information on updates and version number can be found here:

<https://pubchemblog.ncbi.nlm.nih.gov/> or <https://pubchemdocs.ncbi.nlm.nih.gov/news>

3.4.1 Using PubChem Identifier Exchange Service

The PubChem Identifier Exchange Service allows conversion between different chemical identifiers, which is useful as different databases may use different identifiers. More information about the PubChem Identifier Exchange Service is available

at <https://pubchemdocs.ncbi.nlm.nih.gov/identifier-exchange-service>

1. To create an input file for uploading to PubChem Identifier Exchange Service:
 1. Enter SMILES into a separate excel file, with each SMILES in its own cell in the first column
 - a. Note that although the example presented here converts SMILES into CIDs, other conversions are possible. The input file may use Registry IDs, SIDs, CIDs, InChIKey, and SMILES. These identifiers can be separated by white space, comma, tab, or carriage return.
 2. Next, save the excel spreadsheet as a Text File. To do so, you can click “File” → “save as” → “Text (tab delimited)”
2. Navigate to <https://pubchem.ncbi.nlm.nih.gov/idexchange/idexchange.cgi>
 1. Select “SMILES” for the Input ID List to be converted into CIDs for the Output IDs. Use “Choose File” to choose the text file of SMILES to be converted into

CIDs. Ensure that Output IDs has CIDs selected, and that “No Compression” is selected under “Compression” (Figure 21).

The screenshot shows the PubChem Identifier Exchange Service input screen. The browser address bar displays the URL: <https://pubchem.ncbi.nlm.nih.gov/idxchange/idxchange.cgi?inputtype=cid&inputdsn=&inputdsnauto=&operatortype=samecid&outputtype=cid&outputdsn=&output...>. The page title is "PubChem Identifier Exchange Service". The form includes the following sections:

- Input ID List**: Includes a dropdown menu set to "SMILES", a text input field, and a radio button. A red arrow points to the "SMILES" dropdown.
- Choose File**: A button labeled "Choose File" with the text "SMILES for Pubchem.txt" next to it. A red arrow points to this button.
- Operator Type**: A dropdown menu set to "Same CID".
- Output IDs**: A dropdown menu set to "CIDs". A red arrow points to this dropdown.
- Output Method**: A dropdown menu set to "Two column file showing each input-output correspondence".
- Compression**: A dropdown menu set to "No compression". A red arrow points to this dropdown.

Figure 21. Example of PubChem Identifier Exchange Service input screen

2. Click “submit job” at the bottom of the page. A text file will be produced with your input SMILES and the output CIDs. To save the text file:
 - a. Highlight and copy the contents of the text file, and paste into Word (or a basic text editor such as Notepad).
 - b. In Word, click File > Save As, and name your file.
 - c. To open the .txt file in Excel, open Excel. In Excel, click File > Open > Browse and navigate to the location where you saved your .txt file. Select your file and click Open.
 1. If you cannot see your file, ensure that the drop-down box to the right of “File Name” says “All Files”
 - d. Excel will open its Text Import Wizard. Ensure that “Delimited” is selected on the first screen, then click Next.
 - e. The Data Preview at the bottom of the window should show two columns of data. Click Next.
 - f. Click Finish, and your .txt file should be open as a spreadsheet in Excel, with two columns: one containing the SMILES notation and one containing the corresponding CIDs.
 1. The CID column can be copied to a new excel file as a list of CIDs for upload into PubChem (see section 3.4.2). Note that this method is no longer used, but is retained as a record.

- Alternatively, the CIDs can be copied into Word and used to create a single URL for bulk download of PubChem bioassay data (see section 3.4.3)

g. Proceed to Section 3.4.3 (skip section 3.4.2)

3.4.2 Using Upload ID List function in PubChem and Downloading Data

Note, this is one method of getting this information. **This method is no longer used**, but this section is retained as a record and to outline an alternate method. **Instead, use CIDs generated in Section 3.4.1 to bulk download bioassay information in Section 3.4.3, and use CASRNs to search CTD for chemical-gene interaction data and other related data, as described in Section 3.4.4. Analytic Laboratory Methods are obtained from market use task.**

- In order to use the “upload ID list” function within PubChem, first create a text file that lists CIDs for each chemical, as described in the previous section. If the CID list was initially created with the corresponding SMILES notations, a new file will need to be saved containing only the CIDs. To create a *de novo* CID list, enter CIDs in the first column of an excel spreadsheet, with each CID in its own cell.
- To save this spreadsheet (either the one created in Section 3.4.1 or a *de novo* file with CIDs) as a text file, click “file” → “save as” → choose the destination → under “save as type,” choose “text (tab delimited).”
- Go to PubChem homepage at <https://pubchem.ncbi.nlm.nih.gov/> and click on “upload ID list” at the bottom (Figure 22).

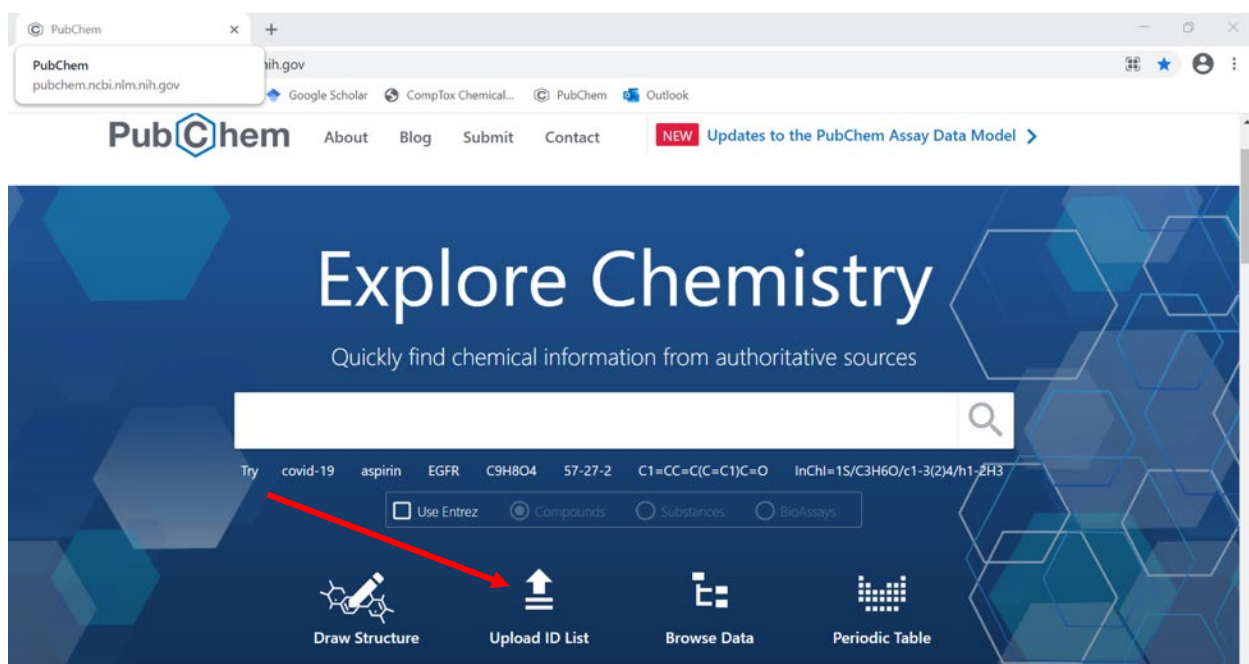


Figure 22. Example of PubChem homepage, showing the upload ID list

4. In the resulting screen (Figure 23), keep the identifier type as “Compound, e.g. CID.” Choose the previously created text file CID list by clicking “Choose file” and navigating to the appropriate text file.

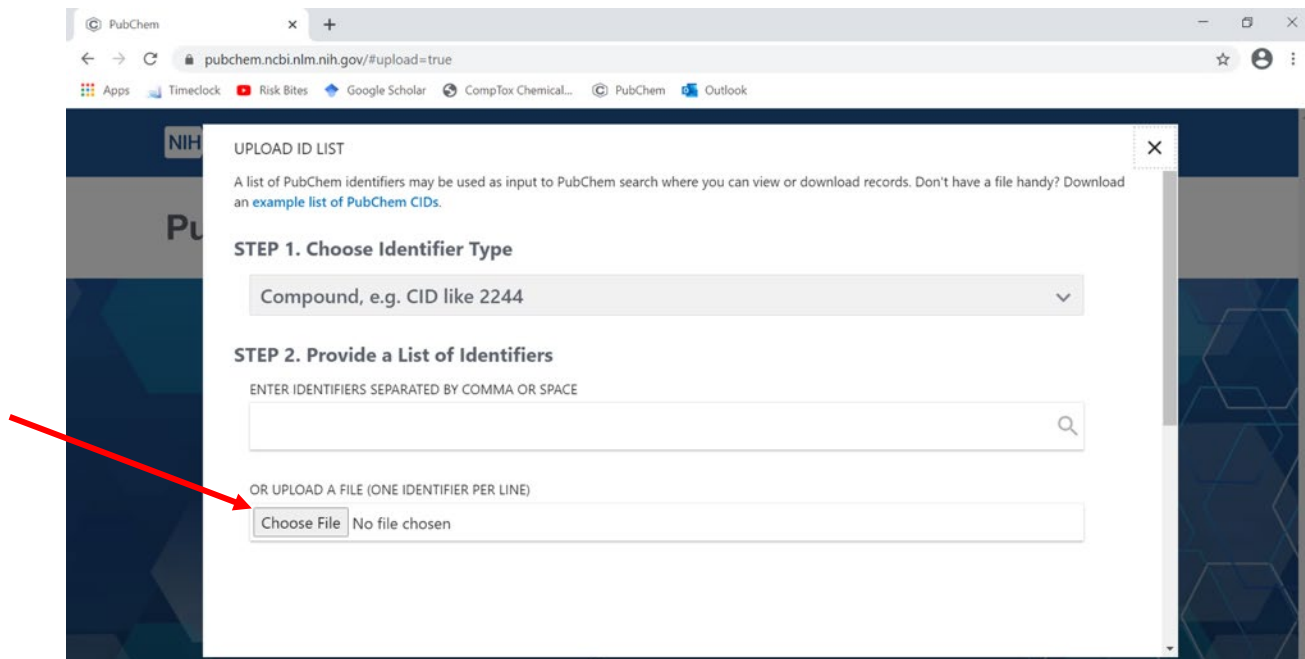


Figure 23. PubChem Uploading an ID list screen

5. After uploading the text file, the screen will preview the CIDs for upload, and indicate either “SUCCESS” (Figure 24 below) if the CID list is formatted correctly or “ERROR” if there is an issue. If “SUCCESS” is displayed, click “Search PubChem With This List” at the bottom of the page. If “ERROR” is displayed, see the error message for information on fixing the error to be able to proceed.

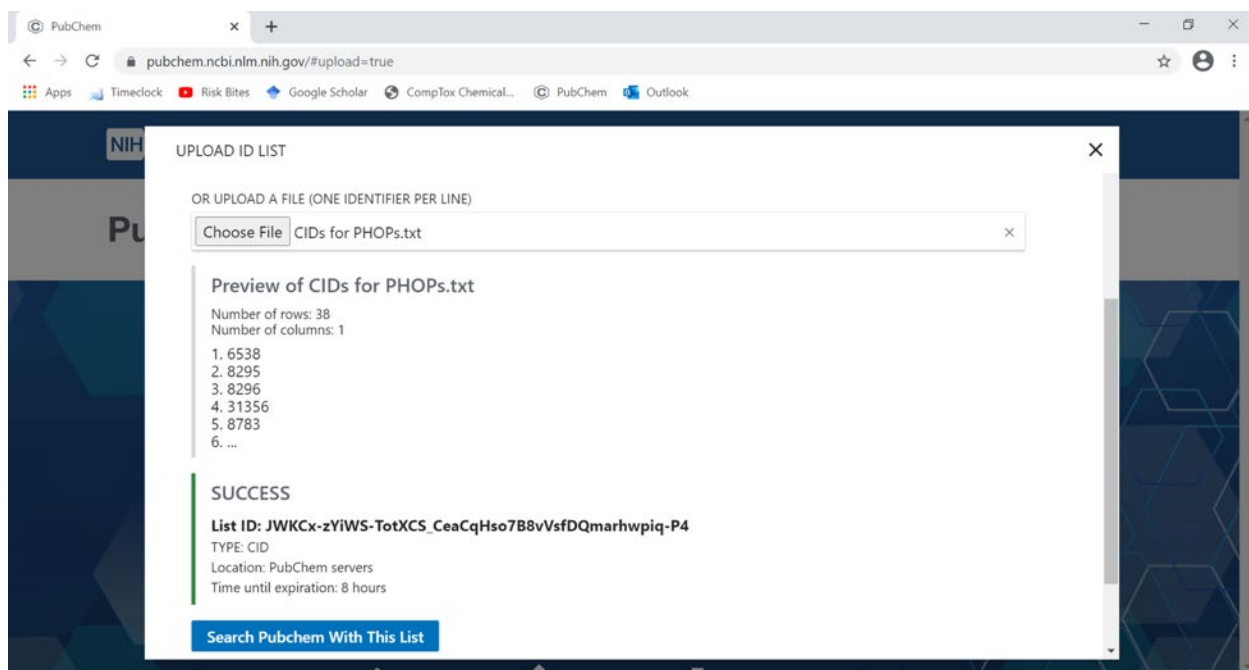


Figure 24. Example of the preview of CIDs after uploading an ID list

6. After searching PubChem, the resulting page will include profiles for each of the CIDs listed in the text file. You can now filter and sort the information. More information on different ways to download/filter data (and much more) can be found at the following link: <https://pubchemdocs.ncbi.nlm.nih.gov/downloads>. A summary of all search results can be downloaded via the Download button at the top right of the search results page.
7. Figure 25 shows the profile for a sample chemical. You can use the “Contents” to access specific sections, or scroll down the page. Note that the sections under “Contents” are often expandable to show more subsections, and that the specific sections listed (as well as the section numbers) are specific to each chemical profile. Many sections can be downloaded individually via the download button under the section headers (Figure 26).
8. Download data for (1) CTD Chemical-Gene Interactions, (2) BioAssay Results, and (3) Analytic Laboratory Methods for each chemical from among the results that PubChem provided. **Note** that there is **no bulk download** option for this step. Downloads must be performed for each chemical **individually**. To do so, click on a chemical result, and follow these instructions for each chemical result. **Note** that the number assigned to each section of the contents will vary based on the chemical, so it is important to pay particular attention to the title rather than the section number.
 1. To download CTD Chemical-Gene Interactions, in the Contents click the downward facing arrow to the right of Biomolecular Interactions and Pathways (Figure 25) in order to expand the contents to show subsections. Click the subsection CTD Chemical-Gene Interactions to access this section in the main part of the page. Under the CTD Chemical-Gene Interactions sub-header in the main part of the page, click the download button (Figure 26) to save a .csv file of

this subsection's data (similar to the example data shown in Figure 27 for BioAssay Results).

2. To download BioAssay Results, in the Contents click the downward facing arrow to the right of Biological Test Results (similar to what is shown in Figure 25) to expand the subsections. Click the subsection BioAssay Results to access this section in the main part of the page. Under the BioAssay Results sub-header in the main part of the page, click the download button (similar to what is shown in Figure 26) to save a .csv file of this subsection's data (Figure 27). Document in the master spreadsheet (see Section 2.3.2) whether either BioAssay or Chem Gene data were available. In addition, in the tracking spreadsheet, document whether each of these two sections exists, the date the site was accessed, and the direct link to either Compound Summary page (if the section does not exist) or the relevant section from the URL bar.
3. Information on whether Analytic Laboratory Methods exist for a given chemical can be useful to help explain the presence or absence of data for a chemical. (That is, if there are no analytic methods available for a chemical it is not surprising if data are not available). To download Analytic Laboratory Methods, in the Contents click the downward facing arrow to the right of Identification in order to expand the contents to show subsections (similar to what is shown in Figure 25). **Note** that the Analytic Laboratory Methods section is not directly downloadable in the same way as the CTD Chemical-Gene Interactions and BioAssay Results, so will require a different method.
 1. If the Analytic Laboratory Methods Section exists:
 1. Click the subsection Analytic Laboratory Methods to access this section in the main part of the page. To the right of the Analytic Laboratory Methods sub-header in the main part of the page, click the box with an arrow in it (located near where the download arrow is shown in Figure 26) to open this section of the page in a new browser window. Capture this newly opened page as a PDF by using the web browser's print function, and saving/printing as a PDF (example PDF in Figure 28).
 2. In the tracking spreadsheet, document that this section exists, the date the site was accessed, and the direct link to the section from the URL bar (this direct link should end in the CID, followed by #section=Analytic-Laboratory-Methods).
 2. If the Analytic Laboratory Methods section does not exist:
 1. In the tracking spreadsheet, document that it does not exist, the date the site was accessed, and the direct link to the chemical's Compound Summary page from the URL bar (this direct link should end in the CID)
4. Repeat for each chemical result on the PubChem search results page.

Tris(2-chloroethyl) phosphate

PubChem CID: 8295

Structure: 2D, 3D

Chemical Safety: Irritant, Health Hazard, Environmental Hazard

Molecular Formula: $C_6H_{12}Cl_3O_4P$ or $(ClCH_2CH_2O)_3PO$

Table of Contents:

- 10 Safety and Hazards
- 11 Toxicity
- 12 Associated Disorders and Diseases
- 13 Literature
- 14 Patents
- 15 Biomolecular Interactions and Pathways
- 16 Biological Test Results
- 17 Classification
- 18 Information Sources

Figure 25. Example chemical profile

PubChem Tris(2-chloroethyl) phosphate (Compound)

15 Biomolecular Interactions and Pathways

15.1 Chemical-Gene Interactions

15.1.1 CTD Chemical-Gene Interactions

124 items View More Rows & Details

Download

PubChem Gene	Interaction	Evidence PMID
ADRA2B	tris(chloroethyl)phosphate results in increased expression of ADRA2B mRNA	27776230
ADRB2A	tris(chloroethyl)phosphate results in increased expression of ADRB2A mRNA	27776230
AKT1	tris(chloroethyl)phosphate results in decreased expression of and results in decreased phosphorylation of AKT1 protein	26378621
AR	tris(chloroethyl)phosphate results in increased expression of AR mRNA	27776230
AR	tris(chloroethyl)phosphate binds to and results in decreased activity of AR protein	23398518

1 2 3 ... 25 Next >

Table of Contents:

- 10 Safety and Hazards
- 11 Toxicity
- 12 Associated Disorders and Diseases
- 13 Literature
- 14 Patents
- 15 Biomolecular Interactions and Pathways
- 16 Biological Test Results
- 17 Classification
- 18 Information Sources

Figure 26. Downloading chemical data from PubChem

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S		
1	baid	acvalue	aid	sid	cid	geneid	taxid	pmid	aidtype	aidmdate	hasdrc	rnai	activity	protacxn	acname	acqualifier	aidsrcnam	aidname	cmpdnam	ta
2	4.62E+08	2.7056	1346799	1.44E+08	8295	NULL	NULL		Confirmat	20181219	1	0	Inconclusi	NULL	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	NU
3	4.49E+08	4.8114	1224841	1.44E+08	8295	2101	9606		Confirmat	20180202	1	0	Active	ADZ17378	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	ES
4	4.48E+08	6.7962	1224843	1.44E+08	8295	7253	9606		Confirmat	20160720	1	0	Active	EAW81333	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	TS
5	4.5E+08	13.5603	1224895	1.44E+08	8295	7253	9606		Summary	20160914	0	0	Inconclusi	EAW81333	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	TS
6	4.38E+08	15.4845	743074	1.44E+08	8295	NULL	NULL		Confirmat	20140820	1	0	Inconclusi	NULL	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	NU
7	4.62E+08	19.3312	1346982	1.44E+08	8295	8856	9606		Confirmat	20190107	1	0	Active	ADZ17384	NULL	NULL	Tox21	Human pr	Tris(2-chlc	NF
8	4.49E+08	21.4916	1224842	1.44E+08	8295	2101	9606		Confirmat	20180202	1	0	Inconclusi	ADZ17378	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	NF
9	2.76E+08	22.3872	588532	17389691	8295	2908	9606		Confirmat	20111017	1	0	Inconclusi	ADP91253	NULL	NULL	National	CqHTS assa	Tris(2-chlc	NF
10	4.38E+08	24.3372	743075	1.44E+08	8295	2099	9606		Confirmat	20140820	1	0	Inactive	AEP43755	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	ES
11	29505165	28.1838	1030	17389691	8295	216	9606		Confirmat	20100315	1	0	Active	AAP35567	NULL	NULL	National	CqHTS assa	Tris(2-chlc	AL
12	4.38E+08	38.5719	743211	1.44E+08	8295	NULL	NULL		Confirmat	20140821	1	0	Inconclusi	NULL	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	NU
13	4.02E+08	48.1136	651631	1.44E+08	8295	7157	9606		Confirmat	20140820	1	0	Inconclusi	P04637	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	TP
14	4.02E+08	48.1136	651631	1.44E+08	8295	7157	9606		Confirmat	20140820	1	0	Inconclusi	P04637	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	TP
15	4.47E+08	54.4842	1159526	1.44E+08	8295	3725	9606		Confirmat	20150725	1	0	Inconclusi	EAX06628	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	JU
16	4.47E+08	60.5715	1159552	1.44E+08	8295	5914	9606		Confirmat	20150917	1	0	Inconclusi	ADZ17337	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	RA
17	4.38E+08	67.9623	743015	1.44E+08	8295	NULL	NULL		Confirmat	20140820	1	0	Inconclusi	NULL	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	NU
18	4.53E+08	67.9623	1259369	1.44E+08	8295	2737	9606		Confirmat	20180214	1	0	Inactive	AAI13617	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	GL
19	4.62E+08	67.9623	1346979	1.44E+08	8295	NULL	NULL		Confirmat	20190107	1	0	Inconclusi	NULL	NULL	NULL	Tox21	Caspase-3	Tris(2-chlc	NU
20	4.54E+08	68.5896	1259380	1.44E+08	8295	NULL	NULL		Confirmat	20180222	1	0	Inconclusi	NULL	NULL	NULL	Tox21	qHTS assa	Tris(2-chlc	NU
21	26716052	79.4328	965	17389691	8295	NULL	NULL		Confirmat	20100706	1	0	Inconclusi	NULL	NULL	NULL	National	CCell Viabili	Tris(2-chlc	NU
22	2502398	NULL	155	69736	8295	NULL	NULL		Other	20040815	0	0	Inactive	NULL	NULL	NULL	DTP/NCI	NCI Yeast	Tris(2-chlc	NU

Figure 27. Example of downloaded bioassay data

5/10/2021 Tris(1,3-dichloro-2-propyl)phosphate | C9H15Cl6O4P - PubChem

PUBCHEM > TRIS(1,3-DICHLORO-2-... > ANALYTIC LABORATORY METHODS

CID 26177

Tris(1,3-dichloro-2-propyl)phosphate

Analytic Laboratory Methods ?

Method: USGS-NWQL O-1433-01; Procedure: gas chromatography/mass spectrometry; Analyte: tri(dichloroisopropyl) phosphate; Matrix: filtered wastewater and natural-water samples; Detection Limit: 0.08 ug/L.

National Environmental Methods Index; Analytical, Test and Sampling Methods. Tri(dichloroisopropyl) phosphate (13674-87-8). Available from, as of March 5, 2013: <http://www.nemi.gov>

[▶ Hazardous Substances Data Bank \(HSDB\)](#)

A fabric sample from children's pajamas was extracted in acetone and resulting extracts analyzed by nuclear magnetic resonance (nmr).

PMID:347576
GOLD MD ET AL; SCIENCE 200: 785 (1978)

[▶ Hazardous Substances Data Bank \(HSDB\)](#)

Food samples (meats, dairy products, fruits, and root vegetables) were analyzed for various phosphate residues incl tris(1,3-dichloro-2-propyl) phosphate by gas chromatography equipped with flame photometric or nitrogen/phosphorus detectors.

