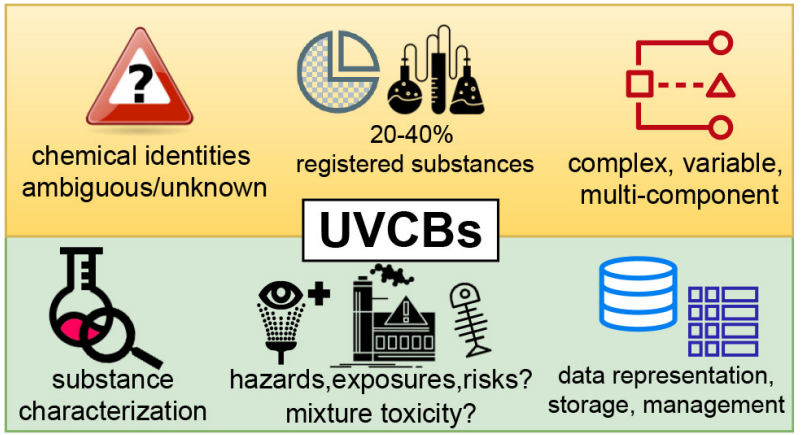
Luxembourg National Research Fund
Funded by the European Union

Integrating UVCBs and Related Data into Open Chemical Knowledgebases

Emma L. Schymanski, Anjana Elapavalore
Environmental Cheminformatics, Luxembourg Centre for Systems Biomedicine, University of Luxembourg
Contact: emma.schymanski@uni.lu Twitter: @ESchymanski

Evan E. Bolton, Qingliang Li, Paul A. Thiessen, Leonid Zaslavsky, Jian Zhang
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health



Lai *et al.* (2022) ES&T. DOI: 10.1021/acs.est.2c00321

HESI® UVCBs and Multi-Constituent Substances (MCS) Committee

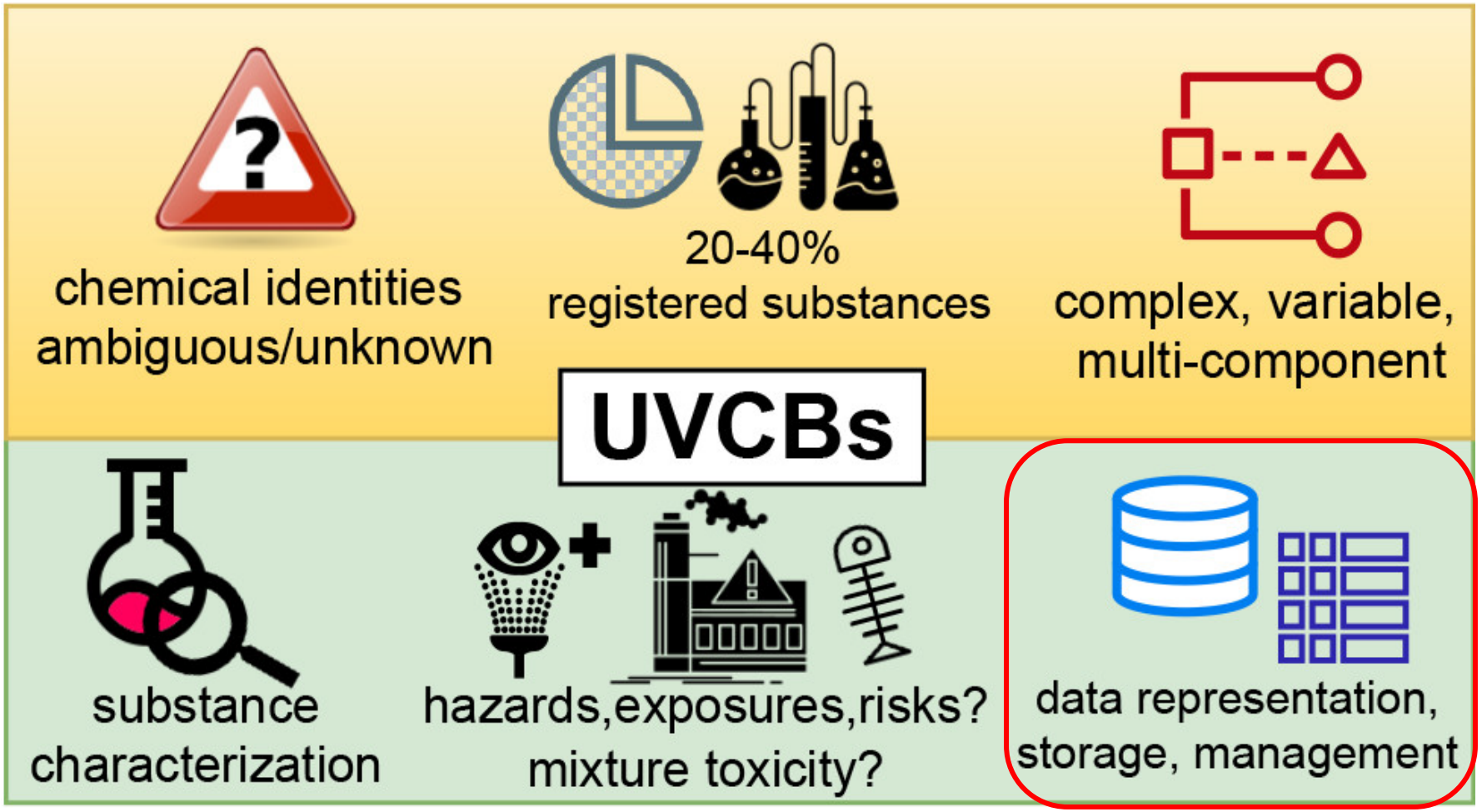
Workshop, 18-19 Sept 2023 (presented remotely)

Slides available at DOI: 10.5281/zenodo.8302440

Today's slides zenodo

What are UVCBs?

Unknown or variable composition, complex reaction products, or biological materials



Lai *et al.* (2022) ES&T. DOI: 10.1021/acs.est.2c00321

Why are UVCBs Challenging?

COMPOUND SUMMARY

Polyethylene Glycol 300

See also: Ethylene Oxide (has monomer).

PubChem CID	Not available because this is not a discrete structure.
Description	Polyethylene glycol 300 (PEG 300) is a water-miscible polyether with an average molecular weight of 300 g/mol. It is a clear viscous liquid at room temperature with non-volatile, stable properties. Polyethylene glycols are widely used in biochemistry, structural biology, and medicine in addition to pharmaceutical and chemical industries. They serve as solubilizers, excipients, lubricants, and chemical reagents. Low molecular weight glycols are observed to exhibit antibacterial properties as well. PEG 300 is found in eye drops as a lubricant to temporarily relieve irritation of the eyes.

