WEBINAR SERIES

Grand Challenges in Environmental Chemistry

WEDNESDAY MAY 17, 2023 | 10:00-11:00 BST / 11:00-12:00 CEST

SPEAKERS

Prof. Yafang Cheng
Max Planck Institute for Chemistry

Prof. Emma Schymanski
University of Luxembourg

HOSTED BY

Environmental Science & Technology

Environmental Science & Technology Letters

ACS Environmental Au
An open access journal of the American Chemical Society
Open Access

Luxembourg National Research Fund

Grand Challenges in Environmental Chemistry

UNIVERSITÉ DU LUXEMBOURG

Funded by the European Union

FAIR/Open Data and Non-target Screening Approaches for Millions of Chemicals

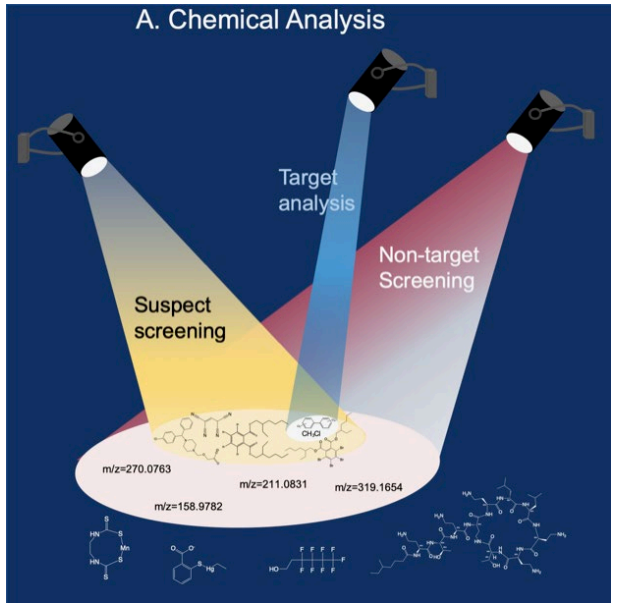


Image mod. from DOI: 10.1126/science.aay6636

Assoc. Prof. Dr. Emma L. Schymanski
(plus many, many colleagues and collaborators!)

Environmental Cheminformatics Group,
Luxembourg Centre for Systems Biomedicine, University of Luxembourg
Email: emma.schymanski@uni.lu and @ESchymanski
https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics/

ACS “Grand Challenges” Webinar, 17 May 2023
https://access.acspubs.org/EnvironmentalGrandChallengesWebinar

Slides available at DOI: 10.5281/zenodo.7944066

Today’s slides
zenodo

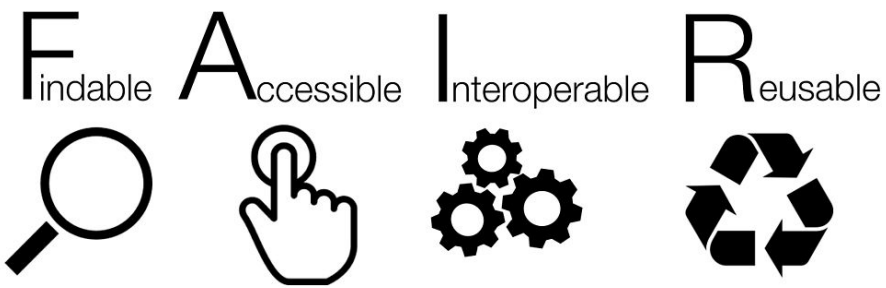
FAIR and Open Science

https://doi.org/10.1038/s44221-022-00014-z

Water science must be Open Science

Emma L. Schymanski & Stanislaus J. Schymanski

Since water is a common good, the outcome of water-related research should be accessible to everyone. Since Open Science is more than just open access research articles, journals must work with the research community to enable fully open and FAIR science