PMID:7126910
DAFT JL; BULL ENVIRON CONTAM TOXICOL 29 (2): 221 (1982)

[▶ Hazardous Substances Data Bank \(HSDB\)](#)

Figure 28. Sample of PubChem Analytic Laboratory Methods saved as a PDF

3.4.3 How to Download PubChem Bioassay Information in Bulk

PubChem provides many options for programmatic and bulk access to data. One of the simplest is via PUG REST (general tutorial page <https://pubchemdocs.ncbi.nlm.nih.gov/pug-rest-tutorial>).

To get all bioassay data from a list of CIDs:

1. Add the relevant list of CIDs, separated by commas, to a URL. Obtain CIDs according to Section 3.4.1.
 - a. For example, to download all bioassay data from CIDs 1000 and 1001 in CSV format, the URL is:
<https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/1000,1001/assaysummary/CSV>.
2. Click the link or copy and paste the URL into an internet browser to download and save the desired results. Be sure to add .csv to the end of the file name to save as a CSV file. Save with a file name of the format CLASS NAME_Bioassay Downloads.csv.

The above is outlined in the following section of the PUG REST tutorial:

[https://pubchemdocs.ncbi.nlm.nih.gov/pug-rest-tutorial\\$Toc458584434](https://pubchemdocs.ncbi.nlm.nih.gov/pug-rest-tutorial$Toc458584434)

3.4.4 The Comparative Toxicogenomic Database (CTD) as an alternative source to chemical-gene data from PubChem

The Comparative Toxicogenomic Database (CTD) contains information on the interaction between chemicals and genes. CTD is the source of chemical-gene interaction data on PubChem. Thus, it can serve a similar purpose to the chemical-gene interactions downloads from PubChem. The main benefit of CTD is that it can be batch-searched, while chemical-gene interaction data from PubChem must be downloaded chemical-by-chemical.

CTD data is updated monthly. Scroll to the bottom of any page, and the lower left corner will indicate the date data was last updated. As of this writing, data was last updated August 5, 2021. Check here for info on most recent data release: <http://ctdbase.org/about/dataStatus.go>

How to download chemical-gene data from CTD

1. Go to <http://ctdbase.org/tools/batchQuery.go>
2. Select input type “Chemicals” (the default selection).
3. Enter a list of CASRNs for search under “Provide query terms”
 - a. This list can be copy-pasted (return, tab, or | delimited) OR uploaded as a tab-separated file
4. Under “Choose data to download”, select Format of CSV. Then select the type of data for download
 - a. Download 6 types of data. The download process will need to be repeated for each data type, using the same list of CASRNs.
 - A. Curated Chemical-gene interactions (Curated Types: ANY)
 - B. Curated Gene associations
 - C. Curated Phenotype associations
 - D. Curated Disease associations (not inferred)

- E. Enriched Gene Ontology associations
 - F. Enriched Pathway associations (not inferred)
5. Under “Select report output”, choose “Write report to a file”
 6. Click Download
 - a. Repeat above steps 4-6 for all 6 data types outlined under step 4 (Can use the same entered/uploaded list of CASRNs for all 6 types).
 7. Copy all 6 downloaded CSV files into an Excel spreadsheet (.xlsx), as tabs labeled A-F. Save as CLASS NAME_CTD_Downloads.
 - a. A readme tab should
 - i. Include the date of download, and identify the file as a download from the Comparative Toxicogenomic Database
 - ii. identify tabs A-F using **Table 3. Example table for tab definitions in CTD downloads**

Table 1. Example table for tab definitions in CTD downloads

Data from CTD
[A] Curated chemical-gene interactions
[B] Curated gene associations
[C] Curated Phenotype associations
[D] Curated Disease associations (not inferred)
[E] Enriched Gene Ontology associations
[F] Enriched Pathway associations (not inferred)

8. Note the date of downloads for metadata file.

3.5 Danish QSAR Database

Data on OFRs and analogs from the Danish QSAR Database (<https://qsar.food.dtu.dk/>) were provided by CPSC. Files were of 2 formats:

- QSAR.food.dtu.dk_[CAS]
- QSARModels.food.dtu.dk_[CAS]_SMILES model

3.6 Additional Grey Literature Sources Treated as Databases

During early screening of PDF data, it was determined that one source originally considered grey literature would be better handled as database data rather than as PDF data. Because it was initially retrieved as PDF data, downloading from this source is described together with the PDF data in Section 5. The source is ECHA Dossier information retrieved via IUCLID (Section 5.4.6).

3.7 MMDB (Multi-Media Monitoring Database)

MMDB pulls data from various sources (e.g., NHANES, CTD, Biomonitoring California), analogous to how ToxVal pulls from a variety of different sources. Version 1.0 of the MMDB was downloaded from GitHub (<https://github.com/HumanExposure/MMDB>) and provided in Excel, last updated 5-7-2021. Data were aggregated by media present. In contrast to the toxicity databases, this exposure database was used only for OFRs, and not analogs.

4. Develop database data format as needed for toxicity data and extract relevant data from databases.

For the toxicity database data, a key challenge is that different databases have different information and different formats. This means that the database downloads cannot be combined using a simple cut and paste approach. (As noted in Section 2.4, exposure database data were provided to the UC team in raw and aggregated form, and so pre-processing was not needed for such data.) For toxicity database data, the extracted fields were designed to match the tagging tree developed for the PDF data and documented in the Screener Guide. As described in the rest of this section, database logic was developed for each database that described the approach for extracting data from the database download in order to develop evidence maps that match the tagging tree. The database logic was then encoded in R, and used to extract the relevant data into data tables with the appropriate number of tags. The data tables were then used to develop evidence maps, as described in Section 7.

4.1 Approach for Database Logic

This section outlines the general approach for developing the database logic files, how decisions were made regarding database formatting and combination, and extraction of data from databases. It contains guidance on how such formatting, combination, and extraction will be done, once such decisions are made. The detailed logic is provided in the file All tox Database logic for evidence maps_2.27.22 in the folder Task 12 - OFRs Lit Survey Info\Task 3 - for CPSC.

Decisions regarding formatting, combination, and extraction of data from databases considered endpoint, species, duration, filling data gaps, and potential duplication of data. When there was a conflict between filling a gap in data and potential duplication of data, filling data gaps was

given a higher priority (i.e., some duplication was acceptable if it meant also filling a gap in data). Information contained within each database and information from documentation about databases were both used to inform these decisions. When multiple formats of data were available, the format most amenable to automated extraction was selected. For example, ICE data is available in 3 formats: wide, long, and summary; long format is most amenable to writing of logic and code for extraction.

To extract data from the databases described in section 3, and to combine that data in a harmonized way, the UC team created a “database logic” for each database. The database logic file serves to 1) harmonize data across databases, 2) document decisions made on handling of data, and 3) communicate between toxicologists (interpreting toxicological data in databases) and biostatisticians (writing and running code consistent with the database logic, to extract data from databases). Design of the database logic files took multiple iterations for each. The output of the developed code for each database logic also went through several iterations of QA, ensuring that the database logic and the code 1) agreed with each other, 2) performed as intended, and 3) addressed characteristics of the data that were unique to different classes.

Note that ECHA/IUCLID Dossier and HPVIS files were initially screened as PDFs, since that was the format of the downloads from the source databases. However, it became apparent during early Distiller screening of these files that it would be more appropriate to treat them as database sources than as grey literature sources. Therefore, these “grey literature” sources also have a “database logic.”

The majority of databases were relevant only to toxicity data, and not exposure data. Databases for toxicity data were also much more complex than the sole exposure database, the Multi-Media Monitoring Database (MMDB). Matching data fields from MMDB to the appropriate tags from the tagging tree was straightforward, and mainly involved identifying environmental and biological media. In addition, there was one field that mapped to Source Characterization/Product Content. One duplication of data was identified in data sourced from the National Health and Nutrition Examination Survey (NHANES). Specifically, NHANES reported data on OFRs in urine, both as the values adjusted for creatinine and un-adjusted values. Since these were based on the same measurements, only one copy of the urine data was counted for evidence maps. Otherwise, mapping the data in MMDB was straightforward and no complex logic files are required. A basic logic file mapping fields from MMDB to the appropriate tags is provided separately from the toxicity database logic files.

Logic files for some toxicity databases are quite complex, often drawing together data from multiple fields to properly categorize data. In particular, logic for toxicokinetics data was often more complex than many other endpoints. Each “database logic” file considers variable names, data types, and species present, and the unique way such information may be presented in each database. The information available in each database was then categorized into appropriate Level 3 and Level 3B tags. When necessary and possible for consistency between database and PDF

data, entries from databases are deduplicated (usually according to a reference or publication provided for the data). However, not all data in databases have an associated reference, so deduplication was not possible in all instances.

Note that in database logic files, Tier 1 and Tier 2 correspond to Level 3 and Level 3B of the tagging tree, respectively. The terminology for the tagging tree was modified after the initial development of the database logic, in order to be consistent with screening levels in Distiller. However, the older “Tier” terminology was retained for the database logic files for ease of communication with the team biostatistician. Similarly, although ICE is the proper name of NTP’s Integrated Chemical Environment, we have retained the previously used term “WebICE” in the database logic files.

In general, the logic files assign Level 3 and 3B tags based on conditions related to the data present in each database. Some broader approaches include:

- if-then logic
 - Used when a data type can be tagged many different ways, depending on a variety of conditions
- always-unless/except logic – a data type is tagged one way, unless another condition is met.
 - Used when most entries for a data type are tagged the same way, with a small number of exceptions
- simultaneously capturing multiple related variables with similar names by using “CONTAINS” logic
 - e.g., BMD, BMDL, BMDL 10 are all captured in a single line of logic pertaining to variable names that “contain” the term “BMD”

4.2 Assumptions and Considerations

Some assumptions and decisions were made in designing database logic. To the extent possible, these are consistent between the logic for database data and the guidance in the screeners guide for PDF data. Some of the key assumptions and considerations are described below, primarily those that 1) apply broadly and are commonly used for many variables present across several databases, or 2) are particular to the interpretation of a specific variable in a specific database, in contrast to other broader assumptions. Documentation of additional assumptions and considerations are included in the database logic files themselves.

- When a toxicity reference value (TRV, such as an RfD) was tagged as Quantitative Hazard Characterization (Level 3):
 - It was also tagged by default as Animal Toxicity (Level 3). For both of these Level 3 tags, the Level 3B tag was Systemic Repeated Dose Toxicity. It was assumed that the existence of the toxicity value meant that animal data also exist,

UNLESS there was additional information that explicitly identified the toxicity value as being derived from human data.

- It is recognized that this is a simplifying approach, but it was followed for consistency across databases, even though it would miss some human data. Many of the databases did not identify the species that was the basis for the toxicity value, and some flagged the value as human, regardless of the species that was the basis for the point of departure. Similarly, the toxicity values were tagged as Systemic Repeated Dose regardless of the critical effect, since the toxicity value is designed to protect from all repeated dose exposures for that route. An exception was that acute DNELs (Derived No Effect Level) from IUCLID were designated as acute.
- For 3B tagging, TRVs were counted separately by route and duration for database data, but not for PDF data.
 - This difference is because the Distiller forms used for Level 3 tagging of PDF data were dichotomous (a data type is present in a reference, or it is not); there is no ability to indicate counts of a data type within a reference. The Excel system used for Level 3B tagging of PDF data was also designed to be dichotomous, for consistency between Levels 3 and 3B of PDF data. A limitation of this approach was the inability to count more than one TRV per reference for PDF data.
- As for Level 3 tags for TRVs, qualitative calls by authoritative groups were assumed to be based on animal data, unless otherwise specified, and so the application of any qualitative hazard tag other than “cannot be determined” or the equivalent meant that a tag for Animal Toxicity was also applied.
- Qualitative calls made for a single study or assay were not tagged as Qualitative Hazard Characterization. In order to be tagged, qualitative calls needed to be based on a review of the literature, such as a weight of evidence approach.
- Relevant types of data on terrestrial mammals (aside from standard laboratory animals) were counted as Experimental Mechanistic (as NAMs). This is because the terrestrial mammal data do provide some information on the effects of the subject chemical.
 - E.g., an LD50 in rats would be counted as Animal Toxicity and Quantitative Hazard Characterization, while an LD50 in deer would be counted as Experimental Mechanistic.
 - This was a late change made to harmonize the approach used for the literature survey with the approach NTP is using in their full literature review. Thus, it is noted that some of the terrestrial mammal data in PDF sources has likely been missed for Classes 1 and 2, for which tagging was largely completed prior to this change. Tags for these data were added when found during QA on Levels 3 and 3B data from PDFs for Classes 1 and 2.
- Zebrafish were counted as Experimental Mechanistic instead of Animal Toxicity (for both database and PDF data). However, LC50s and other points of departure from zebrafish were tagged as Quantitative Hazard Characterization.
- The screeners guide defines acute as up to 14 days (based on the approach used by ATSDR and EPA Health Advisories), and that approach was followed for classifying the

database data. However, since the tagging tree did not include a category of short-term, this means that studies of ≥ 15 days are classified as systemic repeated dose toxicity, which could lead to over-estimation of the number of studies that could be the basis for (chronic) RfDs. This means that the systemic repeated dose category includes studies ≥ 15 days, up to >90 days, 2 years, or longer.

- In ToxVal, LD50s, LC50s, and similar values are assumed to be acute, and no check of the study duration is used. Entries that are not explicitly identified as acute (e.g., identified as chronic or reproductive) were determined to be either not OFRs or are in excluded species.
- An additional 3B tag, “*In vitro* toxicokinetics,” was added late. This addition applies for PDF data and database data. Though it is noted that the late addition **may** have led to the exclusion of some PDF sources from the *in vitro* toxicokinetics tag, the UC team does not anticipate this to have a significant impact on evidence maps from Classes 1 and 2. In the absence of a 3B tag for *in vitro* toxicokinetics studies, it is likely that *in vitro* toxicokinetics data from PDFs for these classes were tagged as “Human, Animal, or Modeled Toxicokinetics/ADME” at Level 3.
- TK/ADME studies (e.g. ToxVal types ABS and GIABS) are assumed to be animal studies if no species is given and not specified as *in vitro*.
- For QSAR and read across data
 - When QSAR results were outside a model’s applicability domain, those data were excluded. Positive, negative, and inconclusive/uncertain results were included.
 - QSAR and read across data were tagged only as QSAR/read across at Level 3; additional Level 3 tags (e.g., Animal Toxicity) were not applied to these data.
- Genotoxicity and mutagenicity data were always tagged as Animal Toxicity, matching handling of the PDF data.
- AC50s from *in vitro* tests were generally captured in ICE data and excluded elsewhere to avoid double counting (since the data were already captured in ICE). However, AC50 data from ToxVal were tagged, because those were from studies conducted in *C. elegans* and were clearly different data from those in ICE. Unlike the *in vitro* data in ICE, there was not a clear way to systematically determine whether toxicity values, points of departure, and individual studies were repeated across databases. It is likely that there is some double or triple counting for the Animal Toxicity and Quantitative Hazard Characterization tags, and possibly for some of the other tags.
- For several data types tagged as Experimental Mechanistic across several databases, no 3B tag was applied because not enough information was provided to easily determine if the assay made a connection to MOA or health effect or not (i.e., to distinguish between the two relevant 3B tags).
- After consultation with CPSC:
 - *In vitro* irritation and corrosion data were tagged as Animal Toxicity or Accepted Alternative. While these *in vitro* studies are not yet fully accepted as an alternative, development of guidelines for their use as an alternative to animal models is well underway.

- *In vitro* sensitization data were tagged as either Animal Toxicity or Accepted Alternative, or as Experimental Mechanistic, according to the “2 out of 3 Key Events” defined approach/integrated testing approaches (DA/IATA), as explained in the screener’s guide.
- Most Acute Toxicity studies in HPVIS were tagged for both Animal Toxicity and Quantitative Hazard Characterization, based on the identification of an LD50/LC50. However, no LD50/LC50 was reported for some Acute Toxicity studies, and therefore these studies did not receive a Quantitative Hazard Characterization tag.
- In data extracted from HPVIS, some studies tagged for more specific endpoints (e.g., Neurotoxicity) may have already been counted for Systemic Repeated Dose Toxicity. Although this would lead to some minor double-counting of studies, separate tags by endpoint is consistent with the approach used in 3B tagging for PDFs.
 - Conversely, studies in repeated dose sections were tagged for Systemic Repeated Dose Toxicity, even if, for example, reproductive effects were noted. Studies received reproductive (or developmental) tags only if they were in the respective section of the file, and therefore were the specified study designs.
- Some studies in ECHA/IUCLID flagged as key studies for a study type did not identify NOAELs/LOAELs. This meant that the study did not receive a Quantitative Hazard Characterization tag even though there may have been quantitative information in the study summary.
- Some data from Danish QSAR incorporate 4 different QSAR models. As long as a data type had an included result (positive, negative, or inconclusive, if inside the applicability domain) from at least one of the models, that data type received one count for the relevant tag(s). Only one count was applied for the relevant tag(s), regardless of the number of models with an included result.
- In Danish QSAR, there are multiple rows related to the Ames test. Only the first was counted; the rest are subsets of the data in the first.
- In ICE data, values of “Flag-Omit” and “QC-Omit” are included in the data as “inconclusive” results, using the same reasoning as for inconclusive results from other databases. That is, the fact that someone was interested enough to look at a chemical is useful information, regardless of whether or not they obtained conclusive results.
- In ICE data, assays can have up to 4 endpoints: ACC, AC50, top of curve, call. A given assay received only a count of 1, regardless of the number of endpoints reported.
- Since HPVIS and ECHA/IUCLID both derive from nearly the same entry system, they are likely to have substantial overlap when a chemical has data in both. However, based on comparing the respective files for a chemical covered by both databases, it appears that each database has the potential to capture studies not covered by the other database. In addition, only ECHA/IUCLID includes DNELS. Because data gaps are more of a concern than double-counting, both HPVIS and ECHA/IUCLID were included in the evidence map when both databases had data for a chemical.

5. Searching for Peer-Review and Grey Literature (pdf Data)

The goal for this searching is to identify literature reviews and completed assessments, in order to characterize the overall availability of data. Searches for these literature surveys and assessments are not intended to identify individual research studies, and do not need to be exhaustive.

5.1 Targeted Searching for Completed Assessments and Literature Reviews in Bibliographic Databases

Two search strategies are employed to maximize the likelihood of identifying relevant literature during screening— one using class-specific terms (Section 5.1.2) and one using chemical-specific terms (Section 5.1.3). These sets of terms are each be combined (using AND) with search terms designed to capture literature reviews and complete assessments (Section 5.1.1). Due to the high likelihood of co-reporting, one comprehensive literature search strategy, comprised of multiple literature searches and deduplication of results across searches, is used to reduce the possibility of screening studies multiple times across multiple searches. When deduplicating, the PubMed entry is preferred, because it includes more meta-data than do other formats. Whenever possible, wildcard characters are used to capture similar terms (e.g., in PubMed, organophosph* will be used to cover organophosphate(s), organophosphorus, etc.). The class- and chemical-specific portions of the search are conducted with “All Fields,” and the portions of the search strings relating to literature type and study type are restricted to the title and abstract. For class- and chemical-specific search terms, search all text/all fields, rather than limiting the search to title and abstract, since a chemical of interest may be mentioned in a minor manner in the text, and not in the title or abstract. However, literature type or study type terms may be limited to title and abstract.

PubMed are searched for each element of the strategy, while only class-specific searches are conducted with the Web of Science, using appropriate modifications to the search terms, as noted below.

Search meta-data are tracked in a file titled “!PR Lit tracking sheet-[CLASS NAME]+Analog.” Meta-data collected include: the date of the search, the database used, the search query, and the number of results returned.

5.1.1 Literature type and study type terms

Because the goal is to gather reviews and completed assessments, rather than individual research articles, an integrated search under each strategy element (e.g., using class-specific search terms vs. using chemical-specific search terms) covers reviews, hazard, exposure, and risk. The set of search terms in PubMed will be ("review"[Title/Abstract] OR "risk assessment"[Title/Abstract] OR "exposure assessment"[Title/Abstract] OR "hazard assessment"[Title/Abstract] OR (toxic* [Title/Abstract] AND profile [Title/Abstract])). We also considered as search terms: “literature

review,” “systematic review,” “literature survey,” and “meta-analysis.” Trial searches using the organophosphate class-specific terms excluded these terms because: 1. “literature review” and “systematic review” are covered by the already included term “review”, and 2. “literature survey” and “meta-analysis” added few or no studies compared to not using these terms.

5.1.2 Class-specific search terms

These searches are useful for identifying reviews that capture multiple OFRs. The search terms are: (([class name terms] AND “flame retard*”) OR trade names) AND review. For example, for the polyhalogenated organophosphate class, the combined search string in PubMed will be: ((Organophosph* AND "flame retard*") OR Fyrol OR antiblaze OR Genomoll OR Niox OR Bromkal OR FireMaster OR Flammex OR Phoscon OR Zetofex) AND ("risk assessment"[Title/Abstract] OR "exposure assessment"[Title/Abstract] OR "hazard assessment"[Title/Abstract] OR "review"[Title/Abstract] OR (toxic* [Title/Abstract] AND profile [Title/Abstract])). The “flame retard*” term² is optionally used only when initial search attempts return many unrelated results – such as search terms for Polyhalogenated Organophosphates returning results for organophosphate pesticides. Note that the full NAS name is not used; we found that including “polyhalogenated” in the search string for the class name was overly restrictive. Note also that the wildcard in the flame retardant term (i.e., flame retard*) aims to capture flame retardant(s), flame retarding chemicals, flame retardance/retardancy, etc. Quotes are used for classes with multi-word names (e.g., “polybrominated diphenyl ether”). Terms derived from trade names are included in the class-specific search terms rather than the chemical-specific search terms since the same trade name is often used with minor modifications for different products. Note that the same component of many trade names is used across different NAS classes (e.g., Saytex).

Specific terms used in the class name portion of searches are adjusted on a class-by-class basis, as many class names are not standardized. For example, (halogenated) organophosphate flame retardants are a well-recognized previously existing group matching reasonably well to the Polyhalogenated Organophosphate class. In contrast, the Polyhalogenated Benzene Aliphatics and Functionalized class has no analogous widely-recognized and -used existing group of chemicals. Chemicals from the Polyhalogenated Benzene Aliphatics and Functionalized class are often grouped with members of several other classes in the literature, such as with the brominated flame retardants (BFRs) or novel brominated flame retardants (NBFRs). Most OFR classes are closer to this latter category than to the former. Thus, different approaches are needed to generate class-based search terms for different classes. Applying relevant chemistry terms is one approach. Functional groups or other aspects of chemical names common to many members (and/or analogs) of a class are good candidates for inclusion as search terms. While class-based

² Note that over the course of the literature searches for different classes, the exact flame retardant-related “limiting term” used varies. This is related to a number of factors, including: evolving understanding of the literature searches, suggestions from CPSC, class size, and class data-richness. Term used include: “flame retard*”, “fire OR flame”, and “(fire OR flame) AND retard*”.

search terms are not specifically designed to capture analogs indicated by the EPA CompTox Dashboard to have no hits in PubMed, functional group/chemical name terms derived from these “no hit” analogs may be included in the class-based search if such terms are common among at last several class members and analogs. A brief look at literature on a more common member of the class may also give alternative existing class names or other keywords that could be included. Following the above approaches, the PubMed search terms used for the class-based search for Polyhalogenated Benzene Aliphatics and Functionalized were: (((halogenated benzene aliphatic*) OR ((bromo* or chloro*) AND (phenyl* OR styrene OR benz* OR toluene OR xylene OR aniline))) AND ("risk assessment"[Title/Abstract] OR "exposure assessment"[Title/Abstract] OR "hazard assessment"[Title/Abstract] OR (toxic* [Title/Abstract] AND profile [Title/Abstract]) OR "review"[Title/Abstract])) AND (flame OR fire).