▸ DrugBank

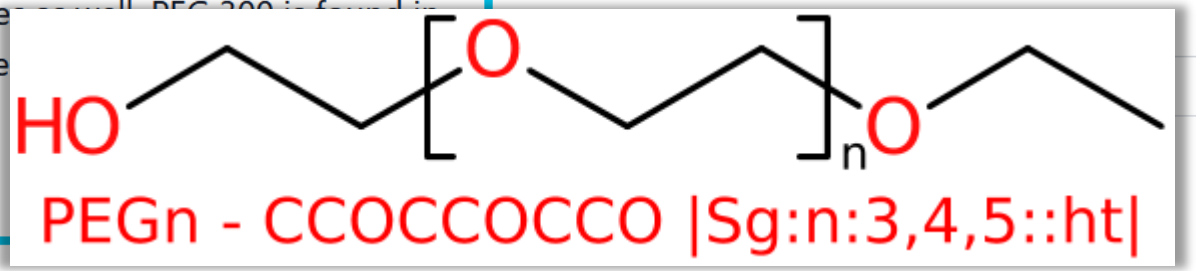
PubChem Ethylene Oxide (Compound)

5.4 Other Relationships

PEG-4 diethylhexanoate (monomer of)	Polyethylene Glycol 600 (monomer of)	Cetyl peg/ppg-10
PEG-8 stearate (monomer of)	Polyethylene Glycol 7000 (monomer of)	Demplatin Pegrag
PEG/PPG-20/6 dimethicone (monomer of)	Polyethylene Glycol 800 (monomer of)	Distearth-100 isc
Pegamotecan (monomer of)	Polyethylene Glycol 8000 (monomer of)	Firtecan Pegol (m
Poloxamer 188 (monomer of)	Polyethylene oxide 2000000 (monomer of)	Glycereth-12 (mo
Poloxamer 407 (monomer of)	Beheneth-10 (monomer of)	Glycereth-18 ethy
Polyethylene Glycol 1000 (monomer of)	Bis-PEG-18 methyl ether dimethyl silane (monomer of)	Glycereth-26 (mo
Polyethylene Glycol 1450 (monomer of)	Bis-peg/ppg-14/14 dimethicone (monomer of)	Glycereth-7 trime
Polyethylene Glycol 4000 (monomer of)	Ceteth-10 phosphate (monomer of)	Hydroxyethyl Cell
Polyethylene Glycol 4500 (monomer of)	Ceteth-20 (monomer of)	Hydroxyethyl cellu

▸ PubChem

- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Related Records
- 5 Drug and Medication Information
- 6 Pharmacology and Biochemistry
- 7 Use and Manufacturing
- 8 Toxicity



https://pubchem.ncbi.nlm.nih.gov/compound/Polyethylene%20Glycol%20300
https://pubchem.ncbi.nlm.nih.gov/compound/Ethylene%20Oxide#section=Other-Relationships

Expert Knowledge: NORMAN Suspect List Exchange

https://www.norman-network.com/nds/SLE/

Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6


NORMAN Database System

NORMAN organises the development and maintenance of var

SEARCH All Databases

Searching for individual substance or group(s) of substances

Note: Click on a link below to go to an individual database h

NORMAN Suspect List Exchang

The NORMAN Suspect List Exchange (NORMAN-SLE) was e
questions. The NORMAN-SLE documents all individual collec
(see Source column in SusDat). NORMAN-SLE versions are t

Comments and contributions are welcome - please email us a

Please refer to our documentation pages for: citation instructi

No.	Abbreviation	Description	Lin
S0	SUSDAT	**Merged NORMAN Suspect List: SusDat**	Inte

Substance Database

A merged list of NORMAN substances; Central Database to access various lists of substances for suspect screening and prioritisation

Suspect List Exchange

Central Database to access various lists of substances for suspect screening and prioritisation

Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmen
matrices

RESEARCH Open Access

Check for updates

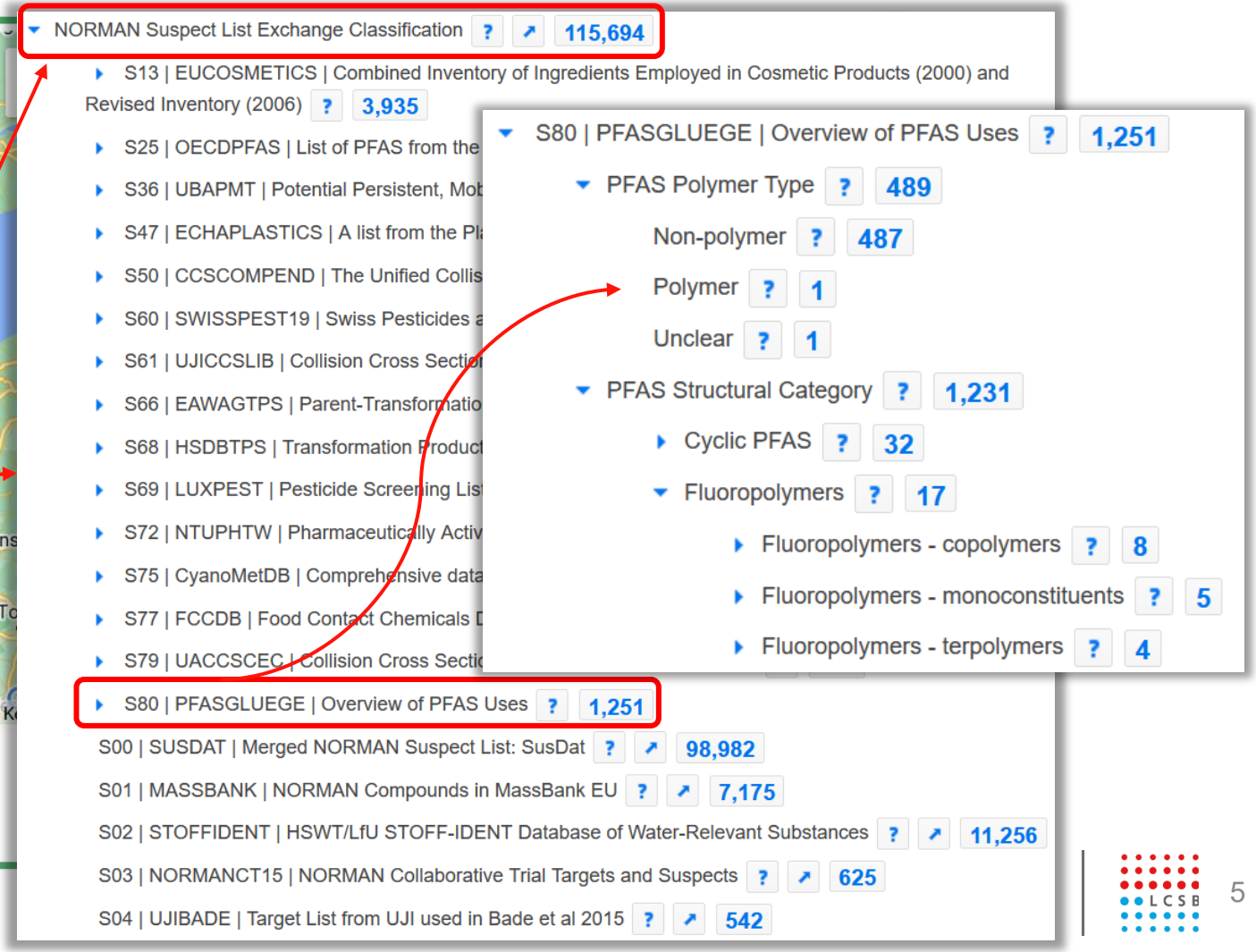
The NORMAN Suspect List Exchange (NORMAN-SLE): facilitating European and worldwide collaboration on suspect screening in high resolution mass spectrometry