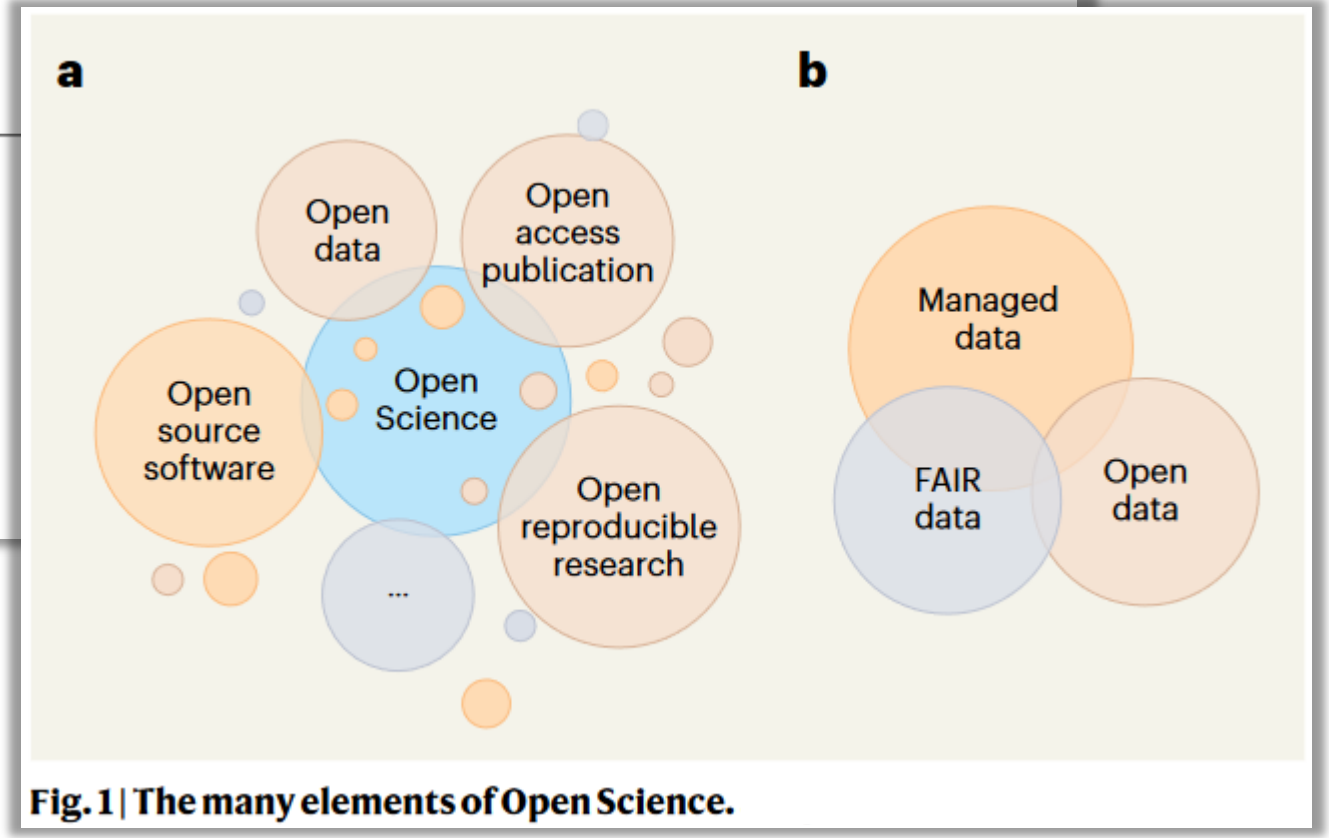


Fig. 1 | The many elements of Open Science.

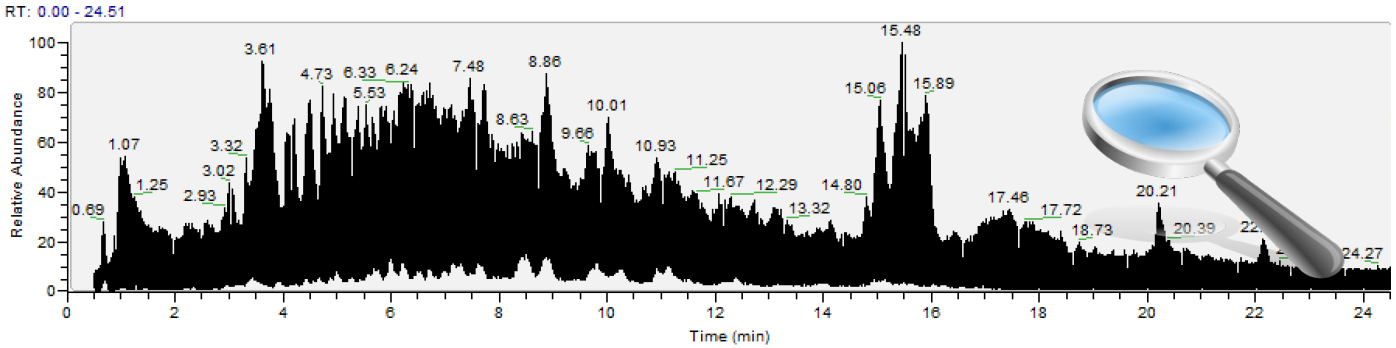
https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