For Web of Science, a similar syntax approach is used. The class portion of the search is conducted with “All Fields,” and the literature and study type portion is restricted to the title and abstract, as for the PubMed searching. For example, for PHOPs, the search terms were: ALL=((Organophosph* AND "flame retard*") OR Fyrol OR antiblaze OR Genomoll OR Niaux OR Bromkal OR FireMaster OR Flammex OR Phoscon OR Zetofex) AND (TI=("risk assessment" OR "exposure assessment" OR "hazard assessment" OR "review" OR (toxic* AND profile)) OR AB=("risk assessment" OR "exposure assessment" OR "hazard assessment" OR "review" OR (toxic* AND profile))). Note that this search string is longer than for PubMed, as Web of Science does not have the same combined title and abstract tag that PubMed has, and so must be tagged separately for title and for abstract. Trial searches in Web of Science were conducted using Research Area and WOS Category to attempt to narrow down a larger number of results from Web of Science than from PubMed. Research Area is searched using SU= and WOS Category is searched using WC=. Initial screening of lists of Research Areas and WOS Categories provided “Toxicology” and “Public, Environmental & Occupational Health” as promising starting points. Initial test searches did not detect any differences between using Research Area or WOS Category. Documentation suggests there is often substantial overlap between the two options, but not always 1:1 correspondence. While these trial searches did narrow the results, the UC team was concerned this narrowing came at the expense of relevant topics that may not have been included in the Research Areas/WOS Categories selected. Thus, this method of further narrowing search results at this stage in the process was rejected in favor of including more results now and potentially screening out more studies later.

Class-based searches in Scopus are also conducted for some classes, when fewer than 50 **unique** references are returned as a result of 1) the above class-based searches in PubMed and WoS, and 2) the below PubMed searches for chemical-specific terms. Scopus follows a similar approach to PubMed and WoS, with the class portion of the search conducted with “All Fields,” and the literature and study type portion restricted to the title and abstract. An example of Scopus search terms, for Class 6 Polyhalogenated Phenol Aliphatic Ethers:

ALL (((("phenolic ether") OR ((bromo* or chloro*) AND "phenol*" AND "aliphatic" AND "ether")) OR ((phenoxy OR anisole OR oxybenz*) AND (halo* OR bromo* OR Chloro*)))

AND

TITLE-ABS-KEY (((("risk assessment" OR "exposure assessment" OR "hazard assessment" OR "review" OR (toxic* AND profile))))))

AND

TITLE-ABS-KEY ((retard* OR fire OR flame))

5.1.3 Chemical-specific search terms

Bibliographic database searching is not conducted for chemicals for which there are no hits on PubMed in the initial EPA CompTox Dashboard download. For chemicals with up to 20 PubMed hits identified in the CompTox batch download (group 1 chemicals), the chemical identity information used for searching is limited to the most common name, acronym (if relevant), and CASRN linked by "OR" and combined with the literature/study type search string using AND, using the syntax described below. If there are 20 or more PubMed hits in the CompTox batch download (group 2 chemicals), a curated list of synonyms (including names, key synonyms, common acronyms and CAS number) is used, linked by "OR", and combined with the literature/study type search string using AND. Due to the complexity of the search string and character limits in search strings, several group 1 chemicals can be combined into one search, while group 2 chemicals need to be searched individually or in smaller groups.

The specific syntax approach for PubMed will be: Use quotes around individual synonyms (unless a single word or acronym, e.g., TDCPP); add syntax for CASRN (i.e., [RN] – as shown in the example in this paragraph). For chemical-specific terms, search all text/all fields, rather than limiting the search to title and abstract, since a chemical of interest may be mentioned in a minor manner in the text, and not in the title or abstract. However, literature type (e.g., review) or other limiting terms (e.g., toxic*) may be limited to title and abstract. As an example, for TDCPP, a partial list of the search string for PubMed is: ("Tris(1,3-dichloro-2-propyl) phosphate" OR "13674-87-8"[RN] OR TDCPP OR "tris[2-chloro-1-(chloromethyl)ethyl] phosphate") AND ("risk assessment"[Title/Abstract] OR "exposure assessment"[Title/Abstract] OR "hazard assessment"[Title/Abstract] OR "review"[Title/Abstract] OR (toxic* [Title/Abstract] AND profile [Title/Abstract])). [The full search string is documented for each OFR and analog, and the documentation is part of the spreadsheet documenting the literature search results.] Note that for analogs, the term "exposure assessment"[Title/Abstract] is dropped, as analogs do not need to be searched for exposure data. The resulting literature and study type portion of the search string for analogs is thus: ("risk assessment"[Title/Abstract] OR "hazard assessment"[Title/Abstract] OR "review"[Title/Abstract] OR (toxic* [Title/Abstract] AND profile [Title/Abstract])).

For two members of the PHOP class (CASRN: 13674-84-5 and 13674-87-8), searches initially returned far more PubMed results than expected based on the “Number of PubMed Articles” indicated by the EPA CompTox Dashboard. In both of these cases, there were terms included in the PubMed search query for which PubMed did not find any results. When this happens, PubMed ignores these terms and uses automatic term mapping (more information here: <https://pubmed.ncbi.nlm.nih.gov/help/#searching-for-a-phrase>). Briefly, automatic term mapping breaks the “not found” term into constituent parts and searches using those instead. For example, a term like "Tris(1-chloropropan-2-yl) phosphate" is turned into: ("tris 2 chloroisopropyl phosphate"[All Fields] OR ("Tris"[All Fields] AND ("phosphated"[All Fields] OR "phosphates"[MeSH Terms] OR "phosphates"[All Fields] OR "phosphate"[All Fields] OR "phosphatic"[All Fields] OR "phosphating"[All Fields] OR "phosphation"[All Fields] OR "phosphatized"[All Fields])). By dropping the “not found” terms from our queries, searches returned expected numbers of articles.

Similar instances of automatic term mapping inflating the number of results should be watched for in future classes. There are two ways to identify when this is happening. First, when viewing the results of a PubMed search, there is a warning above the list of results indicating “Quoted phrase not found:” and then listing the phrase. Second, in the Advanced view, under History and Search Details, clicking the down arrow under the Details column reveals the automatic term mapping as in the example above, as well as a similar “Quoted phrase not found” warning.

While not an issue for the PHOP class, another issue to check for in the event that a search returns far more results than expected is search terms that map to a supplementary concept. This can be checked in the Advanced View of PubMed searches. Under History and Search Details, clicking the down arrow under the Details column will reveal whether terms are mapped to a supplementary concept. To check if the supplementary concept is adding unexpected search results, compare the query as is vs. the query modified to tag the offending phrase with [Text Word] or [tw]. If the modified query has fewer results, the supplementary concept may be adding results to the search.

In an effort to limit the scope of searching and repetitive screening of results, a modified chemical-specific search strategy is applied to especially data-rich analogs. This modified search strategy is triggered when a single analog returns more than 250 results from the standard chemical-specific search process. The overall approach is as follows:

1. Limit by date of study release, using a recent major assessment to determine year cutoff (this recent major assessment is also included for screening).
 - a. Trigger this approach when >250 hits for a single analog, after searching with literature and data type terms.
 - b. E.g., for Lindane, there is a 2005 ATSDR Toxicological Profile, and a 2002 EPA Office of Pesticide Programs (OPP) Reregistration Eligibility Decision (RED). A conservative approach to selecting a year cutoff resulting in only including

references more recent than the year 2000, and including both the Tox Profile and the RED as additional grey literature sources.

2. Then if still >250 hits, limit to reviews (reviews, systematic reviews, meta-analyses).only. (This criterion can be applied first in cases where there is no clear assessment to use to set a year cutoff).

5.1.4 Documenting Search Meta Data

For each search, document meta data: the date of the search, database used, search string, any filters applied and number of hits. In PubMed, documenting most meta data is easily done for many searches at once from search history/advanced search page. At the bottom of the page, there is a box containing a history of searches. At the top of this box is a “download” button. After ensuring all correct searches are present, click “download” to save a .csv file that includes most of the above meta data. The required meta data are then entered into the tracking spreadsheet. Note that the search string, number of results, and any filters applied can be copied directly from the .csv file downloaded from PubMed, to minimize potential for error. Date of the search and the database used (here, PubMed), must be manually entered into the tracking file, as they are not included in the .csv downloaded from PubMed.

In Web of Science, meta data must be documented manually. As with PubMed, the Web of Science meta data are entered into the tracking spreadsheet, including a note on the name and location of the Web of Science search history file. In both PubMed and Web of Science, the results of each individual search can be downloaded; these downloads should be performed and kept as records (outlined in Section 5.1.5 below).

5.1.5 Saving the search and uploading to Endnote

Each search is saved and then uploaded into Endnote. Saving each search separately allows tracking to ensure that all searches have been uploaded, as well as comparing the number of hits using different search terms. After a PubMed search is conducted, to save all search results to EndNote, at the top of the search results page, click “Send to”, then select Citation manager. Ensure that Selection reads “All results”, and click create file to create a file (.nbib) that can be directly imported into EndNote (see below).

In Web of Science, this process is similar. After the search is conducted, there is a button labeled Export at the top of a search results page. Click on the Export button to show export options. Select Export to EndNote Desktop to create a file (.ciw) that can be directly imported into EndNote (see below).

Upload the results into a literature management program (such as EndNote), using the .nbib and .ciw files generated above, and removing duplicates (outlined in section 5.3). [The UC team uses

Endnote, which has a “remove duplicates” option when new data are uploaded, although this approach does not catch all duplicates.

To upload search results into EndNote:

1. Create an EndNote library named for the chemical class of interest, and title it “CLASS NAME_Archive.” Keep the library open.
 - a. E.g., for PHOPs, “PHOPs_PR_Archive_DATE”
2. Locate and open the downloaded EndNote file (.nbib from PubMed, .ciw from Web of Science. Both will open in EndNote).
3. EndNote will import the references from the downloaded file (.nbib or .ciw) into the open EndNote library, and place them in a temporary folder called “Imported References”.
 - a. Note: if no EndNote library is open when opening .nbib or .ciw files, EndNote may automatically import references into the last open library. It is worth checking EndNote settings and testing imports before proceeding with importing. EndNote will prompt to select and open the proper EndNote Library.
4. EndNote will then import all results of that search into a temporary folder called imported references.
5. Create a group by right-clicking on “My Groups” in EndNote, on the left side.
6. Name the group to correspond with the imported search. With the search number from the tracking spreadsheet.
 - a. For class-specific searches, name the group according to the database used (either PubMed, Web of Science, or (if applicable) Scopus) and indicate it was a class search
 - i. E.g. “PubMed – Class”
 - b. For Group 2 chemical-specific searches, name the group according to the CASRN
 - i. E.g., for searches of TCEP, “115-96-8”
 - c. For Group 1 chemical searches, name the group “Combined”
7. Select all references in the temporary Imported References folder, drag them into the newly created group.
8. Repeat for as many files as need to be imported into EndNote. The newly imported references will replace the previously imported references in the “Imported References” folder.
9. **Note: For ease of deduplication, import WOS searches last.** EndNote assigns each reference an ID number. Having higher ID numbers for Web of Science searches helps retain PubMed searches in EndNote’s automatic deduplication feature (see below). Otherwise, the order of importing does not matter.
 - a. **Note:** When Scopus searches are included, they should be imported last. Thus, the order of importing should be:
 - i. PubMed
 - ii. WoS
 - iii. Scopus (when included)

5.2 Peer-reviewed (and grey) literature deduplication, screening, and full-text retrieval

As noted, initial deduplication will be conducted at the time of literature search uploads into EndNote. The deduplication feature of Endnote is used to remove duplicates. This is followed by a preliminary manual screen in the reference management software to identify and remove any remaining duplicates, as well as any results that are not in English. When deduplicating, the PubMed entry is preferred, because it includes more meta-data than do other formats. Record the final number of unique references after all duplicates have been removed. Document the articles removed in this screen and the basis for the screen. Following this preliminary screen, further screening and tagging is done with the Distiller software.

[Note – UC has “one-click” downloading from a large number of journals via Endnote, which is why the pre-screening is in Endnote. Other organizations may wish to move to Distiller for the pre-screening.]

The process used for removing duplicates in EndNote is designed to preserve a record of the references that were removed at each stage prior (see “how to” section below). First, results of each search are individually loaded into EndNote and added to a group that indicates which search it came from. This EndNote library is saved as an archive of all results from all searches, prior to any deduplication and screening (See Table 4 below for naming scheme of EndNote files). Second, all references from the archive EndNote library are copied to a second EndNote library for deduplication using EndNote’s Find Duplicates feature. The Find Duplicates feature should be set to match references on title and year. Trials using author, title, and year found that including author as a criterion for duplicates makes EndNote miss many duplicates due to minor differences in how author names are recorded (e.g., first initial or full first name).

To change this setting:

1. Go to Edit > Preferences.
2. On the left side, click “Duplicates”.
3. In the box under “Compare references based on the following fields,” ensure that Author is NOT checked, and that Year and Title ARE checked.
4. Click Apply, then OK to save the changes.

Duplicates found by EndNote are first copied into a third EndNote library as a record of what was removed by deduplication. Remaining references are then manually screened (see “how to” section below”) for a few additional criteria: 1) any duplicates that EndNote missed in its automatic search, and 2) references that are not in English. For PHOPs, a third category was also screened at this stage, for references that clearly have nothing to do with chemical risk assessment. For all other classes, screening based on relevance is left for the first round of screening (title/abstract) in Distiller. For this third category, screening for the PHOPs was highly

conservative, leaving in references that could plausibly be relevant. A few examples of references excluded using this criterion from the PHOP class are presented in the Table 5 below.

Table 2. Names of EndNote libraries generated at each step of import, deduplication, and initial screen in EndNote

Stage	EndNote library	Purpose
Initial import into EndNote	CLASS NAME_Archive_DATE	Retain a record of all references found by all aspects of search strategy
Deduplication (automatic and manual) and initial screen (language, conference proceedings, etc.)	CLASS NAME_removed	A record of references removed by deduplication and initial screen
Deduplication date	CLASS NAME_dedup_DATE	The deduplicated list of references found for an OFR class, for screening in Distiller

Table 3. Examples of references removed in initial screen of PHOP class peer-reviewed literature for purposes of relevance, with notes on reasons and DOI for reference

DOI	Reference	Note
10.3390/e22010046	Kurthen, M. and T. Enßlin (2019). "A Bayesian Model for Bivariate Causal Inference." <i>Entropy</i> (Basel) 22(1).	Uses a dataset called Tuebingen Cause Effect Pairs (TCEP)
10.1080/15592324.2015.1044192	Li, S. (2015). "The Arabidopsis thaliana TCP transcription factors: A broadening horizon beyond development." <i>Plant Signal Behav</i> 10(7): e1044192.	Review of a family of transcription factors called TCP transcription factors

10.3390/brainsci10080522	de Carvalho, P. S. T., et al. (2020). "Feasibility of Using Intraoperative Neuromonitoring in the Prophylaxis of Dysesthesia in Transforaminal Endoscopic Discectomies of the Lumbar Spine." <u>Brain Sci</u> 10 (8).	Includes a term, transcranial motor evoked potentials (TCEP)
--------------------------	--	--

In each of these instances, acronyms for chemicals of interest are used for unrelated terms. Some general categories of references that were NOT excluded at this stage include: those discussing general chemistry, manufacturing, engineering, or performance of flame retardants; those which discuss non-flame retardant uses of flame retardant chemicals, including those discussing their use as reagents in chemistry or biochemistry studies. As with references removed by EndNote's Find Duplicates feature, references removed at this stage are also copied to a new EndNote library and saved as documentation.

Thus, there are 3 EndNote libraries that result from this process: 1) an archive library including all references found from all searches for a class, 2) a library containing references removed by automatic deduplication, manual deduplication, and an initial screen for language and conference proceedings in EndNote, 3) a library containing the remaining references after deduplication and screening. The first 2 of these libraries serve as documentation of the process.

Note that ideally, the final deduplicated and screened library of peer-reviewed literature are combined with the final library of grey literature (described in section 5.3 below) to upload a single .ris file into DistillerSR. This combined library should also be deduplicated both automatically and manually. It is possible that the same reference is identified in both the peer-reviewed and grey literature search process (e.g., a review may be found in a bibliographic database and via searching the Cochrane website). Gathering PDFs using EndNote's Find Full Text feature is done after title/abstract screening in DistillerSR, and thus does not need to be done at this point.

5.2.1 How to remove duplicates and screened-out references in EndNote, while preserving a record of individual searches and their results

1. Create a new EndNote library titled "CLASS NAME_dedup".
2. From the "CLASS NAME_Archive" library created in Section 5.1.5, click "All references" along the left side. Select all references and drag them into a new "CLASS NAME_dedup" EndNote library. Close the Archive library.
3. In the "CLASS NAME_dedup" library, in the top menu click References > Find Duplicates. EndNote will present a screen comparing duplicates one-by-one.
4. To more quickly delete many references, click Cancel in the top right of the Find Duplicates screen that EndNote presents. EndNote will then return to a list of all duplicates, with one copy highlighted. To remove all highlighted references at

once, in the top menu click References > Move References to Trash.

1. **This is why importing WOS/Scopus searches last in the previous section is important. If imported last, the highlighted references should all be WOS/Scopus references**
5. After automatic deduplication, a second round of manual deduplication will be needed (described in initial screen section below). EndNote sometimes misses duplicates that a human will catch.
6. **Sort the dedup library by title. This will help in identifying duplicates missed by EndNote's Find Duplicates feature.**
7. Screen by title alone when possible. Check against abstract or other reference info if unsure. If still unsure, do not remove the reference. Particular care is needed when deduplicating grey literature; see Section 5.2.2.2.
8. Mark any references to be screened out with 5 stars. This is used to sort and copy these references to a separate EndNote library (CLASS NAME_removed_DATE) to be kept as a record.
9. For each reference, screen for the following criteria:
 1. Duplicates missed by EndNote.
 1. If two articles with the same or similar title remain, check the authors, the journal, the volume and page numbers to verify it is a duplicate
 2. Common situations in which EndNote misses duplicates:
 1. A symbol vs. the name of that symbol spelled out
 1. E.g., β vs. beta
 2. Differing naming conventions for isotopes
 1. E.g., C – 14 vs. [14C]
 3. The number of dashes used
 1. E.g., a single dash with spaces “ - ” vs. a double dash with no spaces “--”
 2. **As an extra safeguard, in addition to screening titles, search the EndNote library for this double dash**
 4. Missing word(s) from the end of a title
 3. Note: priority is to keep entries from PubMed over Web of Science. An easy way to tell which is from Web of Science is to look in the URL field. References from Web of Science have the URL field of the format:
 1. <Go to ISI>://WOS:000383930400110
 2. Non-English
 1. Many references that are not in English have brackets around the titles. Other references can be checked by clicking on the reference and scrolling to the very bottom of the reference information panel to find the Language field. If non-English language, remove now. If not indicated, this is screened later.
 3. Conference proceedings
 1. Manually search for Conference proceedings, which are identified by checking the titles of references and by searching the EndNote

library for the search terms: conference, proceedings, symposium, meeting. It is not always obvious from title that a reference is a conference proceeding. If it is obvious, remove now. If not obvious, this is screened later.

10. Once all references have been screened, sort by rating. Select and copy all screened references (5-starred references) to a new library titled CLASS NAME_removed. After copying, delete these references from the dedup library. The dedup library is now ready to be sent to DistillerSR. The process for exporting a library in the proper file format is described below.
11. In case of an error, references in trash can be removed by clicking on Trash near the top of the left panel. Within the Trash, select the references to be restored to the library. With the desired references selected, in the top menu click References > Restore to Library.

5.2.2 Deduplicating peer-reviewed against previous classes and grey literature against previous classes and against peer-reviewed literature

Note for the first class, this section can be skipped. An .ris file for import into DistillerSR can be created directly from the original Class 1 EndNote Library (Section 5.2.3).

Note that deduplication should be performed prior to uploading any references to Distiller for Level 1 screening.

5.2.2.1 Automatically deduplicating peer-reviewed and grey literature against previous classes, prior to title/abstract screening in DistillerSR, and creating an .RIS file

1. Create an EndNote Library for each class search of original results and deduplicate across literature databases. Save as a backup. (For more on creating an EndNote library, see section 5.1.5 of this guide; for more on EndNote deduplication, see section 5.2.1 of this guide).
 - i. The EndNote library for each class that serves as the “input” for these steps is CLASS NAME_dedup_DATE, as described in Table 4 and resulting from the process in 5.2.1)
2. Create a master EndNote Library for deduping across class searches (this master library only needs to be created once).
 - i. Copy Class 1 search results to the master EndNote Library.
 - ii. Create a Group for Class 1
 - iii. Import unique references into Distiller.
3. Run a search for a new class, import new class search results into an EndNote Library and remove duplicates between literature databases, save as back up. (This is the same process as step 1 here, but for a subsequent OFR class; see sections 5.1.5 and 5.2.1 of this guide).
4. Copy new class search results to the master EndNote Library – leave in Unfiled.
 - i. Note: include both peer-reviewed literature and grey literature
5. In master EndNote Library, deduplicate between classes. After running Find Duplicates, duplicates found in the new class results are highlighted; deleting highlighted references removes those references from Unfiled & Duplicate References. (For more on the EndNote deduplication process, see section 5.2.1 of

this guide).

- i. Remaining references in Duplicate References are the record already associated with previous classes
 - ii. Copy these refs into a Group for the new class. This set of references is now in (at least) two groups: at least one previous class, and the new class.
 1. This can be checked by highlighting a reference > References > Reference Summary > Custom and Smart Groups will name each group a reference is in.
6. **To ensure that all duplicates are caught, in the master EndNote Library, skim titles of newly added references for duplicates from previous classes.** When skimming, **sort the “All references” folder by title** and focus on 1) titles with a) numbers and symbols (e.g. 1 and one, and alpha and α will be missed as duplicates by EndNote), b) isotopes or other similar notations, c) dashes (single and/or double), d) missing words from the end of a title (See Section 5.2.1 step 9 for more detail), and 2) grey literature sources. Additional suggestions for deduplicating grey literature sources can be found in Section 5.2.2.2; particular care is needed when deduplicating grey literature, since the title may vary. Confirm by checking in Distiller and against files from the previous class. Depending on the potential duplicate identified, the following may be checked:
 - i. Check in Distiller Datarama reports. Include data from Level 1 screening. Search for specific text in titles, authors, etc. under Reference Criteria. If a likely overlap from a previous class, then proceed to ii.
 - ii. Check files on OneDrive, for the “upload class” indicated by the likely overlap from Distiller. When checking against files, check both the title AND the PDF itself. Verify the text is actually the same before eliminating.
 - iii. Delete confirmed additional duplicate references from Unfiled & Duplicate References.
 - iv. Remaining references in Duplicate References are the record already associated with at least one previous class.
 - v. Copy these refs into the set of references the Group you made in the previous step for the new class. These references are now associated with both the previous class(es) and the new class.
7. **Note that some references will remain in Unfiled.** These are the references that will be used to create an RIS file for import into DistillerSR (Step 8 of this section), **after additional steps are taken to remove duplicates EndNote may have missed (Step 6 of this section and Section 5.2.2.2 below).**
8. In the master EndNote Library, create an RIS file for references from the new class in Unfiled after completing deduplication processes outlined in Sections 5.2.2.1 and 5.2.2.2 (See Section 5.2.3 of the guide for how to create an RIS file). These are the references that are unique to the new class. (Note there is no need to assign a RefID in the Label field in EndNote. DistillerSR will automatically assign a RefID to newly uploaded references.)
 - i. This RIS file is used to import the unique references for the new class into DistillerSR. Send to appropriate DistillerSR admin for upload.
9. In master EndNote Library, move the new class references from Unfiled to a

Group for the new class. All references from the new class are now in the Group for the new class (regardless of whether they were identified in previous classes or not).