Hiba Mohammed Taha[1], Reza Aalizadeh[2], Nikiforos Alygizakis[3,2], Jean-Philippe Antignac[4], Hans Peter H. Arp[5,6], Richard Bade[7], Nancy Baker[8], Lidia Belova[9], Lubertus Bijlsma[10], Evan E. Bolton[11], Werner Brack[12,13], Alberto Celma[10,14], Wen-Ling Chen[15], Tiejun Cheng[11], Parviel Chirsir[1], Ľuboš Čirka[16,3], Lisa A. D'Agostino[17], Yannick Djoumbou Feunang[18], Valeria Dulio[19], Stellan Fischer[20], Pablo Gago-Ferrero[21], Aikaterini Galani[2], Birgit Geueke[22], Natalia Głowacka[3], Juliane Glüge[23], Ksenia Groh[24], Sylvia Grosse[25], Peter Haglund[26], Pertti J. Hakkinen[11], Sarah E. Hale[5], Felix Hernandez[10], Elisabeth M.-L. Janssen[24], Tim Jonkers[27], Karin Kiefer[24], Michal Kirchner[28], Jan Koschorreck[29], Martin Krauss[12], Jessy Krier[1], Marja H. Lamoree[27], Marion Letzel[30], Thomas Letzel[31], Qingliang Li[11], James Little[32], Yanna Liu[33], David M. Lunderberg[34,35], Jonathan W. Martin[17], Andrew D. McEachran[36], John A. McLean[37], Christiane Meier[29], Jeroen Meijer[38], Frank Menger[14], Carla Merino[39,40], Jane Muncke[22], Matthias Muschket[12], Michael Neumann[29], Vanessa Neveu[41], Kelsey Ng[3,42], Herbert Oberacher[43], Jake O'Brien[7], Peter Oswald[3], Martina Oswaldova[3], Jaqueline A. Picache[37], Cristina Postigo[44,14], Noelia Ramirez[45,39], Thorsten Reemtsma[12], Justin Renaud[46], Pawel Rostkowski[47], Heinz Rüdel[48], Reza M. Salek[41], Saer Samanipour[49], Martin Scheringer[23,42], Ivo Schliebner[29], Wolfgang Schulz[50], Tobias Schulze[12], Manfred Sengl[30], Benjamin A. Shoemaker[11], Kerry Sims[51], Heinz Singer[24], Randolph R. Singh[1,52], Mark Sumarah[46], Paul A. Thiessen[11], Kevin V. Thomas[7], Sonia Torres[39], Xenia Trier[53], Annemarie P. van Wezel[54], Roel C. H. Vermeulen[38], Jelle J. Vlaanderen[38], Peter C. von der Ohe[29], Zhanyun Wang[55], Antony J. Williams[56], Egon L. Willighagen[57], David S. Wishart[58], Jian Zhang[11], Nikolaos S. Thomaidis[2], Juliane Hollender[23,24], Jaroslav Slobodnik[3] and Emma L. Schymanski[1*]

NORMAN Suspect List Exchange in PubChem

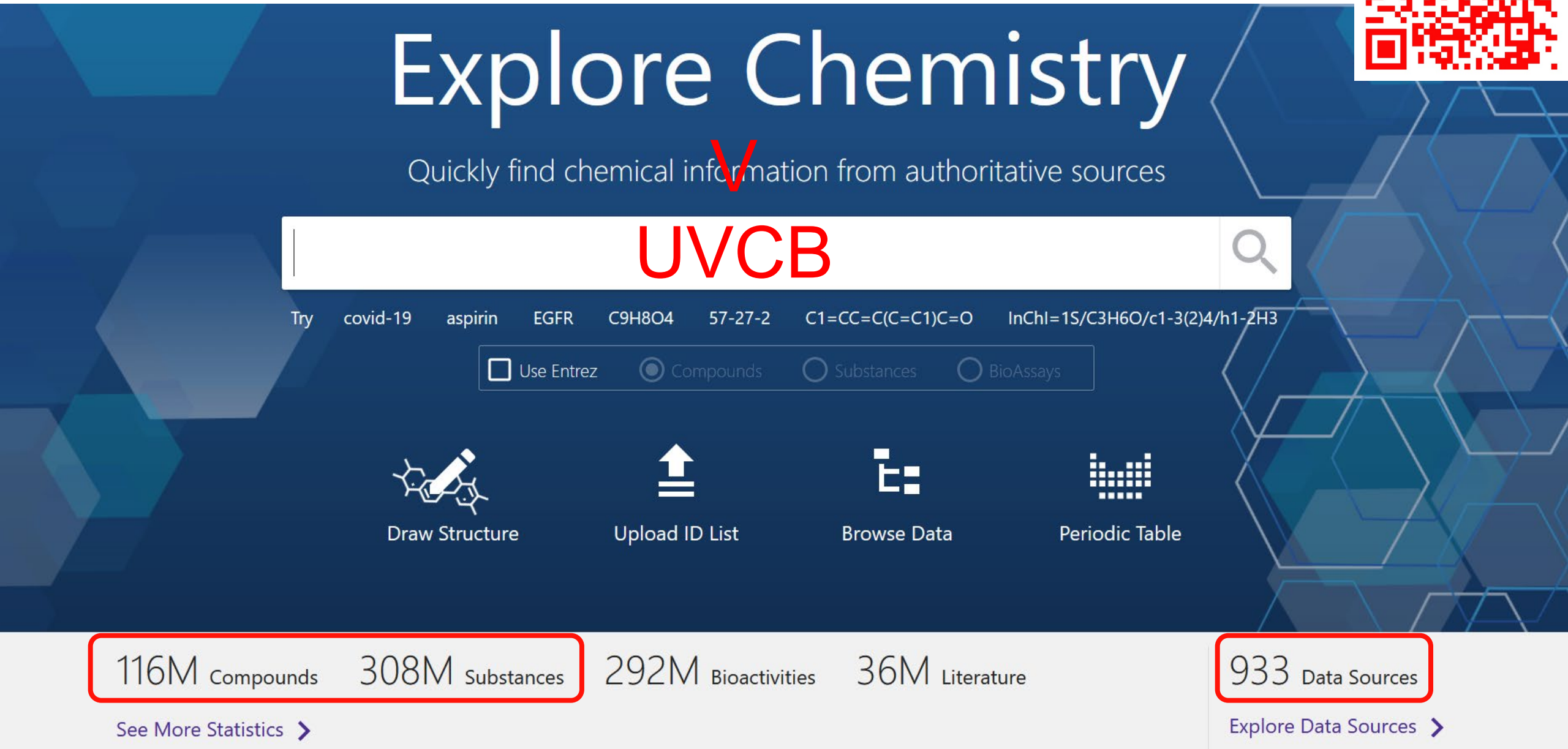
The NORMAN network enhances the exchange of information on emerging environmental substances, and encourages the validation and harmonisation of common measurement methods and monitoring tools so that the requirements of risk assessors and risk managers can be better met. The NORMAN Suspect List Exchange (NORMAN-SLE) is a central access point to find suspect lists relevant for various environmental monitoring questions, described in DOI:10.1186/s12302-022-00680-6

Organization	NORMAN Network (c/o UniLu)
Category	Research and Development
URL	https://www.norman-network.com/nds/SLE/
License Note	Data: CC-BY 4.0; Code (hosted by ECI, LCSB): Artistic-2.0
License URL	https://creativecommons.org/licenses/by/4.0/
Contact Name	Emma Schymanski
Address	6 avenue du Swing, Belvaux, Luxembourg, 4367
Data Source ID	23819
Data in PubChem	118,487 Live Substances 22,314 Annotations 1 Classification
Last Updated	2023/08/23



https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101

What is PubChem? https://pubchem.ncbi.nlm.nih.gov/



Recognising UVCBs & Automated Curation

- Substance name is the most widely available identifier of UVCBs across all databases[1] => text-based recognition (aka “RegEx”)

Gather authoritative UVCB names

GSRS

Chemicals subject to TSCA

ECHA EUROPEAN CHEMICALS AGENCY

~62,000 “UVCB concepts”

~500,000 synonyms (names)

[1]Lai *et al.* (2022) ES&T. DOI: 10.1021/acs.est.2c00321. Datasets: FDA GSRS, US EPA TSCA, ECHA

Recognising UVCBs & Automated Curation

- Substance name is the most widely available identifier of UVCBs across all databases[1] => text-based recognition (aka “RegEx”)