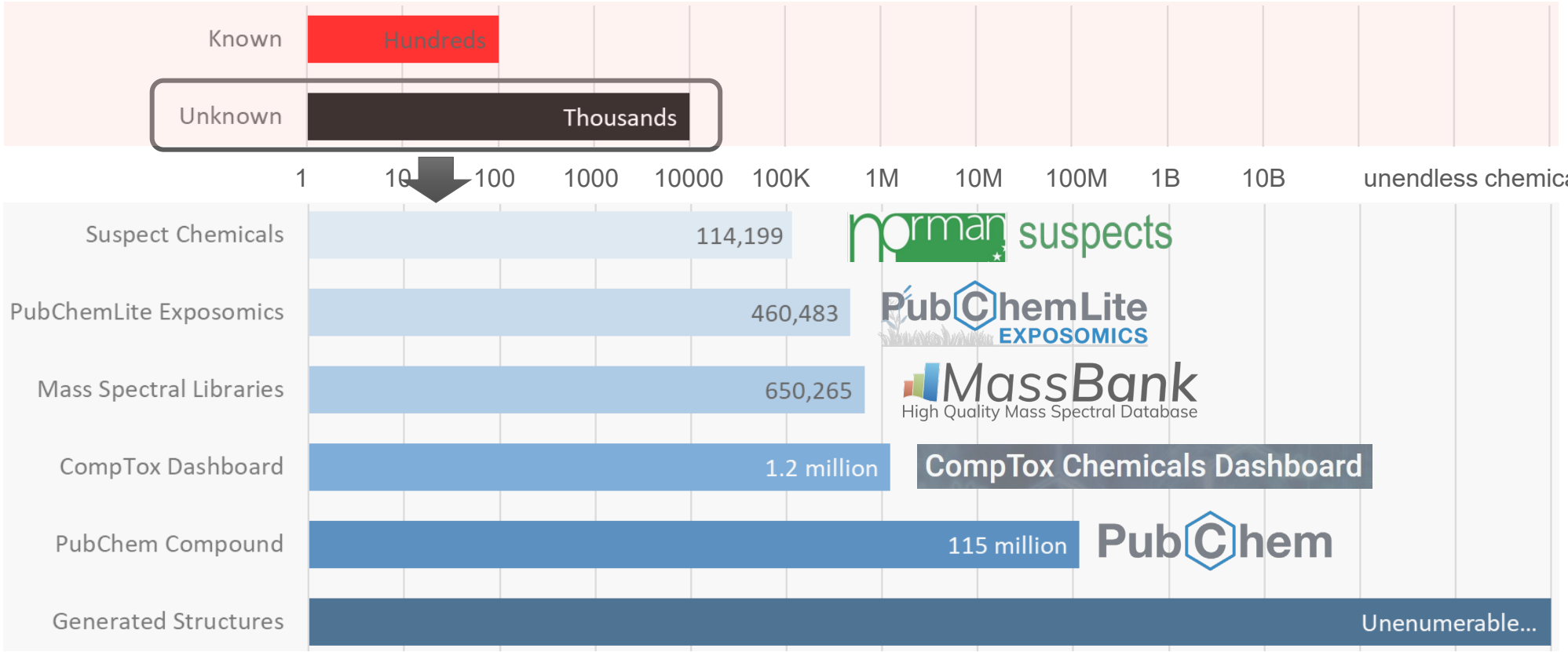
Schymanski & Schymanski (2023) Nature Water. DOI: 10.1038/s44221-022-00014-z

Environmental Cheminformatics & Non-target HR-MS

High resolution mass spectrometry

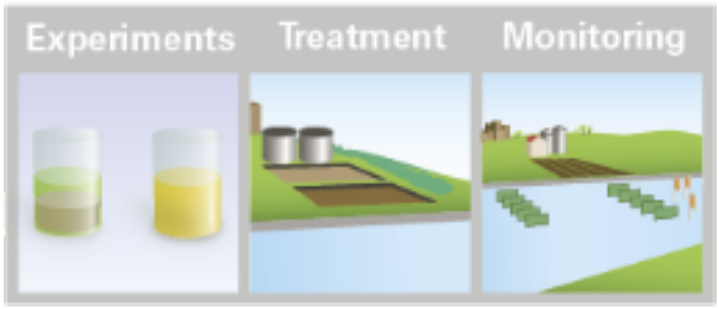
AND connecting chemical knowledge





Chromatogram: Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Resource numbers updated 16 May 2023.