5.2.2.2 Considerations for manually deduplicating grey literature against previous classes and peer-reviewed literature, prior to title/abstract screening in DistillerSR.

Particularly when comparing grey literature files across classes, it is important to check BOTH the title (metadata and file name) AND the contents of the PDF itself. Since grey literature metadata are being manually generated (as opposed to open literature, where metadata is retrieved from a bibliographic database), it is important to verify that suspected duplicate files are indeed duplicates. Ideally the following steps will be checked as early in the process as possible (prior to uploading any references to Distiller for Level 1 screening), but can be applied if duplicates are suspected later in the process as well.

Special attention should be paid to grey literature documents sourced from NTP and from the Cochrane database. While all grey literature should be checked against the peer-reviewed literature, documents from these two sources are perhaps most likely to be found by both peer-reviewed and grey literature searching. For NTP, note that author metadata is likely to differ between sources. For example, an NTP report may have any of the following (or others) in the author field:

- NTP
- National Toxicology Program
- Program, National Toxicology
- Or even a blank field for author metadata, with NTP (or a variation) being indicated in another metadata field

Variation like this can make deduplicating manually in EndNote more difficult. For this reason, grey literature files are best checked against other classes outside of EndNote, by comparing the PDFs themselves. As with titles, authors, etc., file names will likely vary. However, the title can often indicate whether the PDF is likely to be specific to one OFR or class, or cover multiple classes. Looking at folders of PDF files may make it easier to limit to just grey literature, instead of having mixed grey literature and peer-reviewed literature, as is the case in an EndNote library. This is especially true as the “master” EndNote library progressively grows with the addition of each subsequent class.

- 1) Manually deduplicating grey literature against grey literature from previous classes.
 - a) Outside of Distiller
 - i) Both reference metadata (EndNote) and the PDFs themselves can be used to check recently downloaded grey literature PDFs against grey literature PDFs from previous classes.

- (1) In EndNote, automatic deduplication will only identify exact duplicate titles. Since metadata for grey literature are being manually generated, there are likely to be small differences in titles. Even when titles can be copied and pasted from PDFs into EndNote, the possibility for variation still exists. For example, some grey literature titles are not informative enough on their own, and so relevant CAS numbers are appended to titles. If the formatting on this is not perfectly consistent (whether between different people, or for the same person working across different classes at different times), titles may thus be different.
 - ii) In File Explorer/OneDrive: The use of common file naming conventions may assist deduplication by looking at the file names themselves, in addition to the reference metadata in EndNote.
 - b) Using Distiller to aid in comparing with previous classes
 - i) It is possible to “search” within Distiller using Datarama. To do so:
 - ii) In the top menu of Distiller, select Reports > Datarama.
 - iii) In the Report Settings Tab, Data to Display box, select Levels 1 and 2.
 - (1) Level 3 may be included, but may not be necessary.
 - iv) In the Advanced Options tab, Advanced Options box, check the box for “Show all versions of form data”
 - v) In the Reference Criteria tab, Reference Fields box, in the “Search for Text” box, enter terms to find the reference(s) of interest
 - (1) This enables searching title, author, and abstract for the search terms. Note that this does not include the comments or notes fields for user-generated notes.
 - vi) In the Reference Display Options tab:
 - (1) Reference Fields box, check the box next to:
 - (a) Author
 - (b) Title
 - (c) Abstract (if desired)
 - (d) Year
 - (e) Other fields may be added as needed, but author, title, and year will often be sufficient.
 - (2) Reference Labels box:
 - (a) Select “Display labels”
 - (b) This will display the “upload labels,” which indicate which class’ **search** a reference came from. This is helpful in identifying which RefIDs are associated with which class search.
 - vii) Run the report by clicking the green “run report” button.
 - viii) Using the generated Datarama report, compare potential duplicates
 - (1) For suspected duplicates, compare contents of PDFs to confirm.
 - (2) When a duplicate is confirmed, also verify that it has been tagged for the appropriate class
 - (a) E.g. if a reference was screened in Class 1 and identified again in searches Class 4, verify that Class 4 was also included in the chemical class tags during Level 2 screening for Class 1 (or if not, why not – e.g., the chemical was evaluated but concentrations were below the detection limit)
- 2) Deduplicating grey literature with peer reviewed literature, both within a class and across previous classes

- a) Many of the same strategies outlined above apply. However, this is an additional challenge, particularly prior to title/abstract screening, because there are not yet PDF files for peer-reviewed literature to compare against. Thus, if there is uncertainty about a reference being a duplicate between grey and open literature, it should be retained until full text is retrieved after title/abstract screening. Additional checks for duplicate files are in place during PDF retrieval (Section 5.3.1).

5.2.3 How to export an EndNote library as a .ris file, for import into DistillerSR

All deduplicated English-language references are uploaded into Distiller for title/abstract screening. References are uploaded to DistillerSR as an .RIS file. Note that the full-text PDF is not required at this stage, and will be retrieved after completion of title/abstract screening in Distiller. Note also that grey literature (Section 5.4) is to be uploaded to Distiller at the same time as peer-reviewed literature.

1. With the EndNote library to be exported open, click File > Export.
2. Choose where to save the exported file.
3. Give the exported file a name. **Note:** The .ris file extension must be manually typed at the end of the file name. Without this, the file will just be saved as a standard .txt file.
4. Ensure “Save as type” displays Text File (*.txt)
5. Ensure “Output style” displays RefMan (RIS) Export.
 - a. If RefMan (RIS) Export is not among the available options in the dropdown menu, click “Select Another Style”
 - b. Find RefMan (RIS) Export either by scrolling the alphabetical list, or using the quick search box and pressing enter
 - c. Click RefMan (RIS) Export in the list, then click Choose.
6. Uncheck “Export Selected References”
7. Click Save

5.2.4 How to retrieve PDFs for references that pass DistillerSR Title/Abstract Screening

The Screener Guide describes in detail the inclusion/exclusion criteria for screening in Distiller SR, as well as the tagging tree. As described there, the title/abstract screening (Level 1) considers whether the data source is related to human health risk assessment as defined by the PECO statement, and whether it could contain information on any OFR. This screening is done prior to retrieval of PDFs. After all PDFs have been retrieved, they are transferred to ICF for highlighting of key words related to chemical classes, as well as toxicity and exposure tags.

For all references that pass the title/abstract screening in DistillerSR, the next step is to retrieve the relevant PDFs. These PDFs are then used for the full-text screening and tagging stage, which is done in DistillerSR. **Throughout this section, note the QA considerations outlined in Section 5.3.**

1. PDF retrieval happens in several steps (outlined overall here, and in more detail below).

Retrieved PDFs are named to match the PDF naming format from EndNote: RefID-FirstAuthorLastName-Year (e.g., 453-Haber-2021). While overall consistency of filenames is important, **it is crucial that the PDF file names start with the RefID number**. First, export an .RIS file from DistillerSR for the references that passed title/abstract screening.

2. This .RIS file is imported into EndNote, using an EndNote import filter provided by ICF.
 1. This import filter maps DistillerSR's unique RefID into EndNote's Label field, thus preserving each reference's unique identification number. This is **extremely important**, as the RefID number is used to automatically match PDFs to DistillerSR records for the correct reference.
3. The first round of PDF retrieval is done using EndNote's Find Full Text feature. Find Full Text automatically retrieves full-text PDFs when available, and URLs when PDFs are not available.
 1. Note that Find Full Text can only retrieve PDFs for batches of 250 references at a time. In practice, smaller batches (~20) often work best.
 2. Note that Find Full Text never finds PDFs/URLs for all references. Thus, manual retrieval of additional PDFs is needed (see next step).
 3. There is an additional ongoing issue with Find Full Text related to retrieval of references from Elsevier/Science Direct and the American Chemical Society, and potentially other sources. EndNote support staff are aware of the issue, and were unable to resolve it. This issue further necessitates manual retrieval of PDFs.
4. The second round of PDF retrieval is done manually. URLs and DOIs from EndNote can be used to manually retrieve PDFs by searching in Google. Google Scholar can also be searched by titles. If that fails, the title can be searched in UC Library. Finally, PDFs that cannot be retrieved in any of the previous ways may be requested via Interlibrary Loan.
5. Retrieval of PDFs is tracked in an Excel tracking sheet, generated from a DistillerSR Datarama report.

5.2.4.1 Retrieving reference metadata from Distiller, and automatic full text retrieval using EndNote

1. In Distiller, in the menu bar, click Reports > Reference Status.
2. In the resulting table showing progress of references through different levels of review, click on the number in the "Included" cell.
3. This runs a Datarama report with all References that have been Included after Level 1 screening.
 - a. Note, this runs a report for all classes. If the number of references is large enough, Distiller will return a warning about exceeding the maximum of 20,000 cells that Distiller can display online. Click "Cancel" on this warning box and proceed to modify the Datarama report criteria in the following steps.
4. In the Datarama screen on the Report Settings tab, in the Basic Options box:
 - a. At the top of the box, for "Report Format" select "RIS"
 - b. At the bottom of the box, there is a line for "Filter References." Click the "Add Filter" button.
 - c. In the Filter References box that appears:
 - i. For "Filter Type", select "Labels Filter"
 - ii. Under "Labels", select the label that corresponds to the upload group of

interest. Here, it will be an OFR class.

1. Note that it is possible to select multiple labels. E.g., in the case of the Polyhalogenated Organophosphates class (PHOPs), for which open literature and grey literature were uploaded separately, it is possible to download metadata for both at once by selecting both corresponding labels.
- iii. Once the proper labels are selected, click “Add/Update” button
 1. Note it may be necessary to scroll down to see this button.
5. In the Datarama screen, on the Reference Criteria tab, in the Completion Status box (bottom right):
 - a. Ensure that “Included” and “Level 1, TiAB” are both selected.
6. At the bottom of the Datarama screen, click the green “Run Report” button to download an RIS file.
 - a. **Do not close this browser tab/window** – it can be used in the next section to quickly generate a tracking sheet for keeping track of the PDF retrieval process.
7. Ensure that the newly downloaded RIS file and the Import Filter provided by ICF are in the same folder. Also in this folder create a new EndNote library.
8. Open the newly created EndNote library. Then open the Import Filter. Then open the RIS file in EndNote.
9. Check EndNote’s Auto-Rename settings
 - a. In top menu, Edit > Preferences
 - b. Among the options along the left side of the Preferences window, select PDF Handling.
 - c. Within PDF Handling, under PDF Auto Renaming Options, select “Custom”
 - d. In the field for entering custom renaming options, enter Label + Author + Year (Figure 29.)
 - i. Setting custom auto renaming settings allows EndNote to automatically rename retrieved PDFs so that Distiller can automatically associate PDFs with the correct existing reference in Distiller.
 - ii. Beginning with Label is the most important part. The Label field contains the RefID that Distiller uses to match PDFs to references.

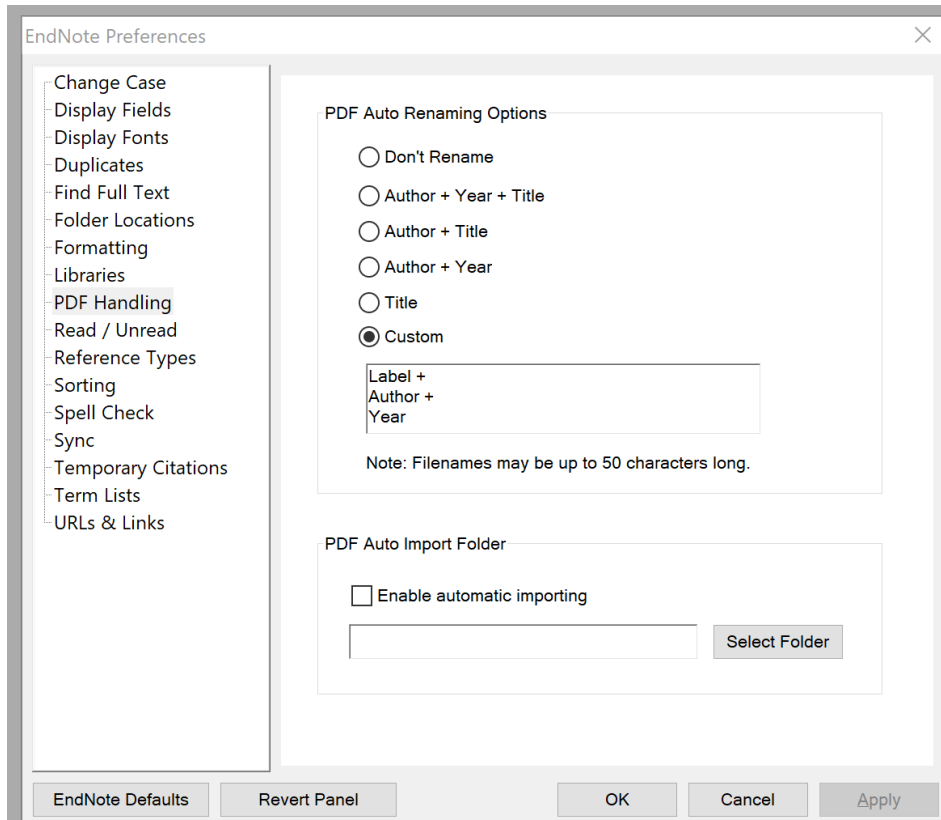


Figure 29. EndNote Preferences for PDF Auto Renaming

10. Use EndNote Find Full Text to retrieve available PDFs.
 - a. While Find Full Text is running, proceed to the next section to prepare for manual retrieval of PDFs that Find Full Text cannot retrieve automatically.

5.2.4.2 Creating a PDF retrieval tracking sheet from Distiller metadata, and manual retrieval of PDFs

While EndNote is automatically retrieving PDFs and URLs, create a tracking sheet for PDF retrieval using Distiller.

1. Return to the Datarama screen in Step 6a of the previous section:
 1. On the Report Settings tab, in the Basic Options box, change Report Format to Excel Spreadsheet.
 1. Occasionally, trying to download an Excel sheet from Distiller will produce an error. It is possible to work around this by downloading a CSV instead, and then saving it as an Excel workbook.
 2. In the Reference Display Options tab, in the Reference Fields box, select the following fields:
 1. Author
 2. Title
 3. Abstract
 4. DOI

5. Issue
 6. Journal
 7. Volume
 8. Year
3. Click the green “Run Report” button and save the resulting Excel spreadsheet with a descriptive file name.
2. In the resulting excel spreadsheet, add columns to track (using “1” to indicate yes/complete) whether PDFs were retrieved and placed into a single folder 1) via Find Full Text, 2) as grey literature PDFs downloaded as part of grey literature searching, 3) PDFs retrieved manually, 4), PDFs received via Interlibrary loan (ILL), or 5) references that were excluded. Adding a column for notes is often useful (e.g., for noting that an ILL request has been submitted, or for noting that a reference has been excluded (see below for exclusion reasons).
 1. If a PDF is found/received and any of the following conditions are met, it can be excluded from retrieval. Its exclusion should be tracked in the appropriate column, and a reason for its exclusion should be noted in the “Notes” column.
 1. Full text is not in English
 2. The reference is a conference proceeding
 1. Note that non-English references and conference proceedings are screened out prior to uploading references to Distiller for Level 1 screening, but these conditions are not always apparent from reference metadata available at that stage.
 3. Grey literature references that have been screened in a previous class
 1. Extra care should be taken here. Many grey literature sources have similar titles, and citation information for grey literature sources has been manually generated (so the same reference may have slightly different titles if collected for 2 different classes). If it is suspected that a grey literature file has been screened before, it should be verified in two ways, and the **contents** of the PDF should always be checked in addition to the title
 1. Checking the grey literature download files for previous classes.
 2. Checking Distiller records via Datarama
 1. Note that titles (or portions of titles) can be searched in Distiller.
 1. Reports > Datarama > Reference Criteria tab > Reference Fields box > Search for Text field
 2. See “?” icon next to Search for Text box for additional info on what can be searched.
 2. If a grey literature PDF has been found in the folder for a previous class, its RefID may be found by searching for the version of the PDF uploaded to Distiller.
3. **Copy** previously gathered grey literature PDFs to the central PDF folder, and **rename** them to match the file format.

4. Once Find Full Text is complete for all references, copy the PDFs from EndNote's internal files (in the .Data folder, PDFs subfolder) into a single central folder. Track that these PDFs have been retrieved and moved to the proper folder.
 1. Note that these PDFs will already be properly renamed with RefID at the beginning of the file name. All other PDFs should mimic this file naming format.
 1. E.g., for a reference with RefID 150: 150-Smith-2020
5. Once 1) previously gathered grey literature PDFs have been copied, and 2) Find Full Text is complete for all references, any remaining references without a PDF will need to be manually retrieved. Retrieve PDFs manually, using DOIs or URLs provided by Endnote (or in the tracking sheet), by searching titles in the UC library or on Google Scholar.

AFTER SEARCHING FOR AND IDENTIFYING PERTINENT PEER-REVIEWED LITERATURE AND DEDUPLICATING AS DESCRIBED IN SECTION 5.2.2, PERFORM ADDITIONAL QA STEPS OUTLINED IN THE NEXT SECTION (5.3) DURING AND AFTER PDF RETRIEVAL.

5.3 QA Review of PDF Sources

The following text outlines various steps to ensure quality of PDFs uploaded to DistillerSR for screening and extraction. Ideally, these steps are conducted as early in the workflow as possible, with later steps acting as an additional check. That is, grey literature PDFs and metadata should be checked against previous classes prior to merging grey literature with open literature, and any grey literature documents that were screened in previous classes should not be uploaded to Distiller for full text (Level 2) screening. A similar check is conducted prior to uploading references to DistillerSR for title/abstract screening. – This additional check prior to Distiller Level 2 screening is an extra precaution – ideally, all “across-class” duplicates will be detected prior to uploading reference metadata for title/abstract screening). (Note that grey literature PDFs are downloaded as part of the initial downloads outlined in Section 5.4. These PDFs will already be downloaded at this stage.)

5.3.1 QA checks during PDF retrieval after DistillerSR title/abstract screening and prior to full text screening.

Despite the QA steps above, it is possible that some duplicate references will still be uploaded to Distiller and pass title/abstract screening without being recognized as a previously screened reference. When this occurs, the next likely opportunity to detect duplicates is during retrieval and collection of PDFs for highlighting and then for upload into Distiller prior to the full text screen (Level 2).

When duplicates are detected here,

1. Confirm with the full text. If confirmed, the full text does not need to be sent for highlighting and uploading into Distiller.

2. To handle the “extra” record in Distiller, confirm the duplicate has indeed been screened, and includes tags for the appropriate class(es) using the manual duplicate check process outlined above in Section 5.2.2.2.
 - a. If the duplicate and appropriate class tags are confirmed:
 - i. The original reference entry in Distiller should be retained.
 - ii. The duplicate entry can be excluded immediately at Level 2 without highlighting or uploading the PDF.
 1. To exclude, choose “no” and enter the following into the comment field:
 - “Previously screened as RefID [###]”

In addition, non-English full text and conference proceedings are initially screened out while preparing metadata (.RIS file) for upload into Distiller for title/abstract screening. However, this is not always apparent from the information available at that time (reference metadata provided by bibliographic databases).

1. Thus, these should be checked again while retrieving PDFs prior to highlighting and full text screening.
 - a. For Non-English language full text: do not download PDF. Make a note in PDF retrieval tracking sheet that the full text is in a language other than English.
 - i. In DistillerSR, these will be excluded with a comment of the standard format:
 1. “Exclude: non-English”
 - b. For conference proceedings (e.g., abstract for a poster presentation, a platform presentation, or a conference session): do not download PDF. Make a note in PDF retrieval tracking sheet that this is a conference proceeding.
 - i. In DistillerSR, these will be excluded with a comment of the standard format:
 1. “Exclude: conference proceeding”

5.3.2 Additional QA steps after PDF retrieval

After retrieval of PDFs, some quality assurance steps are needed, particularly due to the potential for error in manual PDF retrieval. Key QA tasks include:

1. Verify only one PDF exists per RefID
2. When there are multiple references with the same first author’s last name, check EACH PDF to ensure the correct PDF is mapped to the correct reference with the correct RefID.
3. Spot check ~10% of the total number of PDFs to verify the correct PDF is mapped to the correct reference with the correct RefID. When spot checking, avoid the already checked references with the same last name of the first author.

5.3.3 Keyword highlighting of PDFs

After QA on PDFs, PDFs from peer-reviewed literature and from grey literature are sent to ICF for automated highlighting of key terms. Note that long grey literature PDFs (> ~50 pages) are not sent for highlighting, due to the intense resource requirement for highlighting these large files. The time required for highlighting scales with number of references, number of key terms to be highlighted, and the length of the PDF. During screening in Distiller, many of the longer grey literature PDFs can be quickly scanned using the table of contents and tables for relevant information. This means that the benefit of highlighting these large files is often small relative to the large amount of computing resources needed.

5.3.4 PRISMA diagrams for each class

While DistillerSR can generate PRISMA diagrams, it does so only at the level of the entire project, and cannot be filtered by upload label. Thus, PRISMA diagrams for this project are generated manually for each class, pulling data from EndNote libraries, Distiller, the Distiller data download, and evidence maps. **Table 6** defines and identifies the source of information for each box in the completed PRISMA diagrams.

Table 4. Information definition and source for counts of references in PRISMA diagrams

PRISMA Box	Information Definition and Source
Records identified through database searching	Number of records identified through bibliographic database searching. Obtained from Archived EndNote library of all records retrieved from bibliographic databases, prior to any deduplication (See Section 5.2)
Additional records identified through other sources (grey literature)	Number of grey literature files identified by downloading according to Section 5.4 of this guide
Records after duplicates removed, screened at Levels 1 and 2	Total number of records uploaded to Distiller, identified from bibliographic databases and grey literature sources, deduplicated against each other and against previous classes. Obtained by counts of records by Distiller label (note some classes have more than one Distiller label).
Records excluded (Levels 1 and 2)	The difference between the number of records uploaded to Distiller (above) and the number of records evaluated for extraction at Level 3 (below).
Records evaluated for extraction (Level 3)	Number of unique references with Distiller data at Level 3, in Excel file exported from Distiller. Note that this exported file will include both references extracted at Level 3, and references excluded at Level 3.
Records excluded (Level 3)	The difference between the number of records evaluated for Level 3 extraction (above) and the number of records included in Level 3 evidence maps (below).

Records in Level 3 evidence maps	The unique count of references included in Level 3 evidence maps. Obtained from the Grand Total cell in Pivot Tables generated from Distiller data download (See Section 6.2.1). The count in this box will include references uploaded with previous classes, if they contain data relevant to the current class. For completeness of the PRISMA diagrams, both the number of references specifically from this class and the number of references from any class will be included in this box.
Records extracted and in Level 3B evidence maps	The number of records selected for Level 3B extraction. This will be either 25 for data rich classes, or 9 for data-poor classes, unless there is an insufficient number of references in data-poor classes (Section 6.2.2).

5.4 Grey literature searching

For some websites, identification of appropriate reviews or assessments is relatively straightforward, while for other, more complex, websites, deeper website searching is needed. We considered using a webscraping tool such as the Google API, but the cost of developing such a tool was prohibitive, and much of the same goals can be accomplished using targeted Google searches.

Meta-data are tracked in a document titled “!Grey lit tracking document-[CLASS NAME]+Analog.” The Meta-data collected vary by source, but in general include, as relevant: date of search/download, number of search result hits, and a link.

The grey literature searching is tailored for each of the sites/assessments being searched. In the list below, “identified from CompTox dashboard” refers to situations where the dashboard indicates that a relevant assessment exists. For the sites where a certain number of hits are noted to be screened, we stop if we locate the information of interest prior to screening that number (e.g., identifying assessments for all chemicals in the search).