Gather authoritative UVCB names

GSRS

Chemicals subject to TSCA

ECHA EUROPEAN CHEMICALS AGENCY

Generate RegEx to recognise UVCBs (excerpt from “c” to “d”)

```
elsif ( $s =~ /charcol/i ) { $is_uvcb = 1; $why = "charcol"; }
elsif ( $s =~ /chlorinated/i ) { $is_uvcb = 1; $why = "chlorinated"; }
elsif ( $s =~ /complex/i ) { $is_uvcb = 1; $why = "complex"; }
elsif ( $s =~ /compounds/i ) { $is_uvcb = 1; $why = "compounds"; }
elsif ( $s =~ /compd\./i ) { $is_uvcb = 1; $why = "compd."; }
elsif ( $s =~ /compds\./i ) { $is_uvcb = 1; $why = "compds."; }
elsif ( $s =~ /concentrate/i ) { $is_uvcb = 1; $why = "concentrate"; }
elsif ( $s =~ /condensed/i ) { $is_uvcb = 1; $why = "condensed"; }
elsif ( $s =~ /condensation/i ) { $is_uvcb = 1; $why = "condensation"; }
elsif ( $s =~ /distillate/i ) { $is_uvcb = 1; $why = "distillate"; }
elsif ( $s =~ /dervs/i ) { $is_uvcb = 1; $why = "dervs"; }
elsif ( $s =~ /derivs/i ) { $is_uvcb = 1; $why = "derivs"; }
elsif ( $s =~ /derivatives/i ) { $is_uvcb = 1; $why = "derivatives"; }
elsif ( $s =~ /diluent/i ) { $is_uvcb = 1; $why = "diluent"; }
```

[1]Lai *et al.* (2022) ES&T. DOI: 10.1021/acs.est.2c00321. Datasets: FDA GSRS, US EPA TSCA, ECHA

Recognising UVCBs & Automated Curation

Gather
authoritative
UVCB names

GSRS

Chemicals subject to TSCA

ECHA
EUROPEAN CHEMICALS AGENCY

Generate RegEx to recognise UVCBs
(excerpt from “c” to “d”)

```
elsif ( $s =~ /charcol/i ) { $is_uvcb = 1; $why = "charcol"; }
elsif ( $s =~ /chlorinated/i ) { $is_uvcb = 1; $why = "chlorinated"; }
elsif ( $s =~ /complex/i ) { $is_uvcb = 1; $why = "complex"; }
elsif ( $s =~ /compounds/i ) { $is_uvcb = 1; $why = "compounds"; }
elsif ( $s =~ /compd\./i ) { $is_uvcb = 1; $why = "compd."; }
elsif ( $s =~ /compds\./i ) { $is_uvcb = 1; $why = "compds."; }
elsif ( $s =~ /concentrate/i ) { $is_uvcb = 1; $why = "concentrate"; }
elsif ( $s =~ /condensed/i ) { $is_uvcb = 1; $why = "condensed"; }
elsif ( $s =~ /condensation/i ) { $is_uvcb = 1; $why = "condensation"; }
elsif ( $s =~ /distillate/i ) { $is_uvcb = 1; $why = "distillate"; }
elsif ( $s =~ /dervs/i ) { $is_uvcb = 1; $why = "dervs"; }
elsif ( $s =~ /derivs/i ) { $is_uvcb = 1; $why = "derivs"; }
elsif ( $s =~ /derivatives/i ) { $is_uvcb = 1; $why = "derivatives"; }
elsif ( $s =~ /diluent/i ) { $is_uvcb = 1; $why = "diluent"; }
```

Validate on
various datasets

PubChemLite EXPOSOMICS

norman suspects

PubChem

UVCB datasets: FDA GSRS, US EPA TSCA, ECHA. Validation datasets: PubChemLite, NORMAN-SLE and PubChem

Recognising UVCBs & Automated Curation

- 113 regular expressions after PubChemLite and NORMAN-SLE validation

Expression	Example - detected expression in **bold**, plus other matches	
“ tree”	Camphor **tree** whole	(see next slides)
“chlorinated”	1-Propene, tetramer, **chlorinated**	
C#-C# Regex	Carboxylic acids, di-, **C4-11**	(see next slides)
“Amines”	**Amines**, di-C14-18-alkylmethyl	(see next slides)

- Only 8 removed following validation on whole PubChem*

```
*fruit, mace, resin, rosin, shell, solvent, steroid, tannin
```

RegEx: https://gitlab.lcsb.uni.lu/eci/pubchem/-/tree/master/annotations/UVCBs/regex
UVCB datasets: FDA GSRS, US EPA TSCA, ECHA. Validation datasets: PubChemLite, NORMAN-SLE and PubChem.

Variable: Connecting Chemicals & UVCBs (e.g. “C#-C# RegEx”)

Expression	Example - detected expression in **bold**, plus other matches	
C#-C# Regex	Carboxylic acids, di-, **C4-11**	(see next slides)
“Amines”	**Amines**, di-C14-18-alkylmethyl	(see next slides)

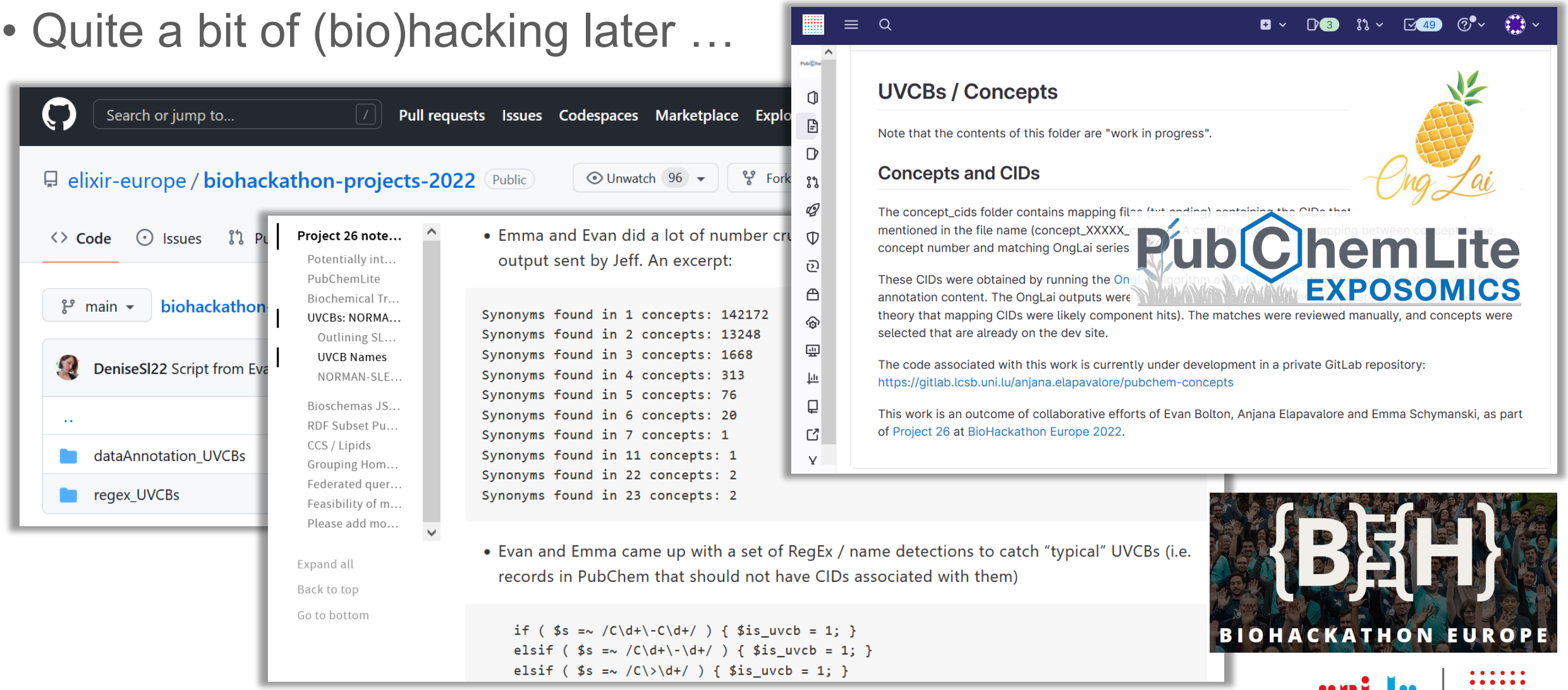
- Homologous series are (relatively) easy to recognise and map up