Non-target High Resolution Mass Spectrometry (NT-HRMS)

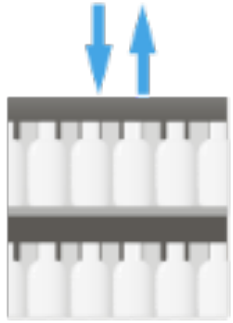


Commercial / Open Software

Databases

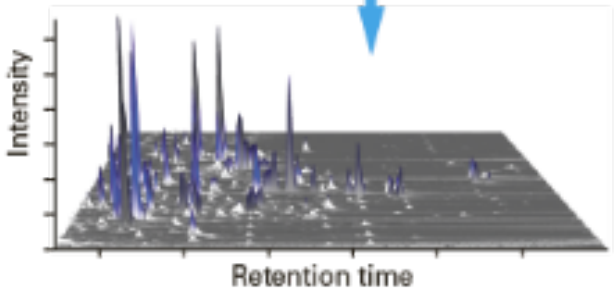
In Silico Prediction

Sampling → Analysis → Data Pre-Processing → Prioritization → Identification

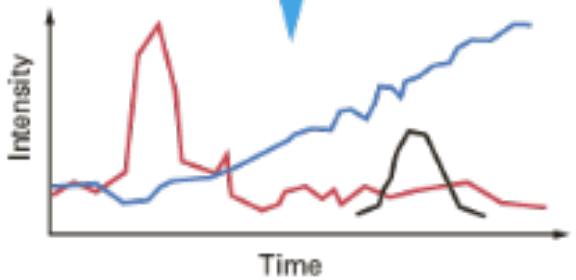


SAMPLE ARCHIVE

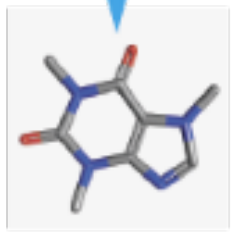
DIGITAL ARCHIVE



POLLUTION OVERVIEW



PATTERN



STRUCTURES

Modified from: Hollender, Schymanski *et al*. 2018, ES&T Feature, 51:20, 11505-11512. DOI: 10.1021/acs.est.7b02184

Open Source Workflows for NT-HRMS: patRoon

Software | Open Access | Published: 06 January 2021

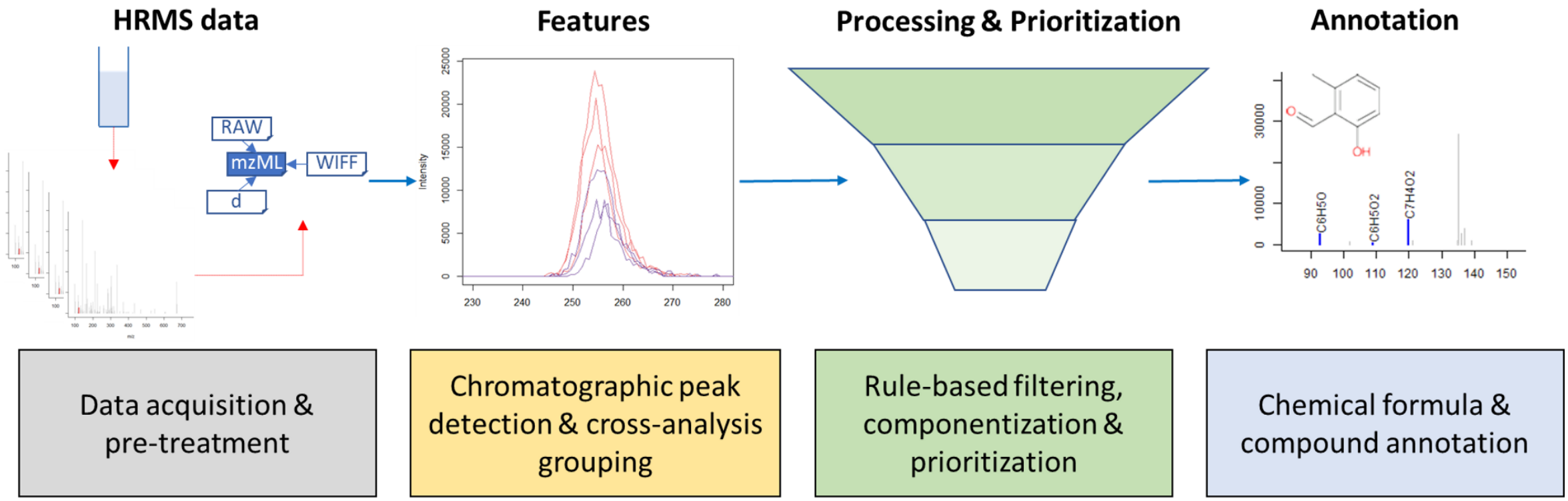
patRoon: open source software platform for environmental mass spectrometry based non-target screening

Journal of Cheminformatics **13**, Article number: 1 (2021) | Cite this article

Rick Helmus ✉, Thomas L. ter Laak, Annemarie P. van Wezel, Pim de Voogt & Emma L. Schymanski

patRoon 2.0: Improved non-target analysis workflows including automated transformation product screening

Rick Helmus[1¶], Bas van de Velde[123], Andrea M. Brunner[24], Thomas L. ter Laak[12], Annemarie P. van Wezel[1], and Emma L. Schymanski[5]