- a. ATSDR Tox Profiles - search for chemicals/classes identified from CompTox dashboard
- b. EPA IRIS – search for chemicals/classes identified from CompTox dashboard
- c. EPA PPRTV – search for chemicals/classes identified from CompTox dashboard
- d. EPA TSCA/HPVIS data –search ChemView by CASRN Can do batch searching <https://chemview.epa.gov/chemview>
 - i. Note data obtained from ChemView is not to be sent to DistillerSR for screening, but is instead processed as database data.
 - ii. Note that although HPVIS data originally were downloaded from ChemView, it was later determined that more detailed information was needed. Thus a more detailed version was downloaded directly from

- HPVIS: https://iaspub.epa.gov/oppthpv/public_search.html_page
- e. General EPA: The above searching is supplemented by searching the National Service Center for Environmental Publications - <https://www.epa.gov/nscep> by CASRN. This captures all relevant EPA publications that have a document number. Search pages are screened until (1) all known EPA documents are identified (e.g., PPRTV and IRIS documents otherwise identified), and (2) an entire 15-reference page is screened with no new useful assessments identified.
 - f. EPA CompTox dashboard – this is done as part of the initial setup for the class
 - g. Health Canada assessments - provided by CPSC for PHOPs. For future classes, do google site search: CASRN site:Canada.ca (There is no separate domain name for Health Canada.) Search only for chemicals identified by the CompTox Dashboard as being on the Canadian Domestic Substances List (DSL). Screen at least the first 40 hits for assessments, until a page is reached with no new assessments (at 20 hits/page). Note that assessments are often class-based. Note also that presence on the DSL does not guarantee the existence of an assessment.
 - h. ECHA – Do Google site search. CASRN site:echa.europa.eu Download assessments and include a link to the dossier. In addition to the link, dossier information will be downloaded as REACH Study Results obtained via IUCLID.
 - i. Note that ECHA assessments ARE to be sent to DistillerSR for screening, whereas dossier info obtained as REACH Study Results via IUCLID are NOT to be sent to DistillerSR, and are instead processed as database data.
 - i. NTP - a targeted search is done by publication type, using CASRN:
 - i. publications - <https://ntp.niehs.nih.gov/publications/index.html>
 - ii. Note: CEBS not included because that is more of a link to primary data. Similarly, WebICE is included as part of the database data search.
 - j. INCHEM is a gateway to a variety of WHO documents. Search chemical identity field by CASRN <http://www.inchem.org/#/search> Download any IPCS Environmental Health Criteria (EHC) documents, JECFA or JMPR assessments, and IARC assessments.
 - k. OECD existing chemicals database <https://hpvchemicals.oecd.org/UI/Search.aspx> Search by CASRN, download most complete available assessment (SIDS are more complete than SIARs)
 - l. Cochrane database of systematic reviews – <https://www.cochranelibrary.com/cdsr/reviews> search by broad term for class (e.g., phos)

Create a unique Endnote citation for each PDF obtained as grey literature. This citation should include a full, descriptive title, the name of the organization that authored the document, and the date. The full, descriptive title is especially important to assist screening in DistillerSR. Many of these sources do not have an abstract, and so are screened on title alone during title/abstract screening. When all of the grey literature citations have been entered into Endnote, an RIS file containing both the grey literature citations and the peer reviewed literature citations is created and uploaded to DistillerSR for Level 1 screening.

When grey literature hits are a detailed data table or database that is unique and does not overlap with data found in the database section (e.g., ECHA dossiers, or detailed HPVIS information), download the PDF. However, this does not need a citation and **does not need** to be uploaded to DistillerSR for screening as it is not a text-based PDF with a results and discussion section. Thus “screening and evaluation” as will be done in DistillerSR is not needed. These PDFs are handled along with database sources for the purposes of evidence mapping, rather than with text-based grey literature PDFs.

When a grey literature “hit” is simply the presence of the chemical’s name on a website or chemical list without a corresponding PDF (e.g., Health Canada has a website with a table listing chemicals on the DSL, but which provides no additional information), there is no need to generate citation information for these hits as this information is not of interest.

Where there is a large number of potentially data-rich analogs, preliminary screening will be done on the analogs prior to doing the analog searching.

5.4.0 How to obtain ATSDR Toxicological Profiles, and EPA IRIS and PPRTV documents

The existence of ATSDR Toxicological Profiles, EPA IRIS documents, and EPA PPRTV documents is shown on the Dashboard download. The Dashboard provides direct links for IRIS and PPRTV – the “Y” indicating the presence of a document will also be a link to a site to download that document. For ATSDR Toxicological Profiles, the Dashboard indicates in 2 ways that there may be documents: either directly indicating there is a Toxicological Profile for a specific chemical, or indirectly indicating there may be a Toxicological Profile for a group of chemicals by indicating the existence of one or more MRLs.

ATSDR Toxicological Profiles

1. In the Dashboard download, find the columns ATSDRPROFILES and ATSDRMRLS.
2. If ATSDRPROFILES indicates the presence of a document:
 - a. Go to <https://www.atsdr.cdc.gov/toxprofiledocs/index.html>
 - b. Scroll down to the “A-Z Index of Tox Profiles” and find the link to the Tox Profile by chemical name.
 - i. Keep in mind possible synonyms, e.g. see Dashboard-provided synonyms.
 - c. Click on the chemical name to go to the Tox Profile page.
 - d. Scroll down to the “Toxicological Profile Access” section, and download the “Complete Profile”
 - e. If the Dashboard indicates the presence of a Tox Profile, but no Tox Profile can be found, follow the steps below for ATSDR MRLs as a check.
3. If ATSDRMRLS indicates the existence of an MRL, this implies the existence of a Tox Profile. If there is one or more MRL, but no Tox Profile indicated, this may mean the MRLs come from a Tox Profile for a group of chemicals instead of for a single chemical. When ATSDRMRLS indicates the existence of an MRL:
 - a. Go to <https://wwwn.cdc.gov/TSP/MRLS/mrlsListing.aspx>
 - b. Search the page (using Control+F) for the associated CASRN.
 - c. Clicking on the associated chemical name will go to the Tox Profile page.

- d. Scroll down to the “Toxicological Profile Access” section, and download the “Complete Profile”
 - e. Example of MRLs from a Tox Profile for a group of chemicals:
 - i. Tris(2-chloroethyl) phosphate (CASRN 115-96-8) and Tris(1,3-dichloro-20propyl) phosphate (CASRN 13674-87-8) both have MRLs. Clicking on either chemical name goes to the Tox Profile page for “Phosphate Ester Flame Retardants,” which includes these two along with several other chemicals.
4. Each group Tox Profile only needs to be downloaded once.

EPA IRIS

1. In the Dashboard download, find the column IRIS_LINK.
2. For any chemicals with a “Y” in this column, follow the link (the “Y” is also a link) to the Dashboard page for that chemical.
3. In the menu along the left side of the page, scroll down to and click “Literature.”
4. In the list that appears, click “IRIS”
5. Download the PDFs of **both** the Toxicological Review and the IRIS Summary on the resulting page.

EPA PPRTV

1. In the Dashboard download, find the column PPRTV_LINK.
2. For any chemicals with a “Y” in this column, follow the link (the “Y” is also a link) to the Dashboard page for that chemical.
3. In the menu along the left side of the page, scroll down to and click “Literature.”
4. In the list that appears, click “PPRTV”
5. Download the PDF of the PPRTV document on the resulting page.

5.4.1 How to obtain EPA TSCA Data from EPA ChemView

1. Go to the following website: <https://chemview.epa.gov/chemview/>
2. On the ChemView home screen, click on the “CHEMICALS” tab above the search features, as shown in Figure 30.. This will allow you to input chemical information for the search.

3. In the search bar under “Chemical Name or Chemical Identifier”, the CAS number of a chemical can be inputted to add a chemical to the batch search.

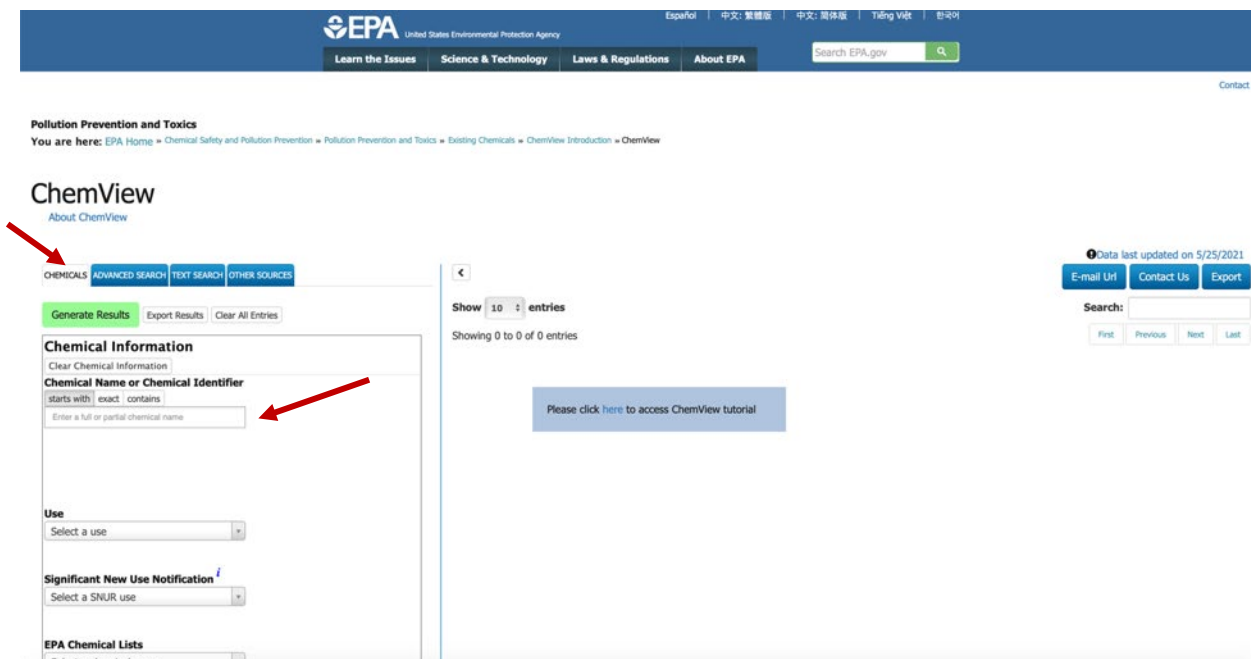


Figure 30. ChemView homepage

4. When typing the CAS number, a dropdown will display the CAS number that was inputted. Select the number from the dropdown for it to be added to the batch search, as displayed in Figure 31.. Each chemical has to be entered individually, but the search can be performed with a batch of chemicals.

CHEMICALS **ADVANCED SEARCH** TEXT SEARCH OTHER SOURCES

Generate Results Export Results Clear All Entries

Chemical Information
Clear Chemical Information

Chemical Name or Chemical Identifier
starts with exact contains
76025-08-6
starts with : 76025-08-6
76025-08-6

Use
Select a use

Significant New Use Notification
Select a SNUR use

EPA Chemical Lists
Select a chemical group

Figure 31. Example of inputting a chemical in ChemView

- After inputting all of the chemicals, click on the “ADVANCED SEARCH” tab and select the green button labeled as “Generate Results”. The results for the search will now be displayed on the right side of the screen, as illustrated in Figure 32..

ChemView
About ChemView

CHEMICALS **ADVANCED SEARCH** TEXT SEARCH OTHER SOURCES

Generate Results Export Results Clear All Entries

Chemical Information
Clear Chemical Information

Chemical Name or Chemical Identifier
starts with exact contains
Enter a full or partial chemical name

Already selected:
[remove] starts with : 10235-09-3
[remove] starts with : 10419-79-1
[remove] starts with : 1067-98-7
[remove] starts with : 1070-42-4
[remove] starts with : 111883-35-3
[remove] starts with : 113289-02-0
[remove] starts with : 13074-89-0
[remove] starts with : 13074-92-5
[remove] starts with : 137888-35-8
[remove] starts with : 14622-42-5
[remove] starts with : 15429-92-2
[remove] starts with : 17077-54-2
[remove] starts with : 1779-45-9
[remove] starts with : 183997-41-3
[remove] starts with : 21661-76-7
[remove] starts with : 22945-76-2
[remove] starts with : 25131-74-2
[remove] starts with : 2799-58-8
[remove] starts with : 3040-58-0
[remove] starts with : 3055-74-1
[remove] starts with : 31066-45-2
[remove] starts with : 36366-20-8
[remove] starts with : 37788-50-4
[remove] starts with : 37884-64-8
[remove] starts with : 42983-32-4
[remove] starts with : 42983-34-8
[remove] starts with : 45733-59-2
[remove] starts with : 494201-57-9
[remove] starts with : 51287-49-1

Chemicals Cases

Data last updated on 5/25/2021

E-mail URL Contact Us Export

Show 10 entries

Structure	Chemical Name / Chemical Identifier	Information Submitted to EPA	EPA Assessments	EPA Actions	Manufacturing, Processing, Use or Release
	Phosphoric acid, mixed 3-bromo-2,2-dimethylpropyl and 2-bromoethyl and 2-chloroethyl esters 125997-20-8	View for All	View for All	View for All	View for All
	Phosphoric acid, bis(2-chloro-1-methylethyl) 2-chloropropyl ester 76025-08-6				
	Phosphoric acid, 2-chloro-1-methylethyl bis(2-chloropropyl) ester 76649-15-5				

Showing 1 to 3 of 3 entries

First Previous 1 Next Last

Please click [here](#) to access ChemView tutorial

Figure 32. Example of search results for a ChemView batch search

- Any results relevant to TSCA will be in the “EPA Actions” column. The results can be exported into an Excel spreadsheet by selecting the “Export Results” button on the top

left side of the screen, as shown in Figure 33.. This will download an Excel spreadsheet with links to relevant results. The spreadsheet format is illustrated in Figure 34..

ChemView
About ChemView

CHEMICALS ADVANCED SEARCH TEXT SEARCH OTHER SOURCES

Generate Results Export Results Clear All Entries

Chemical Information
Clear Chemical Information

Chemical Name or Chemical Identifier
starts with exact contains
Enter a full or partial chemical name

Already selected:
[remove] starts with : 10235-09-3
[remove] starts with : 10419-79-1
[remove] starts with : 1067-98-7
[remove] starts with : 1070-42-4
[remove] starts with : 111883-35-3
[remove] starts with : 112399-82-0
[remove] starts with : 13074-89-0
[remove] starts with : 13074-92-9
[remove] starts with : 137888-35-8
[remove] starts with : 14622-42-5
[remove] starts with : 15429-92-2
[remove] starts with : 17677-54-2
[remove] starts with : 1779-45-9
[remove] starts with : 183997-41-3
[remove] starts with : 21661-76-7
[remove] starts with : 22945-76-2
[remove] starts with : 25131-74-2
[remove] starts with : 2799-58-8
[remove] starts with : 3040-58-0
[remove] starts with : 3055-74-1
[remove] starts with : 31066-45-2
[remove] starts with : 36366-20-8
[remove] starts with : 37789-50-4
[remove] starts with : 37894-64-8
[remove] starts with : 42983-32-4
[remove] starts with : 42983-36-8
[remove] starts with : 45733-58-2
[remove] starts with : 494201-57-9
[remove] starts with : 51287-49-1

Chemicals Cases

Data last updated on 5/25/2021

E-mail Us Contact Us Export

Show 10 entries

Structure	Chemical Name/ Chemical Identifier	Information Submitted to EPA View for All	EPA Assessment s View for All	EPA Actions View for All	Manufacturing, Processing, Use or Release View for All
	Phosphoric acid, mixed 3-bromo-2,2-dimethylpropyl and 2-bromoethyl and 2-chloroethyl esters 125997-20-8				
	Phosphoric acid, bis(2-chloro-1-methylethyl) 2-chloropropyl ester 76025-08-6				
	Phosphoric acid, 2-chloro-1-methylethyl bis(2-chloropropyl) ester 76649-15-5				

Showing 1 to 3 of 3 entries

First Previous 1 Next Last

Please click [here](#) to access ChemView tutorial

Figure 33. Example of search results being downloaded from ChemView

A1	CAS Number	Chemical Name	Structure	8E	SNUR	CO	12(b)	8D	CDR
1	125997-20	Phosphoric acid, mixed 3-bromo-2,2-dimethylpropyl and 2-bromoethyl and 2-chloroethyl esters							
2	76649-15-5	Phosphoric acid, bis(2-chloro-1-methylethyl) 2-chloropropyl ester							
3	76025-08-6	Phosphoric acid, 2-chloro-1-methylethyl bis(2-chloropropyl) ester							
4									
5									
6									
7									

Figure 34. Example of exported results from EPA ChemView

- In addition to exporting the results, a PDF of the results from the EPA assessments column can be saved as well. Click on the “View for All” button in the “EPA assessments” column heading. This will open up a separate window with the results for that column. In the top left, click on the “Print” icon and save the document as a PDF, as shown in Figure 35..

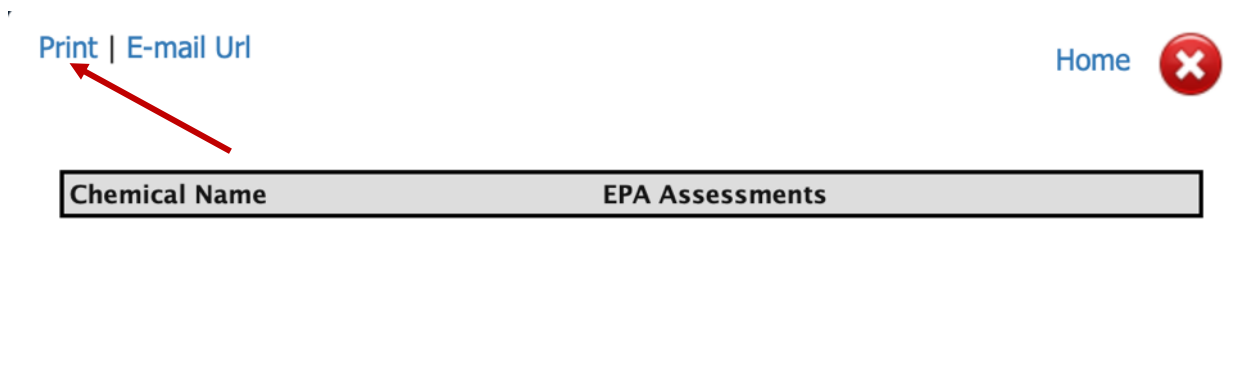


Figure 35. Example of EPA Assessments window to save as a PDF document

5.4.2 How to get detailed HPVIS information

1. Enter CAS number on this page: https://iaspub.epa.gov/opptppv/public_search.html_page
The field for searching by CAS number is about halfway down the page, under “Chemical Selection” header. The field is titled “Chemical (CAS No. – Chemical Name; **Figure 36.**)

Figure 36. Example of entering a CAS number to search HPVIS

2. Scroll to the bottom of the page to click the green “Search” button.
3. The search result will look like **Figure 37.** (Note: The bottom of this page is similar to the more basic HPVIS information downloaded from ChemView in Section 5.4.2.1 of this guide – not shown here)

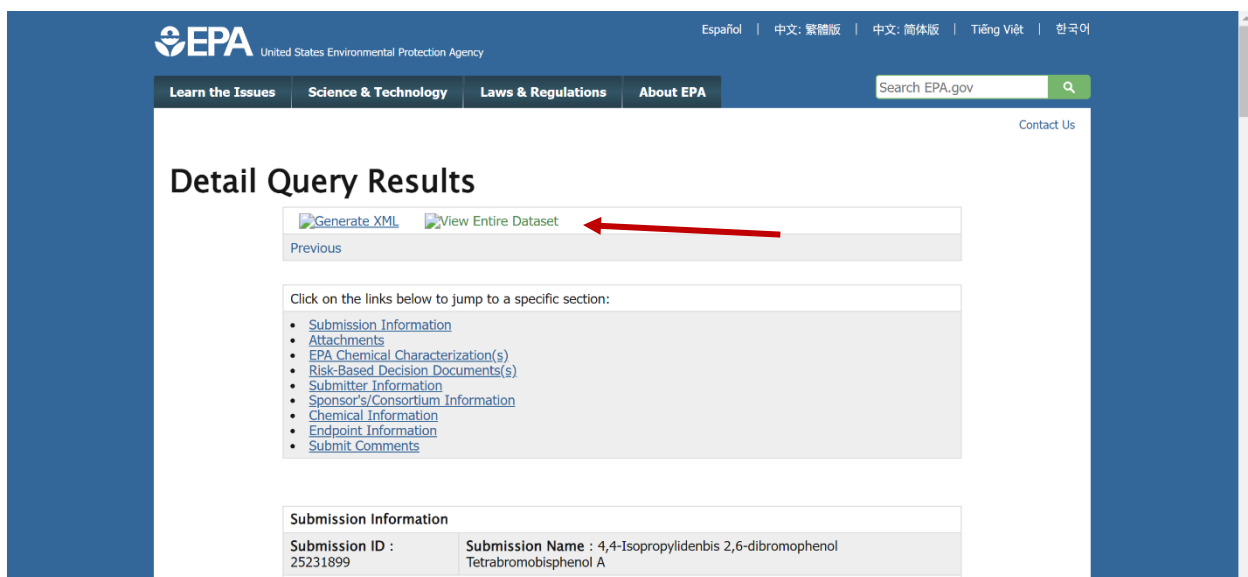


Figure 37. How to access the "Entire Dataset" in HPVIS

- Then click on "View Entire Dataset" at the top, and get something like Figure 38.

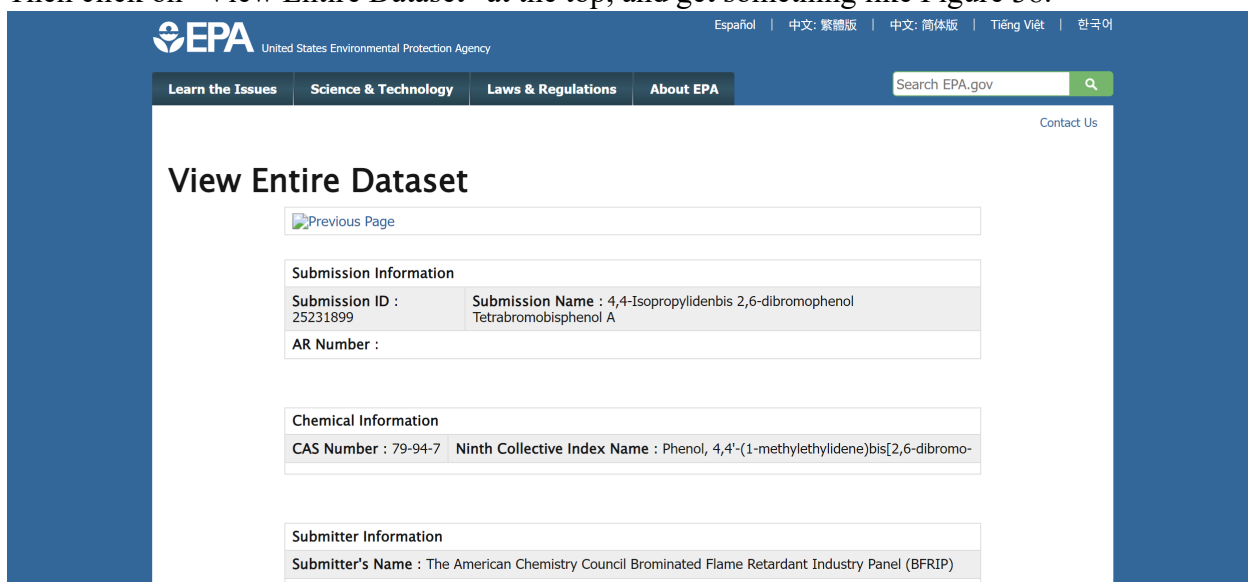


Figure 38. Example of "Entire Dataset" page in HPVIS

- Further down the page are sections that contain more detailed information, related to the simpler information obtained from ChemView. For example, a section on acute toxicity in Figure 39.