- Many concepts had ≥1 “representative” already associated with them

https://github.com/adelenelai/onglai-classify-homologues
Lai *et al.* (2022) *J Cheminform.* 14:85. DOI: 10.1186/s13321-022-00663-y

Variable: Connecting Chemicals & UVCBs (e.g. “C#-C# RegEx”)

- Quite a bit of (bio)hacking later …



https://github.com/elixir-europe/biohackathon-projects-2022/tree/main/26/ & https://hackmd.io/CR8cv0XpSbOu_1j1oirNHQ
& https://gitlab.lcsb.uni.lu/eci/pubchem/-/tree/master/annotations/UVCBs/

Variable: Connecting Chemicals & UVCBs (e.g. “C#-C# RegEx”)

PubChem Carboxylic acids, di-, C4-11 (Compound)

5.2 U.S. Production

Aggregated Product Volume

2019: <1,000,000 lb

2018: <1,000,000 lb

2017: <1,000,000 lb

2016: 1,000,000 lb - <20,000,000 lb

https://www.epa.gov/chemical-data-reporting

▸ EPA Chemicals under the TSCA

5.3 General Manufacturing Information

Industry Processing Sectors

All Other Basic Organic Chemical Manufacturing

▸ EPA Chemicals under the TSCA

EPA TSCA Commercial Activity Status

Carboxylic acids, di-, C4-11: ACTIVE

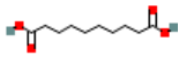
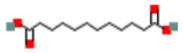
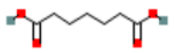
4 Related Records

UVCB Carboxylic acids, di-, C4-11

8 items

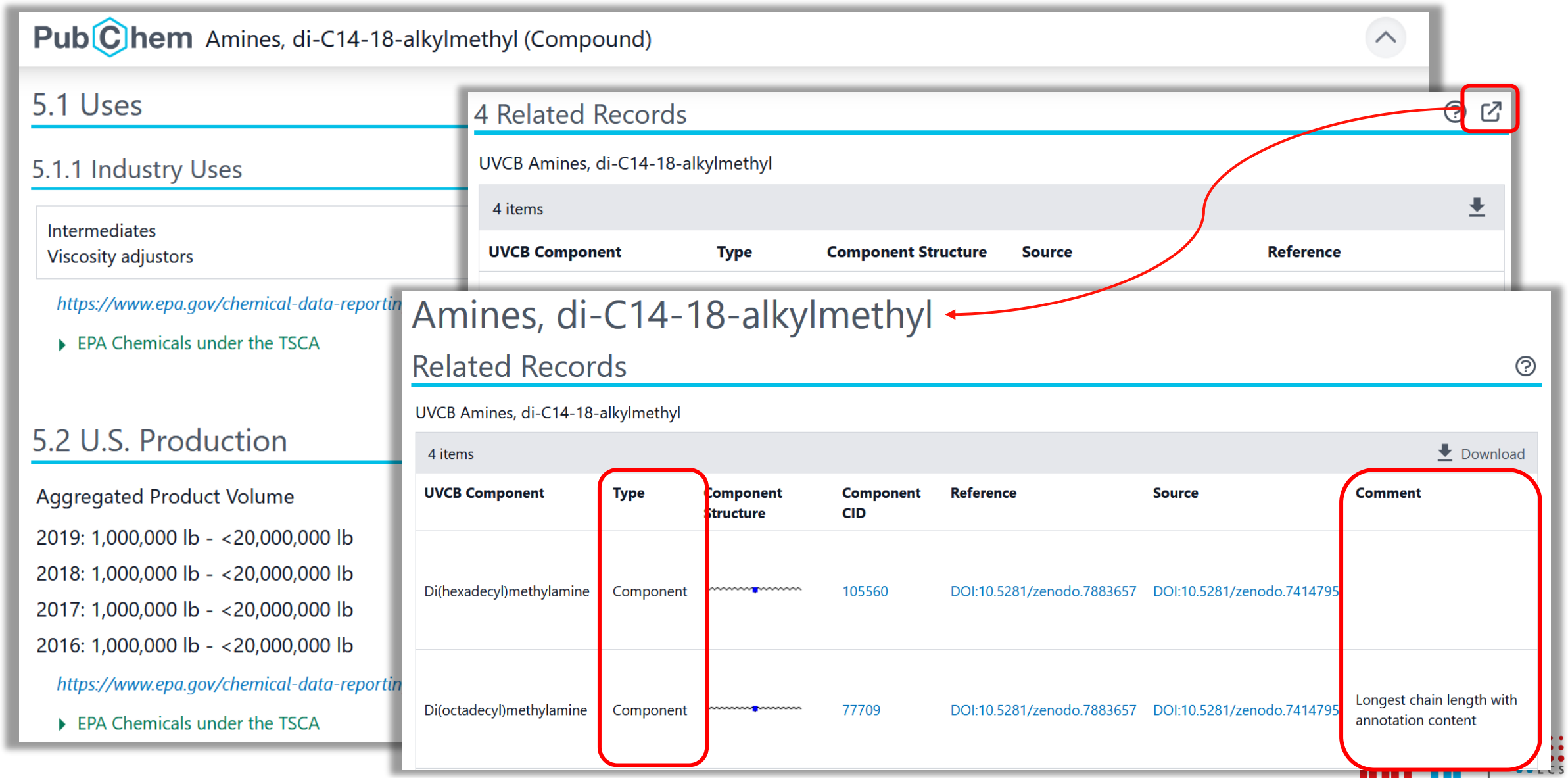
Search

SORT BY UVCB Component - A to Z

UVCB Component	Type	Component Structure	Source	Reference
Decanedioic acid	Component		DOI:10.5281/zenodo.7414795	DOI:10.5281/zenodo.7883657
Dodecanedioic acid	Component		DOI:10.5281/zenodo.7414795	DOI:10.5281/zenodo.7883657
Heptanedioic acid	Component		DOI:10.5281/zenodo.7414795	DOI:10.5281/zenodo.7883657

https://pubchem.ncbi.nlm.nih.gov/compound/Carboxylic%20acids,%20di-,%20C4-11#section=U-S-Production

Variable: Connecting Chemicals & UVCBs (e.g. “C#-C# RegEx”)



https://pubchem.ncbi.nlm.nih.gov/compound/Amines,%20di-C14-18-alkylmethyl#section=Uses

Biological: Associating Biological Information & UVCBs

Expression	Example - detected expression in **bold**, plus other matches
“ tree”	Camphor **tree** whole

PubChem Cinnamomum camphora (camphor tree) (Taxonomy)

3.2 Natural Products

246 items View More Rows & Details Download

SORT BY Compound CID

Structure	Compound CID	Compound	Evidence IDs	Data Source
	2537	Camphor		NPASS
	2537	Camphor	DOI:10.1590/S1516-89132000000300011 DOI:10.1080/10412905.1998.9700962 DOI:10.1093/JAT/9.1.24 DOI:10.1080/10412905.1993.9698258	LOTUS
	2758	Eucalyptol	DOI:10.1007/S11418-006-0039-1	LOTUS

PubChem camphor tree whole

Substances (37) | Proteins (2) | Taxonomy (1) | Pathways (2) | Literature (130)

Searching chemical names and synonyms in the substance records submitted by PubChem's contributors. Read

37 results Filters SORT BY Relevance

Camphor; D-CAMPHOR; DL-Camphor; Camphora; ...; Camphora Tree; ...