Helmus *et al.* (2022) patRoon 2.0 *J. Open Source Software*, DOI: 10.21105/joss.04029
Helmus *et al.* (2021) patRoon. *J. Cheminformatics*. DOI: 10.1186/s13321-020-00477-w

Mass Spectral Libraries: MassBank (Open Source & Data!)

https://massbank.eu/MassBank/ https://github.com/MassBank/MassBank-data/

MassBank Search Contents Download Accession Go More▼

MassBank Europe

MassBank
High Quality Mass Spectral Database

>> Search Spectra

Search for:

Basic Search | Peak List | Peaks

Compound Information

Compound name

Exact Mass

AND

Formula (e.g. C6H7N5, C5H*N5, C5*)

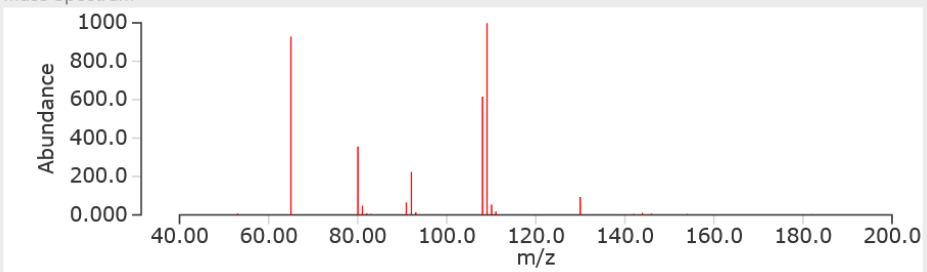
AND

Search

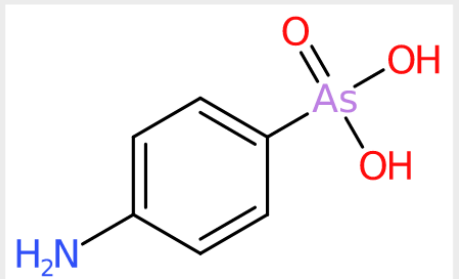
MassBank Record: LU040605

(4-Aminophenyl)arsonic acid; LC-ESI-QFT; MS2; CE: 75; R=17500; [M+H]+

Mass Spectrum



Chemical Structure



Gathering Expert Knowledge: NORMAN-SLE

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

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

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

Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmen
matrices

RESEARCH Open Access

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*]

Example: Retrospective Screening / Early Warning Systems

www.norman-data.eu

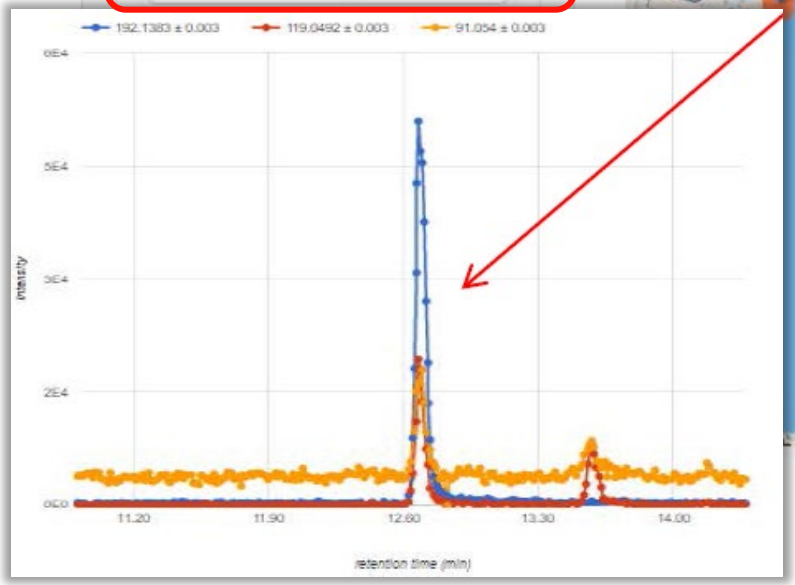
NORMAN Digital Sample Freezing Platform | Main Page | Batch mode