Acute Toxicity	
CAS Number :	79-94-7
Sponsored Chemical Name :	Phenol, 4,4'-(1-methylethylidene)bis[2,6-dibromo-
Test Substance – Acute Toxicity	
Test Substance:	(79-94-7) Phenol, 4,4'-(1-methylethylidene)bis[2,6-dibromo-
Test Substance Purity/Composition and Other Test Substance Comments:	TBBPAMeasured
Method – Acute Toxicity	
Route of Administration:	Inhalation
Type of Exposure:	
Species:	Guinea pig
Mammalian Strain:	
Gender:	Both M/F
Number of Animals per Dose:	10

Figure 39. Example of acute toxicity section of "Entire Dataset" page in HPVIS

- Download this entire dataset by saving the “View Entire Dataset” webpage as a PDF
- In Chrome, click the “three-dot” menu, then select Print.
- In the Print window that opens, for Destination select “Save as PDF” then click the blue “Save” button at the bottom of the page.
- Name the PDF HPVIS_full_CASRN. Note that the file name should indicate that this is a more complete HPVIS download by including the word “full” (rather than the higher-level summary that was previously downloaded using section 5.3.2 of the guide).
- File name e.g., HPIVS_full_79-94-7.pdf

5.4.2.1 How to get HPVIS data counts from ChemView

This is a record of a previous approach used to retrieve HPVIS data. **Do not use this method. Use the method outlined above (Section 5.4.2 How to get more detailed HPVIS information) to retrieve detailed HPVIS information.**

- After generating the exported Excel sheet containing links from the batch search, open the Excel sheet.
- In the column titled HPVIS, open any links present.

3. In the window that opens, under the header “Information Submitted to EPA,” click HPVIS (Figure 40.).

Print | E-mail Url Home

ClCC(Cl)OC(=O)OP(=O)(OC(Cl)CC(Cl)C)OC(Cl)CC(Cl)C

Chemical Name: Tris(1,3-Dichloroisopropyl) phosphate CAS #: 13674-87-8

Information Submitted to EPA

High Production Volume Information System
HPVIS

Manufacturing, Processing, Use, and Release Data Maintained by EPA

Chemical Data Reporting
[CDR Summary](#)

Figure 40. Navigating to HPVIS section on Chem/View

4. In the window that opens, click View High Production Volume Information System (Figure 41.)


<h3>High Production Volume Information System</h3> <p>Chemical View High Production Volume Information System</p> <p>Name: 2-propanol, 1,3-dichloro, phosphate</p> <p>Chemical Identifier: 13674-87-8</p>	
<p>Submitter name: Akzo Nobel Functional Chemicals LLC</p>	
<p>PDF file: Tris [1,3-dichloro-2-propyl] Phosphate (FYROL FR-2)_groupfile.pdf</p>	

Figure 41. Navigating to HPVIS website from ChemView

5. Print the resulting page (Figure 42.) as a PDF. Title this PDF “EPA HPVIS [CASRN] detailed chemical results” – replace [CASRN] with the appropriate CAS number.

Detailed Chemical Results

Chemical Name: 2-Propanol, 1,3-dichloro-, phosphate (3:1)
CAS Number: 13674-87-8

Click on the endpoint link to see the data on a tab page.

Physical-Chemical SIDS

- [Melting Point\(1\)](#)
- [Boiling Point\(1\)](#)
- [Vapor Pressure\(1\)](#)
- [Partition Coefficient\(1\)](#)
- [Water Solubility\(1\)](#)

Fate SIDS

- [Photodegradation\(1\)](#)
- [Stability in Water\(2\)](#)
- [Biodegradation\(2\)](#)

EcoToxicity SIDS

- [Acute Toxicity to Aquatic Vertebrates\(1\)](#)
- [Acute Toxicity to Aquatic Invertebrates\(1\)](#)
- [Acute Toxicity to Aquatic Plants\(1\)](#)

Mammalian Health Effects SIDS

- [Acute Toxicity\(3\)](#)
- [Repeated-Dose Toxicity\(1\)](#)
- [Genetic Toxicity in vitro\(6\)](#)
- [Reproductive Toxicity\(1\)](#)
- [Developmental Toxicity/Teratogenicity\(1\)](#)

Mammalian Health Effects Other

- [Skin Irritation\(1\)](#)
- [Eye Irritation\(1\)](#)
- [Skin Sensitization\(1\)](#)
- [Neurotoxicity\(2\)](#)

Figure 42. HPVIS data type counts

5.4.3 How to search for EPA documents via the National Service Center for Environmental Publications (NSCEP)

1. Go to <https://www.epa.gov/nscep>
2. Scroll down and select Advanced Search. Search by CASRN for each chemical individually.
3. Document the date on which the search was performed, and the total number of hits returned.
4. Screen through the results for useful documents. “Useful” includes any mention of the chemical which 1) is relevant to chemical risk assessment, 2) provides more detailed information than simply the presence of a chemical on a list, and 3) is not simply an addition to the federal register.
5. In the search results page, clicking on the title of a document will open a page for that document, showing one page at a time. Near the top of the page, below the title, the number of hits in the present document for the searched CASRN is shown. At the right

and left side of the document page are red arrows for “Next Search Term” and “Prior Search Term.” These buttons will navigate between occurrences of the searched CASRN in the document. Using these buttons can help to quickly evaluate the quality and relevance of the document for a literature survey. **Note:** it can sometimes be faster to download a PDF (as outlined in step 6) and then search within it rather than navigate search terms on the website, but not all PDFs provided via NSCEP are searchable.

6. Relevant documents can be saved using the “Adobe PDF” button to the right of the displayed document page. Files should be saved in the format: “US EPA_DATE_SHORT TITLE,” where SHORT TITLE indicates a shortened version of the document title.
7. For an individual chemical, screening can stop when: (1) all known EPA documents are identified (e.g., PPRTV and IRIS documents otherwise identified), AND (2) an entire 15-reference page is screened with no new useful assessments identified.
8. Track relevant results in the tracking document. Include number of hits. Document titles and links should be tracked in a separate document, and the name of this tracking document and the presence of EPA documents of interest should be noted in the tracking spreadsheet.

5.4.4 How to search for documents from Health Canada

Note: Search only for chemicals identified by the CompTox Dashboard as being on the Canadian Domestic Substances List (DSL).

1. In the Google search bar, enter the CAS number for the chemical being searched followed by “site:canada.ca”, as shown in **Figure 43**. Note that there is no specific domain for Health Canada. Note also that only chemicals for which the EPA CompTox Dashboard download (Section 2.2.1) indicates a substance is on the Domestic Substances List (DSL) need to be searched.

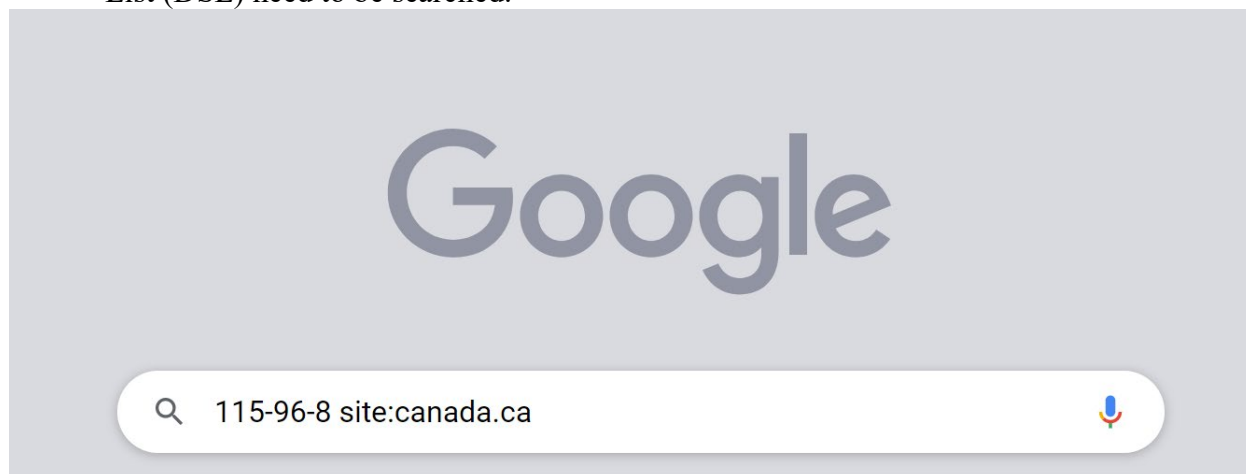


Figure 43 Example of searching the Canadian government websites for a certain chemical

2. Document the number of hits resulting from each search.
3. Look for documents that are completed assessments. A non-exhaustive list of phrases to look for include “screening assessment” (as used in “Draft Screening Assessment”, “Screening Assessment Report”, and “Screening Assessment for the Challenge”), and

“Identification of Risk Assessment Priorities”.

4. Download and save a PDF when available. If a PDF is not available, and the report is presented as a web page instead, print the entire web page as a PDF.

5.4.5 How to search for ECHA assessments and links to dossiers

1. In the Google search bar, enter the CAS number for the chemical being searched followed by “site:echa.europa.eu”, as shown in Figure 44..

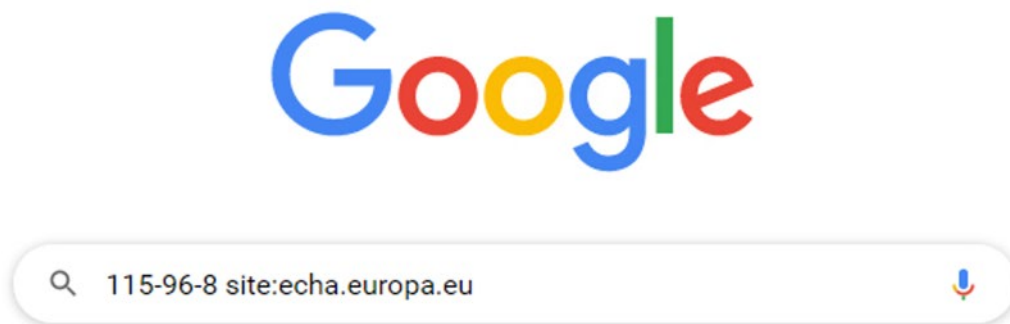


Figure 44. Example of searching the ECHA website for a certain chemical

2. Click on the link from the Google search results that says “...Registration Dossier” to access the website with ECHA dossier information – copy and save the link to the Dossier, to accompany the PDF information downloaded from IUCLID (see section 5.4.6). Click on the PDF document from the search results to view and download the ECHA assessment, as shown in Figure 45..

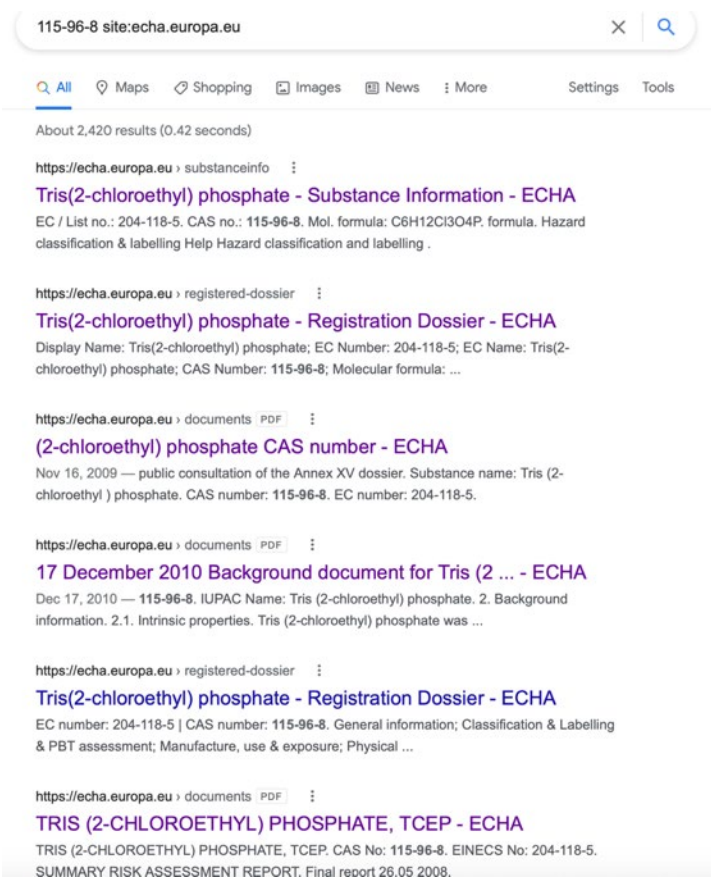


Figure 45. Example of search results when finding ECHA documents

5.4.6 How to download ECHA REACH study results (dossier info) via IUCLID

REACH study results can be obtained via IUCLID. Doing so requires creating an account, downloading some software and files (Section 5.4.6.1), searching the files for the desired dossier (if it exists) and importing the dossier file into IUCLID software (Section 5.4.6.2), and then downloading the desired dossier info from IUCLID (Section 5.4.6.3).

5.4.6.1 Installing IUCLID and downloading files

1. Go to <https://iuclid6.echa.europa.eu/reach-study-results>
2. First, download “REACH Study Results as IUCLID 6.5 dossiers” (9 GB), as displayed in Figure 46.. This will take some time (about 6 hours to download, and about 2 hours to unzip), but requires minimal interaction. This downloading can be done in the background while working on other tasks, or overnight.
 - a. Note: it will be necessary to create an account and sign in to download the 9GB file.

The screenshot shows the IUCLID 6 website interface. At the top left is the IUCLID 6 logo. To the right is a search bar and a 'Logga in' link. Below the logo is a navigation menu with 'Home', 'IUCLID Product', 'Download Software', 'Support', and 'News'. A breadcrumb trail reads 'IUCLID > Support > Get IUCLID data > REACH Study Results'. On the left, under 'Get IUCLID data', there are three links: 'IUCLID inventories', 'Reference substances', and 'REACH Study Results'. The main content area is titled 'REACH Study Results' and contains a paragraph explaining the data source. Below this is a 'Downloads' section with a blue box containing text about data refreshes. Underneath, there are two download links: 'REACH Study Results as IUCLID 6.5 dossiers (9 GB)' and 'List of substances (5.6 MB)'. A red arrow points to the first link. Below the download links is a 'Content of the data' section with a paragraph and a screenshot of a 'REACH Complete' dossier structure showing sections like '1 General information', '2 Classification & Labelling and PBT assessment', and '3 Manufacture, use and exposure'.

Figure 46. Example of downloading IUCLID REACH Study Results

3. Download the “List of substances”
4. Go to <https://iuclid6.echa.europa.eu/download>
5. Download the “IUCLID 6 Desktop installer (refer to Figure 47.). Once it has downloaded, open it and follow instructions to install IUCLID.

Sign Out | My Account

Search the IUCLID 6 website

Home | IUCLID Product | Download Software | Support | News

IUCLID > Download Software

Download Software

- > Cloud, Desktop or Server - which version of IUCLID is best for me?
- > IUCLID 6 installer and updater - which one should I download?

The latest version of IUCLID 6 published for all users on this website is **5.15.0**. The ECHA Cloud services users have access to version **5.15.0**.

Release notes

The main improvements and bug fixes included in the IUCLID versions available for download on this website, or accessible via the ECHA Cloud services, are listed in the [release notes](#).

Available downloads

In connection with your IUCLID 6 download, you might want to consult the following links: [IUCLID 6 manuals](#) | [FAQ](#) | [IUCLID 6 webinars](#) | [checksums for downloads](#)

- **IUCLID 6 Desktop installer**: this download provides all you need to run the IUCLID 6 as a desktop application. If you need a 32-bit compatible installer, please contact the [ECHA Helpdesk](#).
 - ➔ [Windows 64-bit \(629 MB\)](#) | [macOS \(654 MB\)](#)
- **IUCLID 6 Server**: this download contains all necessary software to run the IUCLID 6 Server version that can be installed in a multi-user environment using client-server architecture, including its application server.
 - [Windows and Linux \(652 MB\)](#)
- **IUCLID 6 updater**: this download provides all you need to update your existing IUCLID 6 application (*the same updaters can be used for IUCLID 6 Desktop or Server*). If you need a 32-bit compatible updater, please contact the [ECHA Helpdesk](#).
 - [Windows 64-bit and Linux \(905 MB\)](#) | [macOS \(905 MB\)](#)

Figure 47. Download page for IUCLID 6 Desktop Installer

5.4.6.2 Finding if dossier info exists in ICULID, and importing file to IUCLID program

Note: IUCLID program can be slow to open – open first, then do other prep while it loads

1. Search downloaded Excel sheet for CAS (Figure 48.). **Note:** some CAS appear more than once, associated with different UUIDs (dossier IDs). Be sure to check all UUIDs that appear for a given CAS number. A clear warning sign is if the file loading into IUCLID (below) is very fast for a chemical that is expected to have a lot of data.

1	DOSSIER UUID (Dossier)	NAME (Substance)	CAS_NUMBER	NUMBER_IN_IU	IUPAC_NAME (Ref. sub.)	M
4590	c24756ca-8af7-4235-949e-1f58a1c3f91e	2,2,4-trimethylpentane	540-84-1	208-759-1	2,2,4-trimethylpentane	
4591	c64c3139-c870-4fc4-97c2-dea2ea9044b6	2,2,4-trimethylpentane-1,3-diol	144-19-0	205-619-1	2,2,4-trimethylpentane-1,3-diol	
4592	27de786c-43ad-4453-932f-c53b80ad540f	2,2,5,5-tetra- <i>tert</i> -butyl-1,4,2,5-dioxadisilane	5895-82-9	700-444-4	2,2,5,5-tetramethyl-1,4,2,5-dioxad	
4593	074c0576-105c-427b-8ab9-023e61f5db0f	2,2,5-trimethyl-5-pentylcyclopentan-1-one	65443-14-3	665-779-3	2,2,5-trimethyl-5-pentylcyclopent	
4594	fde49a5d-e285-43e2-8cd0-1e12-4e9999	2,2,6,6,7,8,8-HEPTAMETHYL-4,5,6,7,8,8B-HEXAHYDRO-3AH-INDENO[4,5-D][1,3]DIOXOLE		457-070-4	2,2,6,6,7,8,8-heptamethyl-2H,3aH,	
4595	c1201818-427f-427e-480-f480e81f544	2,2,6,6,7,8,8-HEPTAMETHYL-4,5,6,7,8,8B-HEXAHYDRO-3AH-INDENO[4,5-D][1,3]DIOXOLE		457-070-4	2,2,6,6,7,8,8-heptamethyl-2H,3aH,	
4598	eb215fe1-5258-4824-8dd0-669ee6e8e3d4	2,2',6,6'-tetrabromo-4,4'-isopropylidenediphenol	79-94-7	91-236-9	2,6-dibromo-4-[2-(3,5-dibromo-4-yl	
4599	6990750-070c-4b12-8022-90821d0c222	2,2',6,6'-Tetrabromo-4,4'-isopropylidenediphenol, oligomeric reaction products with 1-chloro-2,3-		500-399-6	Reaction mass of 2,2'-(propane-2,	
4598	cc38fc5c-6c82-4022-b381-e9128cb3218e	2,2',6,6'-Tetrabromo-4,4'-isopropylidenediphenol, oligomeric reaction products with 1-chloro-2,3-		500-399-6	NA	
4599	7ac3e354-2f29-4f1c-93db-b966f87b1d46	2,2',6,6'-Tetrabromo-4,4'-isopropylidenediphenol, oligomeric reaction products with Propylene ox		926-564-6	2,2',6,6'-Tetrabromo-4,4'-isoprop	
4600	8378e07d-9e85-43f0-889e-5e9dbc06a2c1	2,2,6,6-tetrakis(bromomethyl)-4-oxaheptane-1,7-diol	109678-33-3		3-[2,2-bis(bromomethyl)-3-hydrox	
4601	00eb37d1-b79e-4f38-bb10-ec1b4364c804	2,2,6,6-tetramethyl-1-[2-[(3,5,5-trimethylhexanoyl)oxy]ethyl]piperidin-4-yl 3,5,5-trimethylhexano	1445870-18-7	941-267-1	2,2,6,6-tetramethyl-1-[2-[(3,5,5-tri	
4602	70a4e8fc-c95d-4efd-9c89-fe865e537aff	2,2,6,6-tetramethyl-4-oxopiperidinoxy	2896-70-0	220-778-7	(2,2,6,6-tetramethyl-4-oxopiperidi	
4603	04bcc573-ffd6-4cc0-90bb-f79efc8d7e95	2,2,6,6-tetramethyl-4-piperidone	826-36-8	212-554-2	2,2,6,6-tetramethylpiperidin-4-one	
4604	5ecf0906-b6a2-4cf8-b367-ac1edc170e48	2,2,6,6-tetramethyl-4-piperidyl acrylate			2,2,6,6-tetramethylpiperidin-4-yl p	
4605	b418a961-2821-4022-b8d5-c61455514041	2,2,6,6-tetramethyl-4-piperidylamine	36768-62-4	253-197-2	2,2,6,6-tetramethylpiperidin-4-ami	
4606	96312b84-2da1-4dc5-9cae-152715030d16	2,2,6,6-tetramethyl-4-substituted-piperidinyl-1-ox			4-butoxy-2,2,6,6-tetramethyl-1-pi	

Figure 48. Example of searching for a CAS number in the downloaded Excel spreadsheet

- Copy each UUID that exists for a given CAS number, search for that UUID in the folder of IUCLID dossiers (downloaded and unzipped) as shown in Figure 49..

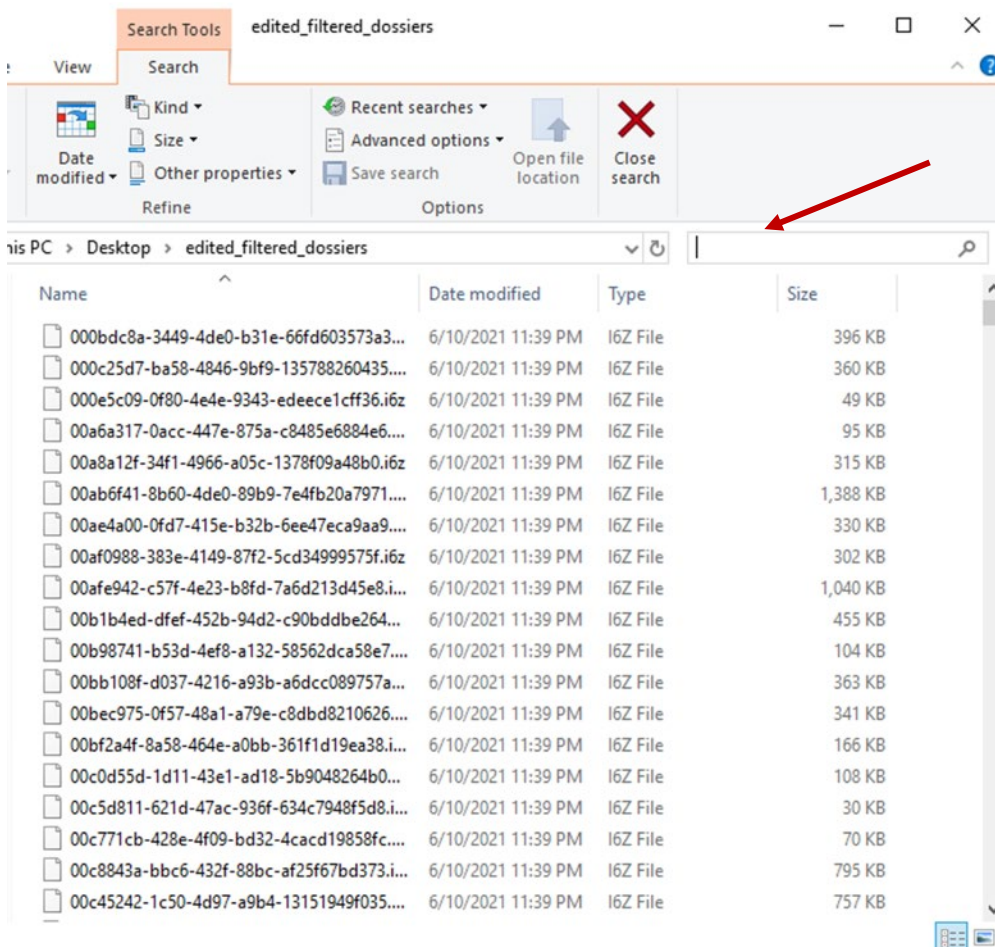


Figure 49. Searching for a UUID in the folder of IUCLID dossiers

3. To load dossier into IUCLID program, click appropriate dossier file and drag into IUCLID (Figure 50.). If loading several dossiers at once, repeat steps 1-3 until all desired dossiers are loaded.