Substance SID: 123094738 Compound CID: 159055

Data Source: ChEMBL External ID: CHEMBL504760

Data Source Category: Curation Efforts; Research and Development

Deposit Date: 2011-06-16 Last Modified Date: 2023-03-02

Structure not available

Camphorwood; Laurus Camphora; Camphor Laurel; Camphora Tree; Camphor Whole; ...

Substance SID: 472387794

Data Source: FDA Global Substance Registration System (GSRS)

External ID: 0B27814T7X

https://pubchem.ncbi.nlm.nih.gov/#query=Camphor%20tree&tab=substance & https://pubchem.ncbi.nlm.nih.gov/taxonomy/13429

Biological: Extracting & Using Related Chemical Information

Expression	Example - detected expression in bold
“ tree”	Camphor **tree** whole (see

PubChem Cinnamomum camphora (camphor tree) (Taxonomy)

3.2 Natural Products

246 items View More Rows & Details

Download

Structure	Compound CID	Compound
	2537	Camphor
	2758	Eucalyptol

3.1 Metabolites

160 items View More Rows & Details

SORT BY Compound C

Structure	Compound CID	Compound	Evidence IDs
	931	Naphthalene	
	3314	Eugenol	PMID:10743224

KNApSAcK

Retrieving *Camphor tree* Chemicals from PubChem

Emma SCHYMANSKI

15/04/2023

Background

Camphor is an example of an ambiguous name that can refer to a discrete chemical entity (camphor) or more abs… ple of a “biologica… products or biologi…) via the embedded … components related to PubChem …

Associate… chemicals … note that not all of …

For *Cinn…* … of most interest (see Figure 1).

Automated extraction & formatting for workflows (e.g. RMarkdown)

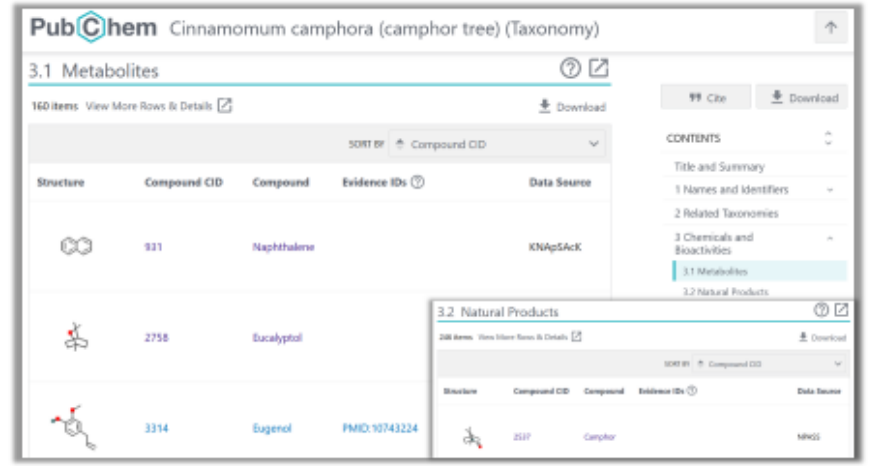
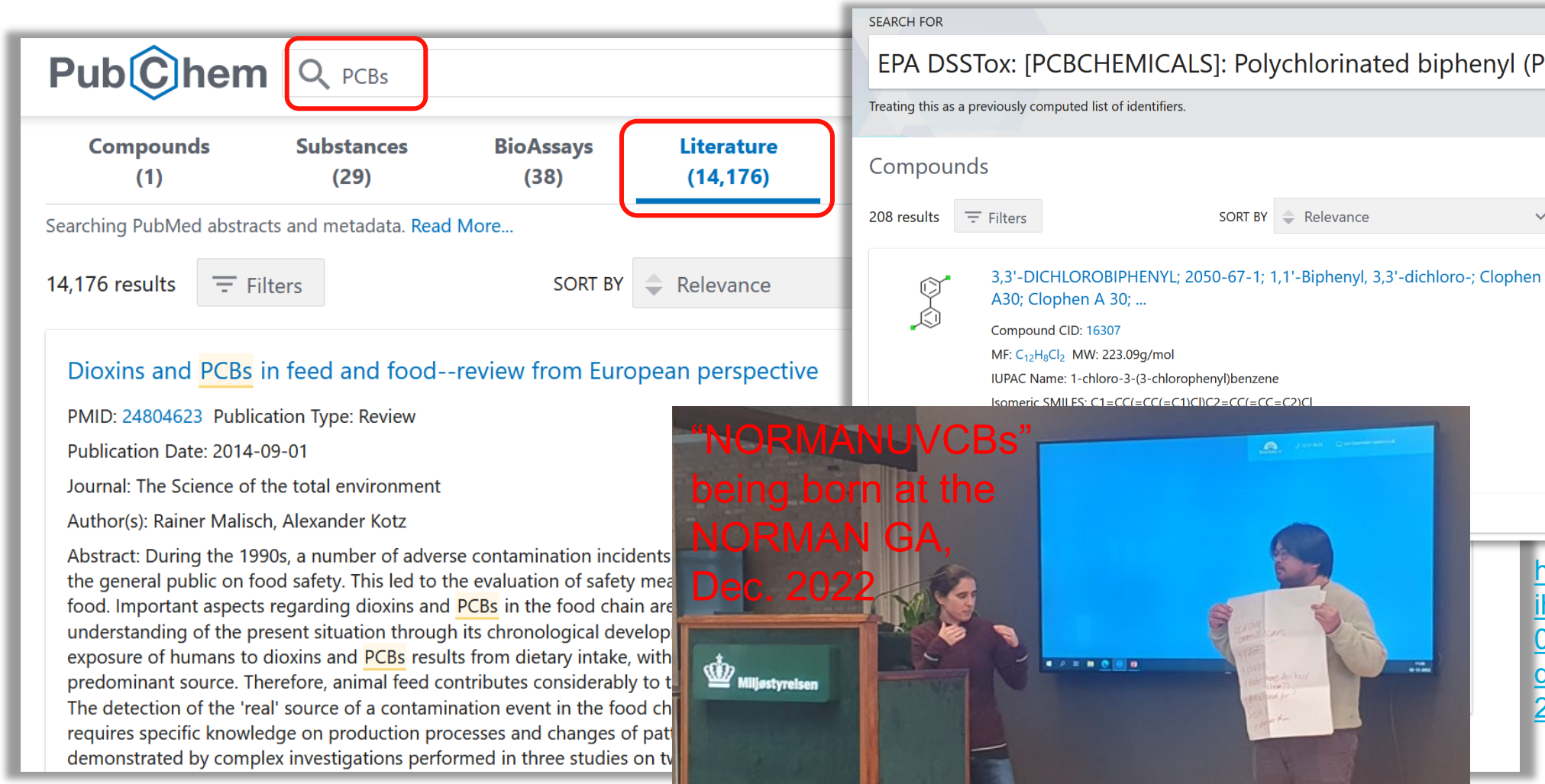


Figure 1: *Associated chemicals of interest for Cinnamomum camphora on PubChem*

https://pubchem.ncbi.nlm.nih.gov/taxonomy/13429#section=Chemicals-and-Bioactivities & https://gitlab.lcsb.uni.lu/eci/pubchem-docs/-/blob/main/taxonomy/Camphor_tree/

Challenges / Perspectives – Example of PCBs

- Authoritative names don't match with "our" community names



https://pubchem.ncbi.nlm.nih.gov/classification/#hid=105&search=Polychlorinated%20biphenyl%20(PCB)%20collection&view=tree

https://pubchem.ncbi.nlm.nih.gov/#query=PCBs&tab=pubmed

NORMAN-UVCBs & UVCB Template (a little complex!)