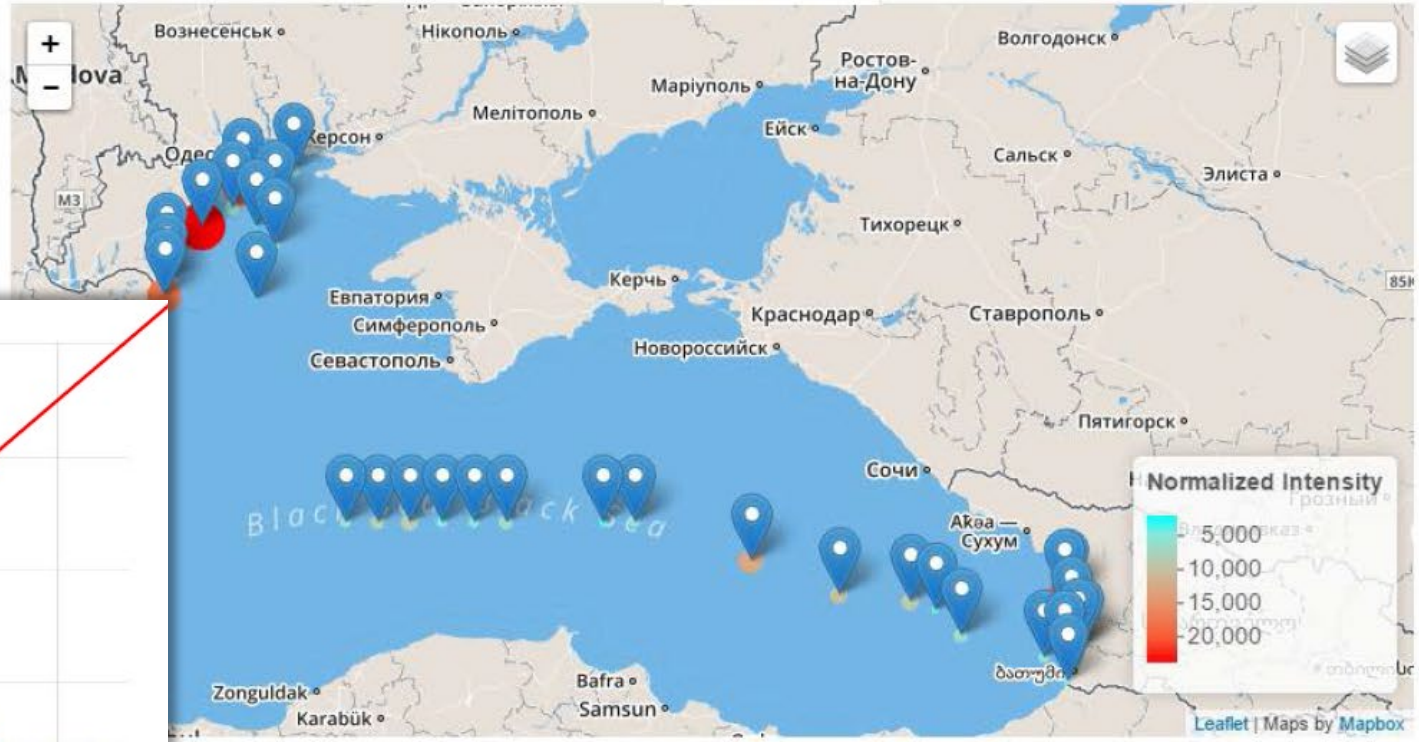
Choose Emerging Substance or input mass of interest and experimental RTI

Substance name or CAS or StdInChIKey

DEET [134-62-3] [MMOXZBCLCQITDF-UHFFFAOYSA-N]