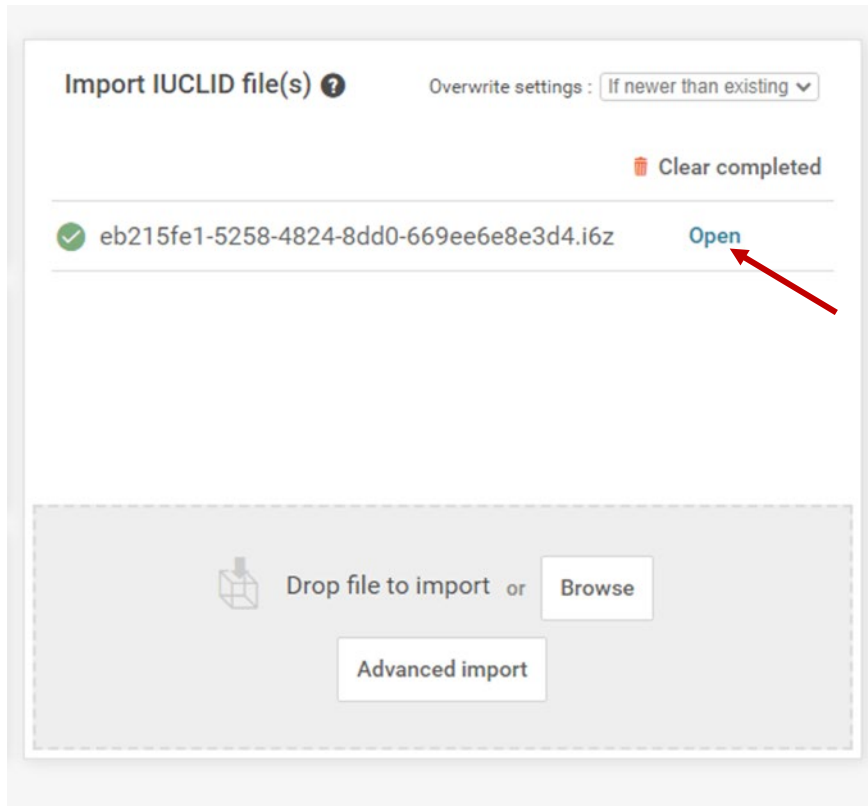


Figure 50. Example of dossier file loaded into IUCLID program

5.4.6.3 Downloading dossier info from IUCLID

1. Once the file loads, click open.
2. Click “REACH Complete” at top left (should already be selected, unless user was navigating through the dossier first)

3. Click the “3-dots in a box” at the top right, as shown in Figure 51..

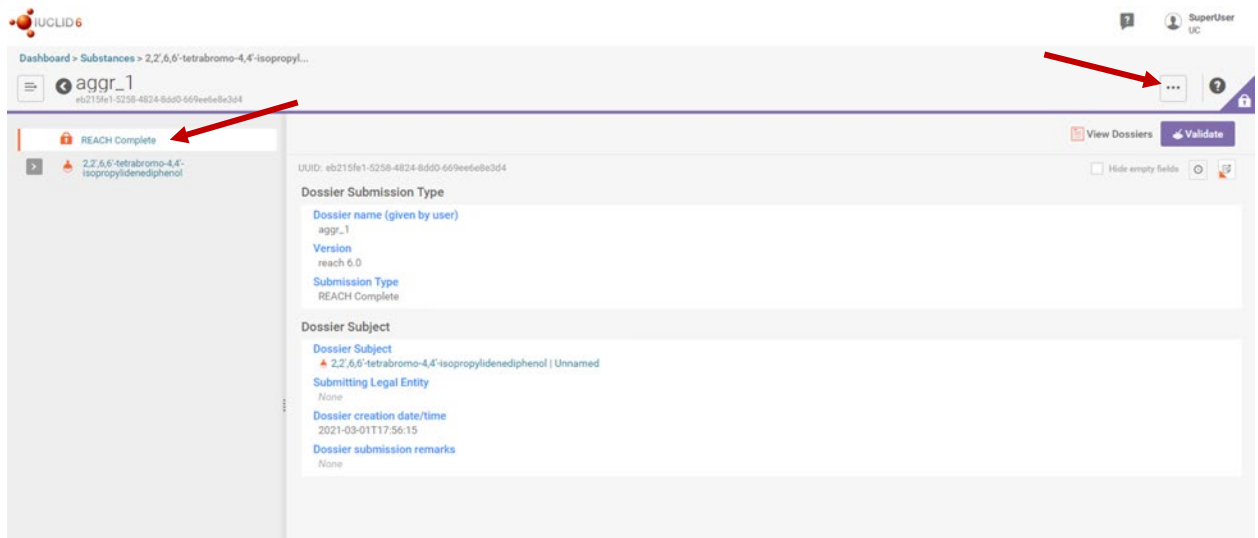


Figure 51. Example of main screen when dossier is opened

4. Select “Create Component PDF” (Figure 52.).

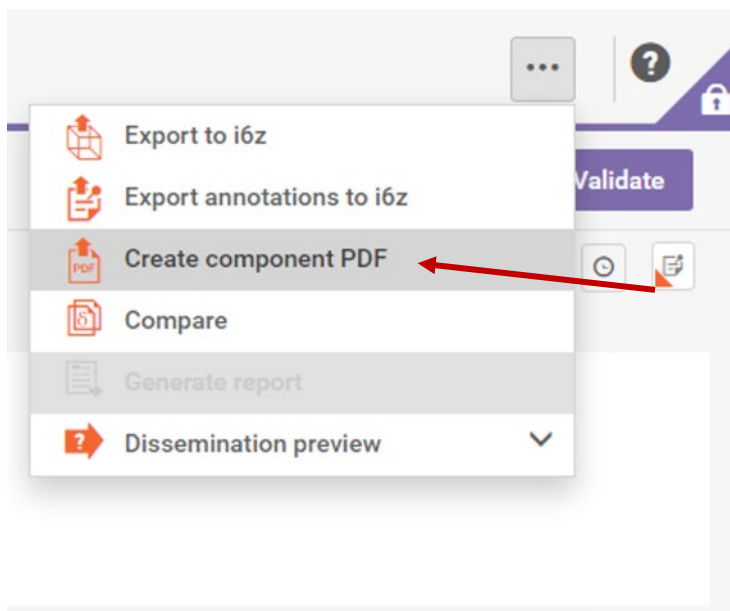


Figure 52. Creating a component PDF of the dossier information

5. Leave all defaults selected.

6. Scroll to the bottom and check “Select documents to be included”, as displayed in Figure 53..

The screenshot shows a 'PDF Settings' dialog box with a close button (X) in the top right corner. The 'Working Context' is 'REACH Complete'. The settings are as follows:

- Include cover page
- Included Annotations**
 - Include annotations
- Include legal entity**
 - Include legal entity
- Detail level of document fields**
 - Detailed fields (e.g. needed for robust endpoint summaries)
 - Fields marked "confidential"
 - Include fields with no value
- Flags for confidentiality**
 - Select information to be included
 - Data for which a confidentiality flag may be set, but it is not.
 - CBI
 - IP
 - no PA
- Flags for regulatory programme**
 - Select information to be included
 - No regulatory purposes
 - AU: AICIS
 - CA: CEPA
 - CA: PCPA
 - EU: BPR
 - EU: CLP
 - EU: PPP
 - EU: REACH
 - JP: CSCL
 - NZ: HSNO
 - OECD: CoCAP
 - US: EPA HDVC

At the bottom, there is a checkbox Select documents to be included and a purple 'Next' button. Two red arrows point to these elements from the bottom right of the dialog.

Figure 53. PDF settings window

7. Click Next

8. Deselect unwanted information, as shown for Physical and Chemical Properties in
 - a. Deselect everything under Physical and Chemical Properties, Environmental Fate and Pathways, and Ecotoxicology
 - b. As a result, only Identification, Composition, and all available Toxicology and Toxicokinetic information will be selected.
 - c. Note that some of the “general information” is not removable, but this includes information that is necessary to identify the substance.

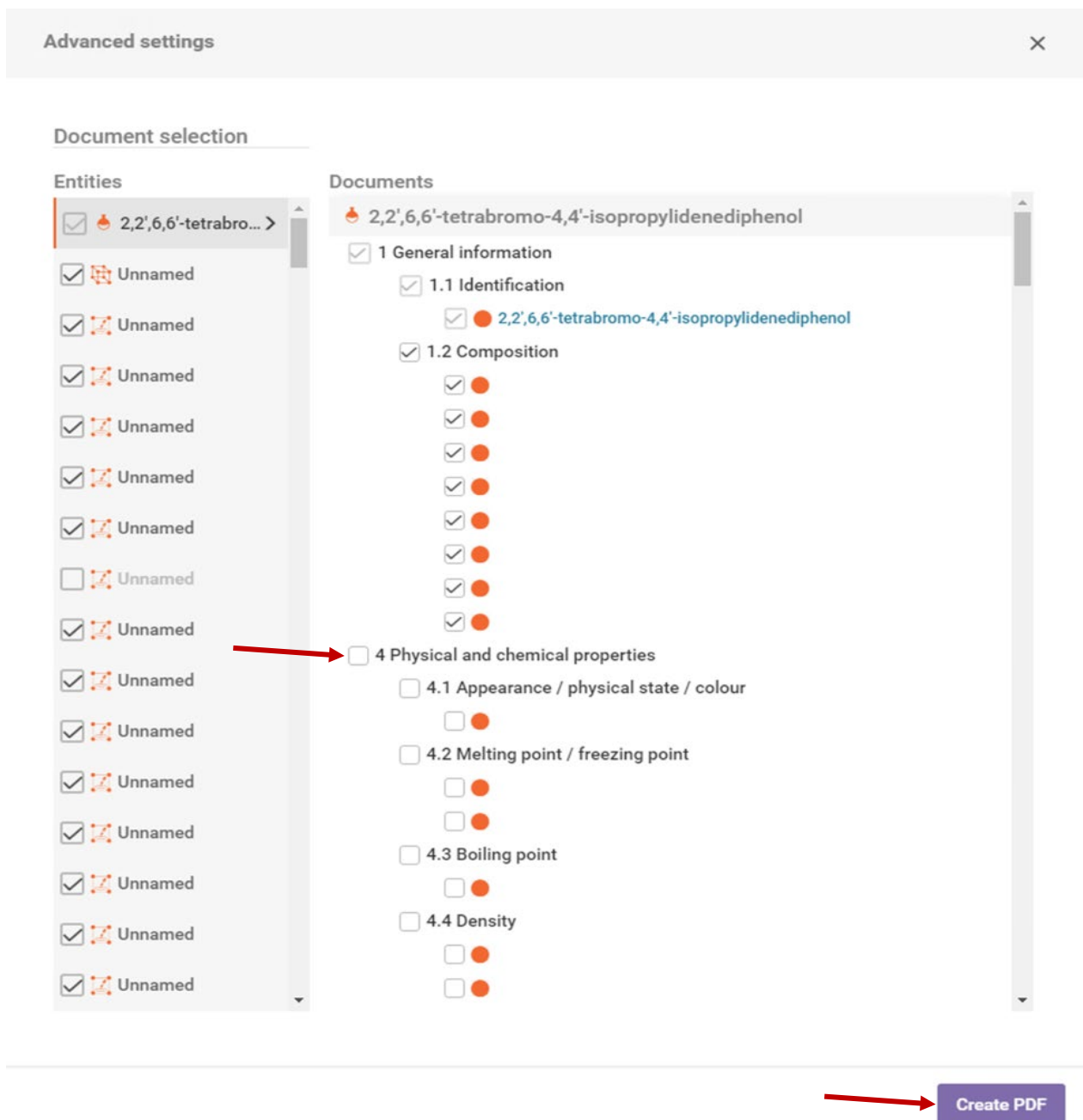


Figure 54. Deselecting irrelevant information and creating the PDF

9. Click Create PDF at bottom right. Generating PDF may be slow, particularly for chemicals with a large quantity of data. Once PDF is generated, download and save with a descriptive file name: ECHA Dossier info via IUCLD [CASRN]

Note: table of contents of PDF has links to click through to specific sections

5.4.7 How to search for NTP publications

1. Go the following website: <https://ntp.niehs.nih.gov/publications/index.html>
2. Scroll to the bottom of the page to find the search bar indicated in Figure 55., and search by CASRN to view the full list of available reports.

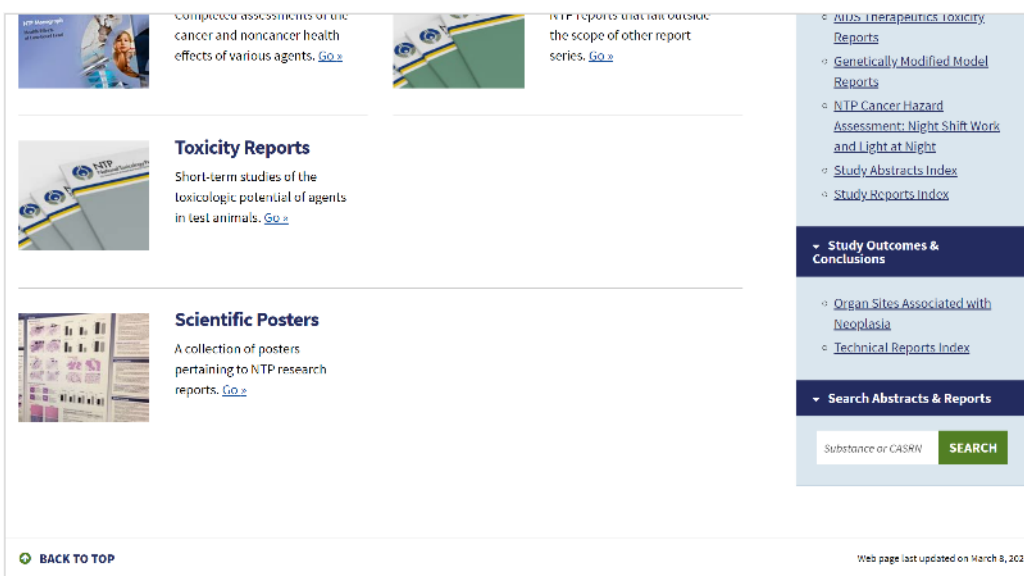


Figure 55. NTP Publications home screen search bar

3. From the list of technical reports, the reports being displayed can be filtered by searching for the chemical of interest. This can be accomplished by searching for the chemical name or CAS number in the search box labeled as “Study Reports”, as illustrated in **Figure 55**. After searching for a chemical, the respective technical report can be opened and downloaded by clicking on the “PDF” link in the “Full Text” column.

5.4.8 How to search for documents on INCHEM

1. Go to the following website: <https://inchem.org/#/>
2. On the left side of the webpage, chemicals can be searched by CAS numbers or by chemical names (Figure 56.).

INCHEM Internationally Peer Reviewed Chemical Safety Information

World Health Organization

BROWSE Collections SEARCH INCHEM for...

Rapid access to internationally peer reviewed information on chemicals published through the International Programme on Chemical Safety (IPCS). All types of chemicals from the full range of exposure situations (environment, food, occupational) are included. Searching across all collections or within individual collections is available.

About INCHEM >

Filter your search by: Showing 1 - 10 of 8,426 Per page 10

Chemical Identity Search

CAS Number

Search...

- 67-66-3 10
- 56-23-5 9
- 7439-92-1 9
- 7439-97-6 9
- 29973-13-5 8
- 7440-38-2 8
- 7440-43-9 8
- 75-15-0 8
- 75-21-8 8
- 100-42-5 7
- 107-13-1 7
- 122-39-4 7

SHOW MORE

Chemical Name or Synonym

Search...

Document Title	Summary
Disulfiram (ICSC)	DISULFIRAM ICSC: 1438 Tetraethylthiuramdisulfide 1,1'-Dithiobis(N,N-diethylthioformamide) bis-(N,N-Diethylthiocarbamoyl)disulfide TETD September 2002 CAS #
GLUCOSE ISOMERASE FROM ACTINOPLANES MISSOURIENSIS (IMMOBILIZED) (JECFA Evaluation)	Summary of Evaluations Performed by the Joint FAO/WHO Expert Committee on Food Additives GLUCOSE ISOMERASE FROM ACTINOPLANES MISSOURIENSIS (IMMOBILIZED) Ch
Trichloroethane, 1,1,1- (WHO Food Additives Series 16)	1,1,1-TRICHLOROETHANE Explanation This substance was evaluated for an acceptable daily intake for man by the Joint FAO/WHO Expert Commit
Tin (ICSC)	TIN ICSC: 1535 October 2004 CAS #: 7440-31-5 EC Number: 231-141-8 ACUTE HAZARDS PREVENTION FIRE FIGHTING FIRE & EXPLOSION Combustible. NO open flames. Use
Acephate (Pesticide residues in food: 1990 evaluations Toxicology)	ACEPHATE EXPLANATION First draft prepared by Dr. E.M. den Tonkelaar, National Institute of Public Health and Environmental Protection,
Cyfluthrin (UKPID)	UKPID MONOGRAPH CYFLUTHRIN SA Cage MSc M Inst Inf Sci SM Bradberry BSc MB MRCP S Meacham BSc JA Vale MD FRCP FRCP E
Aramite (IARC Summary & Evaluation, Volume 5, 1974)	International Agency for Research on Cancer (IARC) - Summaries & Evaluations ARAMITER VOL.: 5 (1974) (p. 39) CAS No.: 140-57-8 5.

Figure 56. INCHEM home screen

- In order to search by CAS number, enter the number in the box under the heading “CAS Number”. Click on the resulting CAS number displayed under the search bar, as displayed in Figure 57.

INCHEM Internationally Peer Reviewed Chemical Safety Information

World Health Organization

BROWSE Collections SEARCH INCHEM for...

Rapid access to internationally peer reviewed information on chemicals published through the International Programme on Chemical Safety (IPCS). All types of chemicals from the full range of exposure situations (environment, food, occupational) are included. Searching across all collections or within individual collections is available.

About INCHEM >

Filter your search by: Showing 1 - 10 of 8,426 Per page 10

Chemical Identity Search

CAS Number

115-96-8

115-96-8 5

Chemical Name or Synonym

Search...

- carbaryl 19
- lindane 17
- captan 15
- fenitrothion 14
- heptachlor 14
- malathion 14
- permethrin 14
- chlorodane 13
- chlorothalonil 13
- dimethoate 13
- acephate 12
- carbendazim 12

SHOW MORE

Document Title	Summary
Disulfiram (ICSC)	DISULFIRAM ICSC: 1438 Tetraethylthiuramdisulfide 1,1'-Dithiobis(N,N-diethylthioformamide) bis-(N,N-Diethylthiocarbamoyl)disulfide TETD September 2002 CAS #
GLUCOSE ISOMERASE FROM ACTINOPLANES MISSOURIENSIS (IMMOBILIZED) (JECFA Evaluation)	Summary of Evaluations Performed by the Joint FAO/WHO Expert Committee on Food Additives GLUCOSE ISOMERASE FROM ACTINOPLANES MISSOURIENSIS (IMMOBILIZED) Ch
Trichloroethane, 1,1,1- (WHO Food Additives Series 16)	1,1,1-TRICHLOROETHANE Explanation This substance was evaluated for an acceptable daily intake for man by the Joint FAO/WHO Expert Commit
Tin (ICSC)	TIN ICSC: 1535 October 2004 CAS #: 7440-31-5 EC Number: 231-141-8 ACUTE HAZARDS PREVENTION FIRE FIGHTING FIRE & EXPLOSION Combustible. NO open flames. Use
Acephate (Pesticide residues in food: 1990 evaluations Toxicology)	ACEPHATE EXPLANATION First draft prepared by Dr. E.M. den Tonkelaar, National Institute of Public Health and Environmental Protection,
Cyfluthrin (UKPID)	UKPID MONOGRAPH CYFLUTHRIN SA Cage MSc M Inst Inf Sci SM Bradberry BSc MB MRCP S Meacham BSc JA Vale MD FRCP FRCP E
Aramite (IARC Summary & Evaluation, Volume 5, 1974)	International Agency for Research on Cancer (IARC) - Summaries & Evaluations ARAMITER VOL.: 5 (1974) (p. 39) CAS No.: 140-57-8 5.

Figure 57. Example of searching by CAS number on INCHEM

4. The relevant documents for the CAS number that was searched are displayed as shown in Figure 57. By clicking on the links in the document titles, the various documents can be viewed. The documents can be exported as PDFs and saved.
 - a. For IARC documents INCHEM only provides the summary of the assessments. This is sufficient to download.

5.4.9 How to search for assessments on OECD Existing Chemicals Database

1. Go to the following website: <https://hpvchemicals.oecd.org/UI/Default.aspx>
2. From the menu on left side of the screen, click on “Search” (as illustrated in Figure 58.).

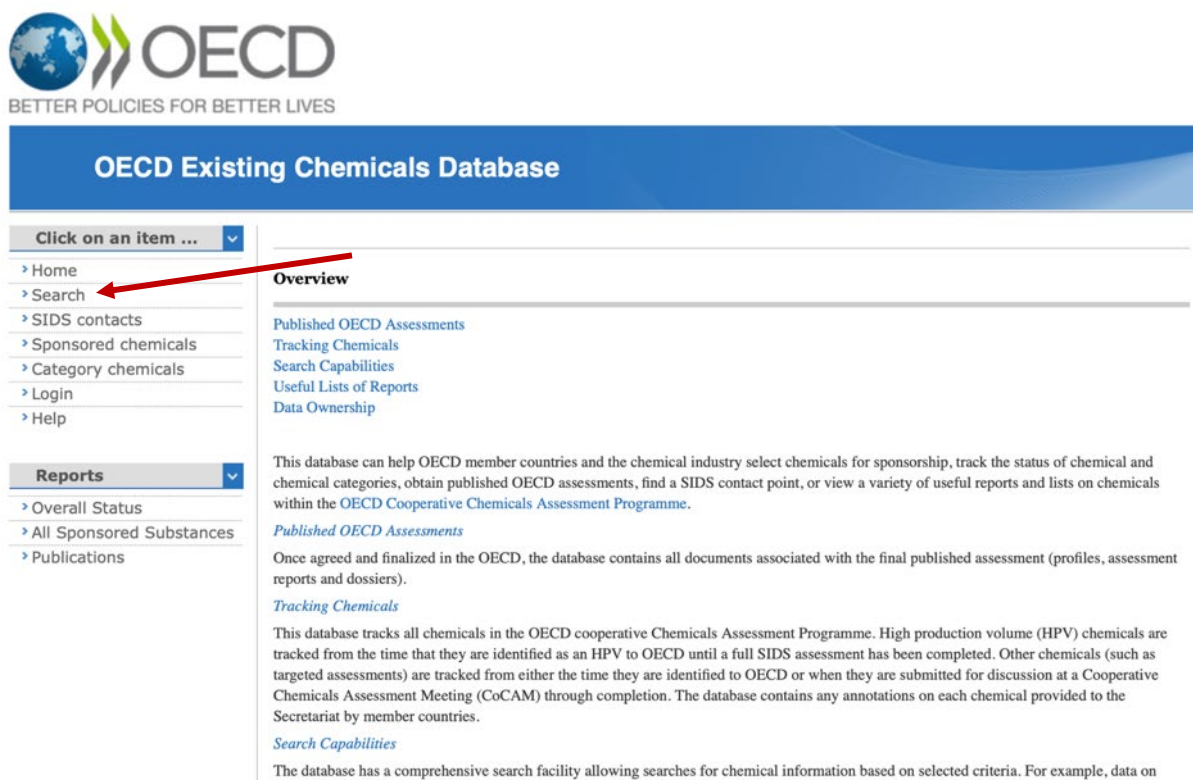


Figure 58. OECD Existing Chemicals Database home screen

3. To search for a chemical based on a CAS number, enter the number in the box labeled “by CAS Number” (refer to Figure 59.).