- Adding more UVCB data with FAIR templates & open license

A	B	C	D	E	F	G	H	I	J
UVCB_Name	UVCB_Synonym	CAS_RN	EC_Num	PubChem_SID	Source	Source_ID	Reference_Description	Reference_ID	Related_UVCB
PCBs	polychlorinated biphenyls				Randolph Singh (IFREMER, France), Emma Schymanski (LCSB	ORCID:0000-0003-4500-3400||ORCID:0000-0001-6868-8145	J Haedrich et al (2021) Environ Sci Eur 33(1):33	PMID:33828936	
DL-PC									

UVCB_Name	Type	Name	Synonym	PubChem_	SMILES	InChIKey	CAS_RN	DTXSID	Compositio	Source	
PCBs	Component	2,3,3'-Trichlo	2,3,3'-Trichloro-	38034	ClC1=CC=CC(=	SXHLTVKPNQVZGL	38444-84	DTXSID8073502		Randolph Sir	
PCBs	Component	2,2',4,5',6-Per	2,2',4,5',6-Penta	63086	ClC1=CC=C(Cl	PQHZWWBJPCNN	60145-21	DTXSID5074189		Randolph Sir	
PCBs	Component	2,3,3',4,4',5,5	2,3,3',4,4',5,5'-H	38306	ClC1=CC(=CC	XUAWBXBYHDRR	39635-31	DTXSID4074144		Randolph Sir	
PCBs	Component	2,2',3,4',6'-Pe	2,2',3,4',6'-Penta	43238	ClC1=CC(Cl)=C	GOFFZTAPOOICFT	60233-25	DTXSID9074193		Randolph Sir	
PCBs	Component	3,3',4-Trichlo	3,3',4-Trichloro-	37806	ClC1=CC=CC(=	JHBVPKZLIBDTJR-	37680-69	DTXSID60865879		Randolph Sir	
PCBs	Component	2,3',5,5'-Tetra	2,3',5,5'-Tetrach	38878	ClC1=CC(=C(C	WBTMFEPLVQOW	41464-42	DTXSID8074152		Randolph Sing	ORCID:0000-0 CompTox PCBCH
PCBs	Component	2,2',3,4,6-Pen	2,2',3,4,6-Pentac	41362	ClC1=CC(Cl)=C	QGDKRLQRLFUJPF	55215-17	DTXSID6074178		Randolph Sing	ORCID:0000-0 CompTox PCBCH
PCBs	Component	2,3,3',4,5'-Per	2,3,3',4,5'-Penta	91704	ClC1=CC(=CC	MPCDNZSLJWJDN	70362-41	DTXSID1074206		Randolph Sir	
PCBs	Component	3,3',4,5'-Tetra	3,3',4,5'-Tetrach	63104	ClC1=CC(=CC	QLCTXEMDCZGPC	41464-48	DTXSID3074155		Randolph Sir	
PCBs	Component	2,3,4,4',5-Pen	2,3,4,4',5-Pentac	53036	ClC1=CC=C(C	SXZSFWHOSHAKN	74472-37	DTXSID9074226		Randolph Sir	
PCBs	Component	3,4'-Dichlorol	3,4'-Dichloro-1,1	18101	ClC1=CC=C(C	CJDNEKOMKXLSB	2974-90-5	DTXSID10863067		Randolph Sir	
PCBs	Component	2,2',3,5,6,6'-H	2,2',3,5,6,6'-Hex	63070	ClC1=CC(Cl)=C	CLODVDBWNVQL	68194-09	DTXSID70867526		Randolph Sir	
PCBs	Component	2,2',3,3',4,4',5	2,2',3,3',4,4',5,6-	40485	ClC1=C(Cl)C(C	JAHJITLFJSDRCG-	52663-78	DTXSID1074171		Randolph Sir	
PCBs	Component	2,3,4',6-Tetra	2,3,4',6-Tetrachl	63110	ClC1=CC=C(C	FXRXQYZZALWW	52663-58	DTXSID3074159		Randolph Sir	
PCBs	Component	2,3,3',4,4',5-H	2,3,3',4,4',5-Hex	38019	ClC1=CC(=C(C	LCXMEXLGMKFLQ	38380-08	DTXSID0052706		Randolph Sir	
PCBs	Component	2,2',5,5'-Tetra	2,2',5,5'-Tetrach	37248	ClC1=CC(=C(C	HCWZEPKLWVAE	35693-99	DTXSID3038305		Randolph Sir	
PCBs	Component	2,4-Dichlorob	2,4-Dichloro-1,1	36399	ClC1=CC(Cl)=C	WEJZHZJJXPXXMU	33284-50	DTXSID8040301		Randolph Sir	
PCBs	Component	3,4',5-Trichlo	3,4',5-Trichloro-	38038	ClC1=CC=C(C	SYSBNFJJSJLZMM-	38444-88	DTXSID40865913		Randolph Sing	ORCID:0000-0 CompTox PCBCH

README | UVCB | Product | **Chemicals** | Types | Notes

Letter to the Editor | Open Access | Published: 07 July 2021

FAIR chemical structures in the Journal of Cheminformatics

Emma L. Schymanski & Evan E. Bolton

Journal of Cheminformatics **13**, Article number: 50 (2021) | Cite this article

3046 Accesses | **10** Citations | **27** Altmetric | Metrics

JOURNAL ARTICLE

FAIRifying the exposome journal: Templates for chemical structures and transformations

Emma L Schymanski, Evan E Bolton

Exposome, Volume 2, Issue 1, 2022, osab006, https://doi.org/10.1093/exposome/osab006

Published: 24 December 2021 **Article history**

Schymanski, Bolton & Singh (2022) NORMANUVCB (S103.0.1.1) Zenodo. DOI:10.5281/zenodo.7584082
Schymanski & Bolton (2022) FAIR-ifying the Exposome Journal: DOI: 10.1093/exposome/osab006
Schymanski & Bolton (2021) FAIR Chemical Structures. *J. Cheminform*. DOI: 10.1186/s13321-021-00520-4

NORMAN-UVCBs & Polychlorinated Biphenyls

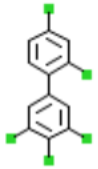
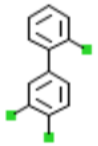
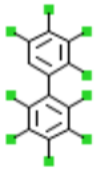
UVCB Polychlorinated biphenyls

209 items

Download

Search

SORT BY UVCB Component - A to Z

UVCB Component	Type	Component Structure	Component CID	Reference	Source
2',3,4,4',5-Pentachlorobiphenyl	Component		47650	https://comptox.epa.gov/dashboard/chemical-lists/PCBCHEMICALS	DOI:10.5281/zenodo.7414795
2',3,4-Trichlorobiphenyl	Component		38036	https://comptox.epa.gov/dashboard/chemical-lists/PCBCHEMICALS	DOI:10.5281/zenodo.7414795
2,2',3,3',4,4',5,5',6-Nonachlorobiphenyl	Component		38411	https://comptox.epa.gov/dashboard/chemical-lists/PCBCHEMICALS	DOI:10.5281/zenodo.7414795

First Previous Page 1 of 21 Next Last

https://pubchem.ncbi.nlm.nih.gov/compound/Polychlorinated%20Biphenyls#section=Related-Records&fullscreen=true