Contributed Samples | Results | Chromatograms | Interactive Map | Help

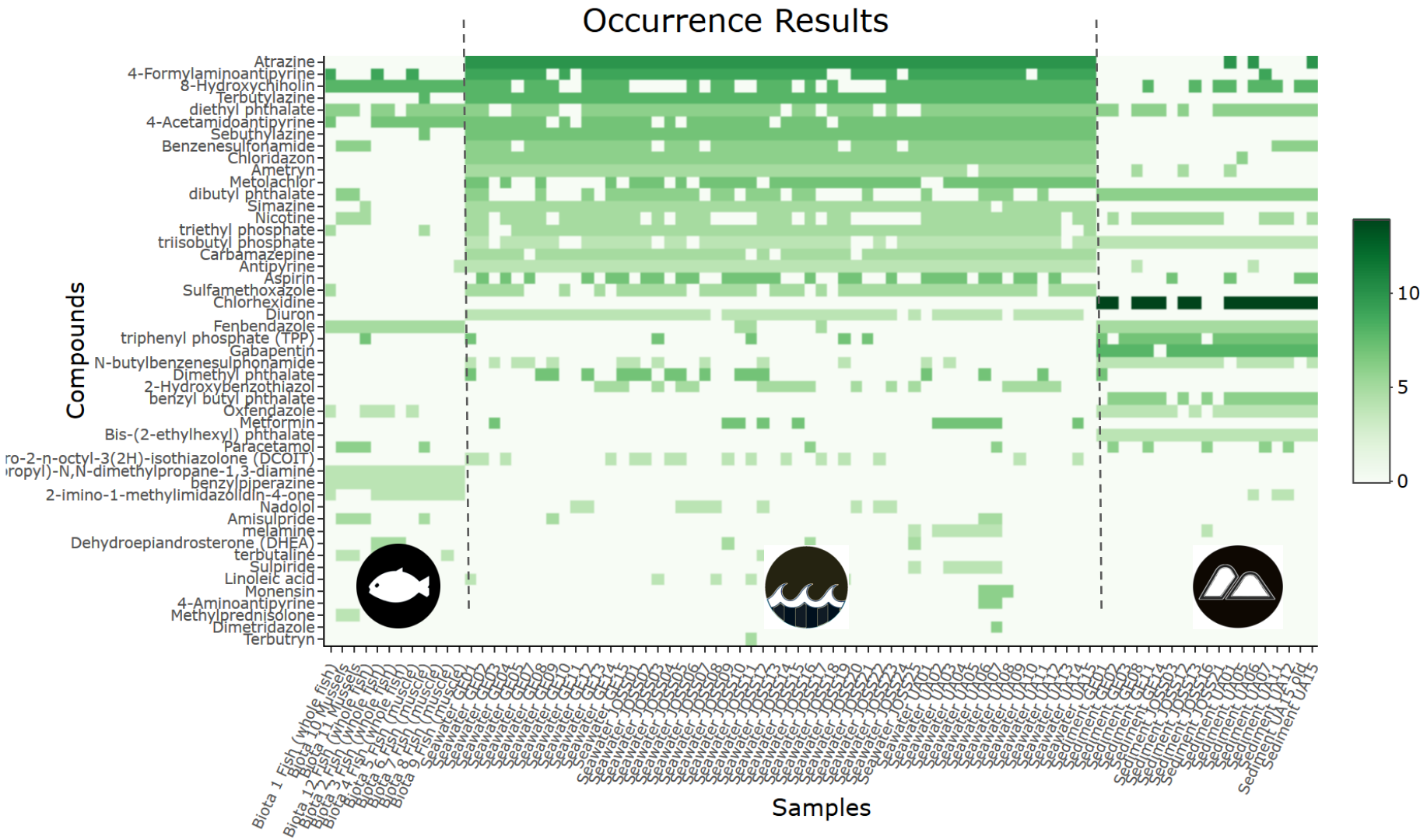




http://norman-data.eu/ & Alygizakis *et al*, 2018, ES&T, DOI: 10.1021/acs.est.8b00365 & Alygizakis *et al*, 2019, TrAC, DOI: 10.1016/j.trac.2019.04.008