4. Next, click on the “Search” icon at the bottom of the screen to perform the search, as displayed in Figure 59.

OECD Existing Chemicals Database

Click on an item ...

- Home
- Search
- SIDS contacts
- Sponsored chemicals
- Category chemicals
- Login
- Help

Reports

- Overall Status
- All Sponsored Substances
- Publications

> HPV Database search

by Chemical Name: by CAS Number: 115-96-8

by Sponsor: by SIDS process Status: Search in:

<No Criteria> <No Criteria> <No Criteria>

Australia Chemical not yet sponsored

Austria Information Gathering & Data Review

Belgium Preliminary Documents for Written Comments

BIAC Preliminary Review Completed

Canada Chemical assessment in discussion

Czech Republic Chemical assessment discussed

Denmark Conclusions Agreed

EEB Publication available

Estonia Provisionally Set Aside

European Chemicals Agency Further Work Underway

European Commission

Finland

France

Germany

Greece

All Chemicals

Sponsored chemicals

Non-sponsored chemicals

<No Criteria>

By HPV chemical only

By Non-HPV chemical only

<No Criteria>

By target assessments

By non-targeted assessments

Return 25 Rows per page Order by: Name

SEARCH

Click here for hints about the search

Information has only been included in the database where it has been notified to OECD by Member Countries

© OECD. All rights reserved. Terms & Conditions | Privacy Policy Contact Us | Site Map | Help/FAQ | MyOECD

Figure 59. Example of search input on OECD Existing Chemicals Database

5. If the chemical that was searched has relevant results available in the OECD Existing Chemicals Database, they will now be displayed. In order to download the SIDS Initial

Assessment Profile for the chemical, click on the link to the PDF aligned with the row labeled “OECD Agreed Conclusions”, as illustrated in Figure 60.

OECD Existing Chemicals Database

Click on an item ...

- Home
- Search
- SIDS contacts
- Sponsored chemicals
- Category chemicals
- Login
- Help

Reports

- Overall Status
- All Sponsored Substances
- Publications

> HPV Database search > Chemical Detailed Results

Identity

Cas Number(s)	115-96-8
Chemical Name	Ethanol, 2-chloro-, phosphate (3:1)
Synonyms (Others Chemical names)	Tris(2-chloroethyl)phosphate

Available information on the chemical

Information has only been included in the database where it has been notified to OECD by Member Countries.

HPV	Yes
Recognized Low hazard	No
Existing Reports	
On the ICCA List	No
See also	
Additional Information	EU 2nd priority substance

SIDS Relevant Information

Sponsors	Germany
Sponsorship Date	08/1995
Current Status	Publication available
Investigated in a Category	
Assessment Meeting (SIAM or CoCAM)	SIAM 23, 17/10/2006, assessed
OECD Agreed Conclusions	115968.pdf
Final Assessment Report	SIAR published by EU
Date Published	10/2009
Targeted Assessment	No
OECD-IUCLID Export File	

© OECD. All rights reserved. Terms & Conditions | Privacy Policy

Contact Us | Site Map | Help/FAQ | MyOECD

Figure 60. Example of chemical detailed results on OECD Existing Chemicals Database

5.4.10 How to search for reviews on Cochrane

The terms used for searching will vary by class. For PHOPs, the broad search term “phos” was initially tried, and returned many results (829), mostly irrelevant. Next, organophosph* was tried. This was much narrower, returning 20 results. However, only 2 were deemed relevant, with the rest focusing on pesticides or pharmaceuticals. Finally, searching “flame retard*” AND organophosph* returned only 2 results, which were the same 2 relevant results that were initially identified by searching only for organophosph*. A similar approach of progressively narrower search terms is used for other classes: first a very general term broadly related to the chemistry of the class; second, a more specific term related to the chemistry of the class; and third, adding “flame retard*” to the more specific chemistry term.

1. Go to <https://www.cochranelibrary.com/cdsr/reviews>

2. In the search bar at the top right, replace “*” with the appropriate search terms. Search terms will necessarily vary by class, and some trial and error may be necessary.
3. Clicking on “Advanced search” allows creating a record of searches attempted, enabling easier comparison between search strategies attempted. **Note** that the number shown for each search in the record includes all publication types covered by Cochrane (reviews, protocols, trials, editorials, special collections, clinical answers), not just reviews. Within the search results page for each individual search, it is possible to obtain numbers of results for each publication type.
4. Record the search terms attempted, the date the searches were conducted, and the number of results **for reviews only** that each set of terms produced.
5. When proper search terms are identified, document the search terms used, the date the search was conducted, the link to any relevant results (DOI or other stable link preferable), and download a PDF of the relevant references.

AFTER SEARCHING FOR AND IDENTIFYING PERTINENT GREY LITERATURE AND DEDUPLICATING AS DESCRIBED IN SECTION 5.2.3, PERFORM ADDITIONAL QA STEPS OUTLINED IN SECTION 5.3 DURING AND AFTER PDF RETRIEVAL.

6. Tagging

Detailed information on the tagging approach is provided in the separate Integrated Screeners Guide. The tagging tree and criteria for each tag were developed in consultation with CPSC, with the aim of having the same set of aggregators for PDF data and database data.

6.1 Tagging and aggregating database data

Based on the tagging tree, database logic was developed to identify the tagging strategy for each database (e.g., the relationship between various fields and the tagging tree), as described in Section 4. The database logic was then used to develop R code for automatic extraction from each database to an EXCEL® spreadsheet.

6.2 Tagging for peer-reviewed and grey literature

Details for the tagging of peer-reviewed and grey literature are provided in the separate Integrated Screener Guide. Screening and early stages of tagging (Levels 1, 2 and 3 in Distiller) are done in the Distiller software; Level 3B tagging is done using Excel, after cleanup and quality assurance of the Level 3 data from Distiller. The overall goal of tagging the PDFs is to develop semiquantitative/categorical characterizations of the number of studies in various categories. The first step was to develop a tagging tree, that is, a list of the tags and the order of

the tags. The tagging tree was developed as an iterative process with a pilot group of data sources in collaboration with CPSC staff. The tagging tree was then applied to the entire set of literature for the first OFR class in the Distiller literature management software. Lessons learned during screening and tagging were applied to update the tagging tree and to refine definitions of its categories on an ongoing basis, as needed. This iterative approach allows for the analysis of data using different approaches for organizing the tags, and aids in ensuring that the key information is obtained first, before proceeding with more time-intensive analyses.

For exposure, only information on the chemicals in the class (i.e., not on analogs) is considered. The exposure data tagging will focus on routes and pathways that are potentially relevant to consumer exposure. Relevant routes, pathways and media will be identified based on guidance from CPSC staff.

The following overall approach was used to tag pdf data from completed assessments and literature reviews. Specifics of the approach, including the tagging tree, are outlined in the “Integrated Screeners guide.” [File name: “Integrated Screener Guide_Literature Survey_2-18-229-13-21]. Detailed criteria for tagging have been developed (e.g., defining terms further for toxicokinetic and NAM studies) to ensure clarity in instructions and uniformity in how the team applies the tagging tree. This text briefly describes the approach at each level of screening to provide context for the remainder of this Guide.

The tagging tree is organized in increasing tiers of complexity, which correspond to different rounds of screening/extraction in Distiller/Excel. Each first- and second-order tag is described in more detail in the Screeners guide.

- Level 1 screening in Distiller: This level of screening/tagging addresses basic relevance of references to the PECO statement, and the possible presence of an OFR.
- Level 2 screening in Distiller: These are tags for organizing later steps. First-order tagging is conducted on completed risk assessments, toxicity assessments, exposure assessments, and literature reviews. The Level 2 tags are: toxicity, exposure, and risk. Classes addressed in the reference are also identified at Level 2.
 - Note that in the SOW, this level of tagging was referred to as zero-order tagging. During the first class (Polyhalogenated Organophosphates), tagging for toxicity, exposure, and risk occurred at both Levels 1 and 2 screening in Distiller (tagged for likelihood at Level 1, and confirmed at Level 2). In discussions with CPSC, this was determined to be duplicative, and so tags for toxicity, exposure, and risk were moved to Level 2 only. Corresponding higher levels of tagging are also now referred to as one order higher than in the SOW.
- Level 3 Extraction in Distiller: Second-order tagging will be conducted for toxicity assessments, exposure assessments, risk assessments, and literature reviews.
 - For toxicity, the second-order tags are: animal toxicity or accepted alternative, human toxicity, qualitative hazard characterization, SAR/QSAR/read across,

quantitative hazard characterization, animal or human or modeled toxicokinetics, and experimental mechanistic. For exposure, the second-order tags are: environmental monitoring, biomonitoring or personal exposure monitoring, modeled concentrations, modeled human dose, source characterization, and environmental epidemiology. For risk, the second-order tag is: quantitative human health risk assessment.

- Level 3B extraction in Excel: third-order tagging is done for selected toxicity assessments, toxicity literature reviews, risk assessments and exposure (monitoring) literature reviews.
 - Prior to conducting the second-order tagging, the UC team developed a prioritized list of tags in consultation with CPSC staff. The full tagging tree is in the Screener Guide.
- Particularly useful data sources to consider for an assessment are additionally flagged as such for later consideration in the assessment phase, using a standard format comment in the comment fields in Distiller. A standard format is used for ease of finding these key studies using Excel at a later date.

Pilot tagging stage. Prior to conducting tagging for a given order (e.g., zero-order tags), we developed specific tagging criteria (e.g., key words or other definitions) for each tag. We tested the criteria by initially reviewing and tagging a group of 100 representative toxicity/risk PDFs and 100 representative exposure PDFs. from the first class using the agreed upon tags and criteria. A team of several scientists each from the UC team and CPSC independently conducted the tagging and then compared their results. Based on the comparison, the tagging criteria were evaluated and modified as needed. Substantial modifications to the tagging criteria underwent similar pilot testing. The process followed and lessons learned have been documented, with the agreed-upon process documented in the Screeners guide. The tagging criteria and process were piloted for each order of tags. The tagging criteria will be updated additionally, as part of the how-to guide, as needed after the pilot, based on team feedback and discussions with CPSC staff.

After the pilot tagging, we proceeded with tagging the remainder of the PDFs class in Distiller. The tagging guide has been updated as needed after the pilot, based on team feedback and discussions with CPSC staff.

Quality assurance. Accurate tagging is important for the evidence maps to be informative, although it is recognized that small variations in numbers will not have significant impact on the conclusions of the literature survey. Level 1 and Level 2 screening are conducted by two independent scientists. When disagreements arise, the conflicts are resolved first by a discussion between the two relevant screeners, with a follow-up discussion including a third or more (senior) screener if the two screeners do not come to a resolution. This serves as an additional QA step.

Level 3 is conducted by a single associate or senior scientist, with initial spot checking of tagging by a senior scientist. and a percentage of the references undergoing QA review. After this round of tagging, 25% of the references that are tagged for an OFR class are checked by a second associate or senior scientist. (Note that not all references uploaded with an OFR class get tagged for that class. Some references that are not ultimately excluded are tagged only for other classes. This might occur if a chemical in a class is mentioned, but there is insufficient data, such as measurements only below the detection limit. References that are not tagged for a given class do not contribute to the total number of references for the purposes of determining the number of references for a QA check). The QA review occurs outside of Distiller. For efficiency, to the degree possible, this 25% QA will overlap with the references selected for Level 3B extraction.

6.2.1 Data cleaning and preliminary visualization using Power Query and Pivot Tables in Excel

After Level 3 tagging in DistillerSR, data extracted from PDFs needs to be exported to Excel and cleaned prior to preliminary visualization and selection of studies to be tagged for Level 3B. The primary Excel tool for this cleaning step is Power Query. Since Power Query retains a record of the steps taken, a detailed step by step will not be included in this guide. Rather, the overall approach will be outlined here.

Before any data is exported from Distiller, verify that there are no "Saved, Unsubmitted" references at Level 3 for the class with data to be exported. "Saved, Unsubmitted" references are reference for which Level 3 extraction has been started, but not completed. To check for "Saved, Unsubmitted" references in Distiller:

1. In the top menu, go to Reports > Reference Status
2. Find the cell of the table for "Saved, Unsubmitted" and "Level 3". If the number in this cell is anything other than 0, click on the number to see the references. This will go to a Datarama search.
3. To verify which class each reference belongs to, go to the Reference Display Options tab > Reference Labels box > and select "Display Labels". Then click the green Run Report button again.
4. If "Saved, Unsubmitted" references exist for the current class of interest, contact the User (indicated in the Datarama report) to finish extraction for the indicated references.

It is noted that the data cleaning process necessarily involves some QA of chemical identity information, and that the data cleaning and visualization occur simultaneously as part of an iterative process. It is further noted that some of this data cleaning of chemical identifiers can be done without the aid of the PDFs, while some will require the PDF in order to confirm identity or other issues. Data cleaning that does not require checking the PDF can be done as part of this iterative process, but data cleaning that does require the PDF should be done as part of the QA steps for Levels 3 and 3B. Note that for efficiency, QA for Levels 3 and 3B are combined to the extent possible (i.e., Level 3 QA happens in conjunction with 3B extraction for the up to 25 studies selected for 3B extraction. While QA of data extracted from PDFs could be performed in

either Distiller or Excel, Excel is used here. This is in part because of the efficiency of combining chemical identity QA with the process of generating a lookup table (described below).

The data exported from Distiller must be transformed from wide format to long format. Specifically, each combination of OFR and Level 3 tag for a given reference must be in its own row. This involves merging the columns containing Level 3 tags into a single delimited column, and then splitting that column into rows by the delimiter. For example, in data exported from Distiller, all tags for Chemical A from Reference 1 are in separate columns in a single row. To transform this into wide data, all tags are merged into a single cell, separated by a delimiter (e.g., Tag 1 | Tag 2 | Tag 3). This cell can then be split by the delimiter (here, "|") into a row for each tag, all associated with Chemical A. Note that some teams may wish to perform this same merging and splitting process with chemical identifier columns (i.e. CASRN, chemical name, and acronym) in addition to data tags; the UC team has chosen not to do this, due to challenges with chemical identification (e.g., non-unique acronyms that could lead to errors in chemical identification for a given PDF).

An additional key task is generating a lookup table for chemical identity. Generally, columns containing chemical identifiers are copied from the original data downloaded from Distiller into a new Excel sheet in the same file. In this new sheet, issues of chemical identity can be addressed (combining tasks for QA and for visualization). These can be addressed in one of a few ways. Here, the UC team used a field from Distiller that combines all available chemical identity information (CASRN, chemical name, and abbreviation) from a reference into a single cell as our "lookup" column, and assigned "correct" CASRN, chemical name, and abbreviation identifiers. An example of the columns used in a lookup table is below (Table 6). Assigning this "correct" information serves two purposes: 1) to QA issues to chemical identity, when issues arise (such as a name and a CAS that don't match, or an abbreviation that is not unique to a single chemical), and 2) to provide a single set of chemical identifiers to be used to create visualizations/heat maps in Excel (including combining PDF data with database data).

- **Note:** when deduplicating the column to be used as a lookup, it is important to remove duplicates using Power Query and not the Remove Duplicates function in Excel. Power Query recognizes the difference between capital and lowercase letters, which the Remove Duplicates function in Excel does not. Since Power Query is what is used to look up values from the Distiller Data in the lookup table, retaining records that differ only by a capital letter is essential; removing such records will result in many missing data.

Table 5. Example columns from a lookup table. Note the removal of a comma from the CAS number, and the replacement of a non-unique abbreviation with a unique one

Combined chemical identifiers	Preferred CAS	Preferred Chemical Name	Preferred Abbreviation
-------------------------------	---------------	-------------------------	------------------------

13,674-87-8-Tris(1,3-dichloro-2-propyl) phosphate-TDCPP-Polyhalogenated Organophosphates	13674-87-8	Tris(1,3-dichloro-2-propyl) phosphate	TDCIPP
--	------------	---------------------------------------	--------

Once the lookup table is complete, it can be merged with the long format of the Distiller data using the Merge Queries function in Power Query. This ensures the correct chemical identity information and the correct tags are applied to the correct references.

Once correct chemical identification has been assigned using the lookup table, a Pivot Table can be used to visualize the data in a heatmap. A key step in setting up a Pivot Table is to check the box to “Add this data to the Data Model.” This enables Excel to “know” when a reference has been tagged multiple times for an OFR or a data type tag, and to account for this in the totals it presents in the Pivot Table (i.e., it enables the values presented in the Pivot Table to be a count of references with each combination of OFR and data tag, rather than the total number of times that combination occurred regardless of which reference(s) it occurred in). For example, if a chemical is entered twice in Distiller (in error), adding data to the data model enables Excel to only count the chemical once, despite the chemical being associated with the reference multiple times. Another example would be wanting to count how many references were tagged for a given data type, regardless of how many chemicals were present in each reference.

Pivot tables are set up to reflect the number of references tagged for each combination of Level 3 tag and OFR or (as appropriate) analog. These pivot tables are the precursors to evidence maps for the PDF data. A color scale for data richness is applied, which is also used in the final evidence maps (Section 7): low (red, <=5 references), medium (yellow, between 5 and 20 (inclusive) references), and high (green, 21 or more references). Modeled data will be indicated by numbers and column headers in italic font.

6.2.2 Selection of studies for 3B tagging

A number of criteria were considered for the selection of studies for 3B tagging. Some studies were flagged as possible key studies for 3B tagging during earlier levels of screening/tagging. For toxicity studies, key authoritative documents (e.g. IRIS assessments) are important. In discussions with CPSC, the following were identified as key criteria for choosing 3B studies:

- For studies tagged for toxicity data, the coverage of many endpoints was deemed important.
- For studies tagged for exposure and risk data, the coverage of many chemicals was deemed important.
- Despite the emphasis on endpoints for toxicity data and number of chemicals for exposure/risk data in the previous two points, both criteria (number of endpoints/tags and number of chemicals) are used for all data types (toxicity, exposure, risk).
- Studies that include data on chemicals of medium or poor data richness were deemed to

be important for inclusion.

- Exposure studies from Asia were generally not included.
- Reviews and assessments on consumer product exposure were emphasized.
- Reviews and assessments focusing on children's risk (toxicity or exposure) were emphasized.
- More recent reviews were emphasized over older reviews.
- Where there were several studies emphasizing different 3B tags (e.g., concentrations in dust, air, and water), a distribution of study types was attempted.
- Similarly, for toxicity data, a goal was to include reviews focusing on different organs or systems.

Based on these criteria, the following process was developed for identifying studies for 3B tagging:

- Identify studies that address the OFRs that are data-poor or have medium data richness.
- Based on these studies, identify the studies most likely to give broad coverage of the OFRs.
- Fill in up to the desired target number, recognizing that many of the remaining 3B references will address primarily the data-rich chemicals. Things to consider in choosing studies:
 - All other things being equal, prefer more recent references
 - Aim for wide diversity in tags, and covering key toxicity and exposure tags (e.g., covering indoor air, dust, reproductive and neurotoxicity)
 - Studies relevant to children (toxicity, exposure) are of greater interest
 - Studies on exposure in Asia are of less interest
 - Studies covering multiple toxicity endpoints are of interest.
 - Studies on consumer product exposure were emphasized.

A total of 25 references (including at least 10 each addressing toxicity and exposure) are tagged at the 3B level for data-rich classes. For data-poor classes, 9 references (including at least 3 each for toxicity and exposure) are tagged for 3B.

7. Evidence Mapping

Evidence maps of database data are prepared on the dataset scale. Separate evidence maps are prepared for toxicity and exposure data, based on Level 2, Level 3, and Level 3B tagging. Note that most of the Level 3B data are from database data, and only a subset of the PDF data is tagged at Level 3B. Evidence maps are created from data tables using conditional formatting. The sample evidence map copied below (Figure 61.) and provided by CPSC staff as part of early scoping, illustrates this type of visualization.

For differentiating modeled vs. experimental data, overlaid shading and patterns were investigated; this approach was rejected because it made the numbers in each cell difficult to read. For the same reason, using data bars to visually highlight the quantity of data indicated by each cell was also rejected.

Polyhalogenated Bisphenol Aliphatics and Functionalized (PHBAFs)	CAS	ANIMAL TOXICITY OR ACCEPTED ALTERNATIVE	HUMAN TOXICITY	HUMAN, ANIMAL, OR MODELED TOXICOKINETICS (ADME)	EXPERIMENTAL MECHANISTIC (CELL OR TISSUE-BASED, TRANSCRIPTIONS; ALTERNATIVE SPECIES)	QSAR, READ-ACROSS, ANALOG	QUALITATIVE HAZARD CHARACTERIZATION	QUANTITATIVE HAZARD CHARACTERIZATION	HUMAN HEALTH RISK ASSESSMENT
3,3',5,5'-Tetrabromobisphenol A	79-94-7	289	15	66	6766	61	92	277	30
Tetrabromobisphenol A-bis(2,3-dibromopropyl ether)	21850-44-2	41	2	12	544	48	14	24	1
1,1'-(Isopropylidene)bis(3,5-dibromo-4-(2,3-dibromo-2-methylpropoxy)benzene)	97416-84-7	27	0	6	1	56	7	24	0
2,2',6,6'-Tetrachlorobisphenol A	79-95-8	10	0	7	1603	54	5	17	0
Tetrabromobisphenol A bis(2-hydroxyethyl) ether	4162-45-2	10	0	6	751	51	20	10	1
2,2'-[(1-Methylethylidene)bis[(2,6-dibromo-4,1-phenylene)oxymethylene]]bis[oxirane]	3072-84-2	17	0	5	6	49	3	9	0
Tetrabromobisphenol A diallyl ether	25327-89-3	15	0	5	389	48	7	8	1
Tetrabromobisphenol A dimethyl ether	37853-61-5	0	0	6	4	47	0	4	0
1,1'-sulfonylbis(3,5-dibromo-4-methoxybenzene)	70156-79-5	3	0	0	0	17	1	3	0
Bis(p-acryloxyethoxy)tetrabromobisphenol A	66710-97-2	1	0	5	0	48	4	0	0

Phenol, 4,4'-(1-methylethylidene)bis[2,6-dibromo-, 1,1'-diacetate	3379 8- 02-6	1	0	5	0	37	1	0	0
4,4'-(propane-2,2-diyl)bis(2-bromophenol)	2942 6- 78-6	0	0	1	0	25	0	0	0
3,3',5,5'-Tetrabromobisphenol A bispropionate	3741 9- 42-4	0	0	0	0	23	0	0	0
(Propane-2,2-diyl)bis(2,6-dibromo-4,1-phenylene) diprop-2-enoate	5520 5- 38-4	0	0	0	0	18	0	0	0

Figure 61. Example of Evidence Map

A color scale for data richness is applied, similar to that used in the Pivot Tables for PDF data (Section 6.2.1): with darker blue indicating more data is available, lighter blue indicating less data is available, and no-color shading when no data is available for that data type and chemical substance.