NORMAN-UVCBs & Polychlorinated Biphenyls

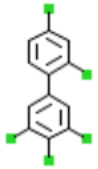
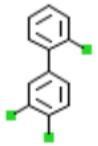
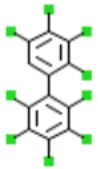
UVCB Polychlorinated biphenyls

209 items

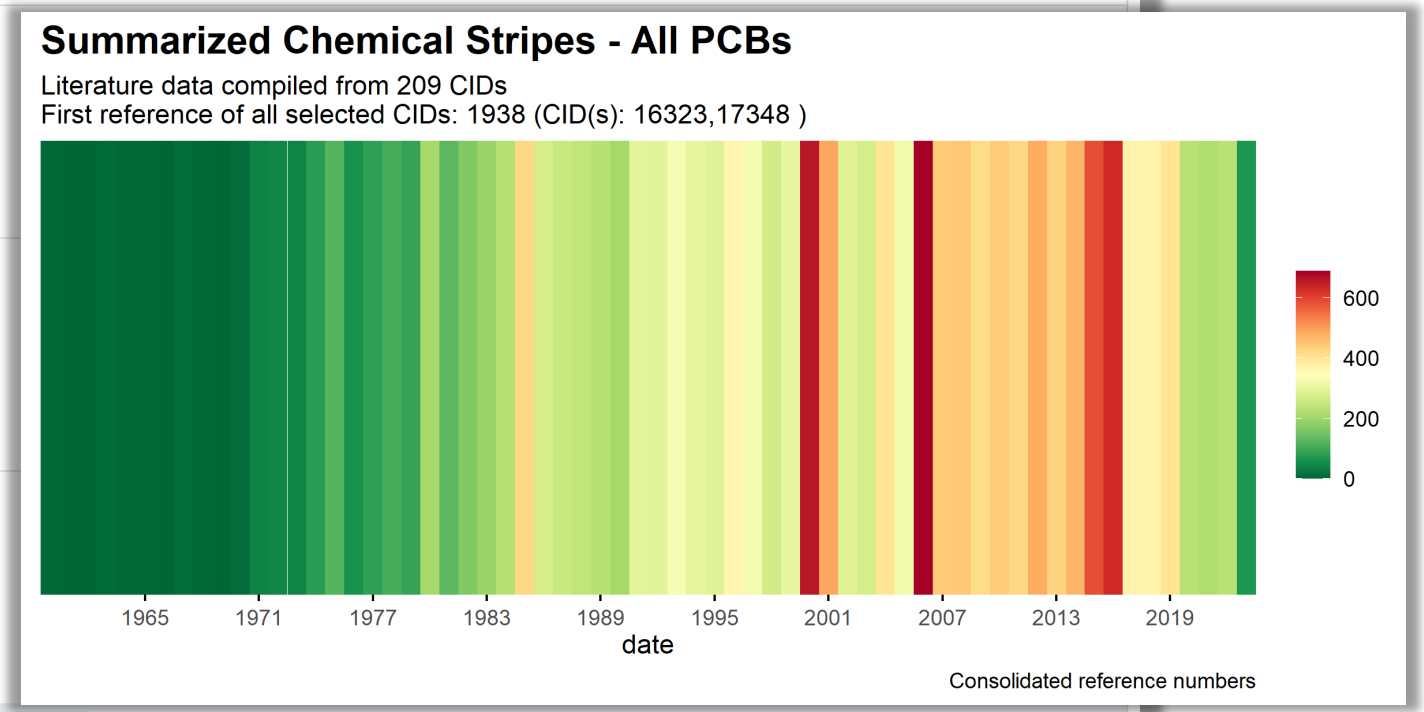
Download

Search

SORT BY UVCB Component - A to Z

UVCB Component	Type	Component Structure	Component CID	Reference	Source
2',3,4,4',5-Pentachlorobiphenyl	Component		47650		
2',3,4-Trichlorobiphenyl	Component		38036		
2,2',3,3',4,4',5,5',6-Nonachlorobiphenyl	Component		38411		

First Previous Page 1 of 21 Next Last



https://gitlab.lcsb.uni.lu/eci/chemicalstripes

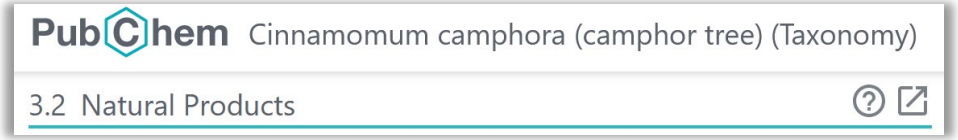
https://pubchem.ncbi.nlm.nih.gov/compound/Polychlorinated%20Biphenyls#section=Related-Records&fullscreen=true

Take home messages

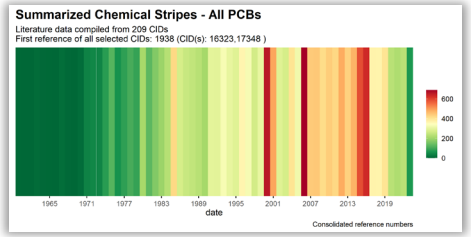
```
elsif ( $s =~ /charcol/i ) {
elsif ( $s =~ /chlorinated/i
elsif ( $s =~ /complex/i ) {
elsif ( $s =~ /compounds/i )
elsif ( $s =~ /compd\./i ) {
elsif ( $s =~ /compds\./i )
```

- Text-based curation to recognise UVCBs
- Algorithms can support cross-linking Variable cases

- Biological cases leverage other disciplines



- Adding in NORMAN Network UVCB data case by case



- We would be happy to work with HESI on the Tier 0 Database!
  - Emma: emma.schymanski@uni.lu or @ESchymanski
  - PubChem: pubchem-help@ncbi.nlm.nih.gov or @pubchem

U.S. National Library of Medicine
National Center for Biotechnology Information

Acknowledgements!

Today's slides:

DOI: 10.5281/zenodo.8302440

Email: emma.schymanski@uni.lu

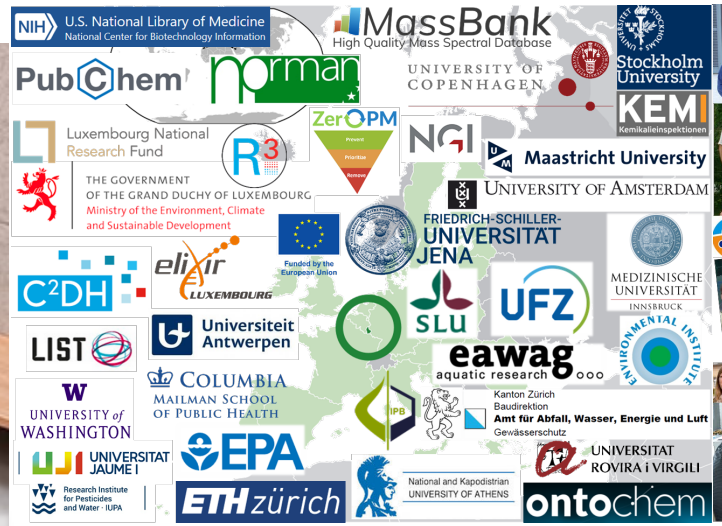
pubchem-help@ncbi.nlm.nih.gov

Twitter: @ESchymanski, @EvanBolton @pubchem

Today's slides zenodo

Evan Bolton, Jian (Jeff) Zhang, Paul Thiessen, Leonid Zaslavsky, Qingliang (Leon) Li, plus Tiejun, Asta, Ben, Siqian + the whole team

This work was supported in part by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM), National Institutes of Health (NIH)



Funded by the European Union

H2020: 101036756

Luxembourg National Research Fund