Examples: Distribution of Chemicals in Various Matrices

Retrospective screening of REACH chemicals in Black Sea samples

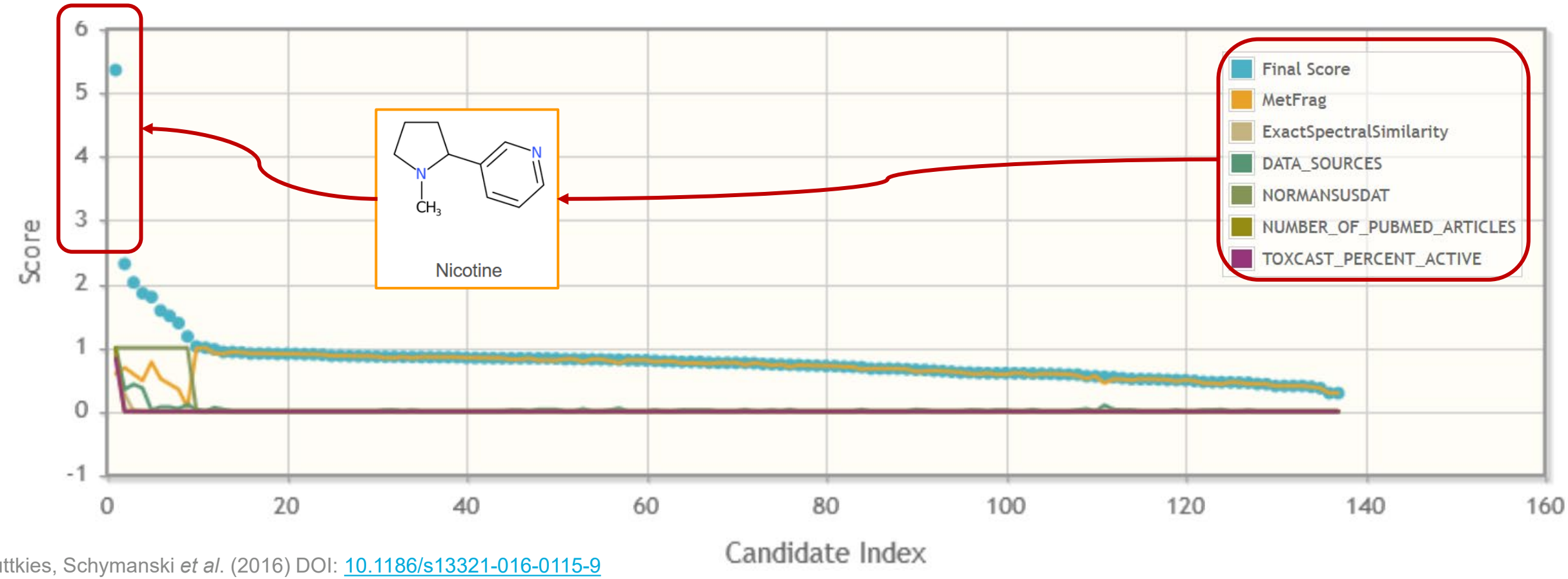


Alygizakis NA *et al* (2019) DOI: [10.1016/j.trac.2019.04.008](https://doi.org/10.1016/j.trac.2019.04.008)

Interactive heat map at http://norman-data.eu/NORMAN-REACH

Connecting Knowledge for Chemical Identification: MetFrag

https://msbi.ipb-halle.de/MetFrag/ http://ipb-halle.github.io/MetFrag/



Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9

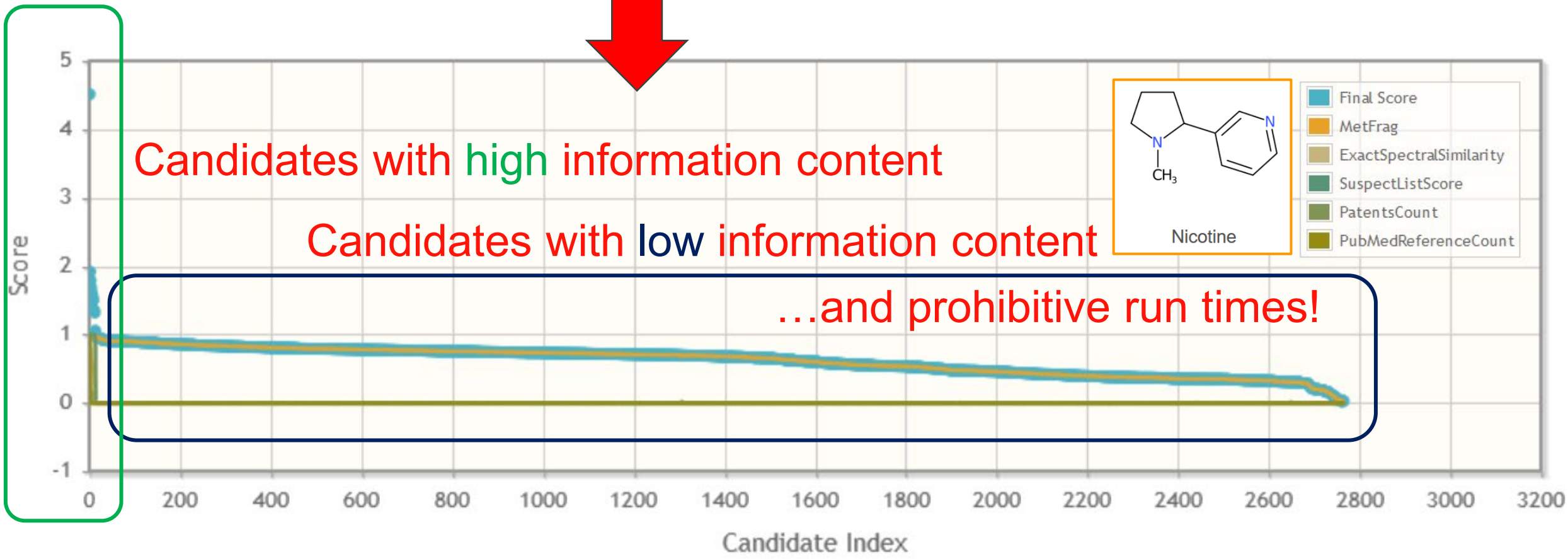
Problem: HR-MS “Chemical Space” is too big!

CAS® A DIVISION OF THE AMERICAN CHEMICAL SOCIETY — 204 million

PubChem — 115 million

ChemSpider Search and share chemistry — 123 million

CompTox Chemicals Dashboard — 1.2 million



Can we break down PubChem into useful bits?

- PubChem Compound TOC 66,469,244
  - Agrochemical Information 3,100
  - Associated Disorders and Diseases 30,075
  - Biologic Description 2,480,712
  - Biological Test Results 4,514,231
  - Chemical and Physical Properties 267,891
  - Classification 22,843,372
  - Drug and Medication Information 19,319
  - Food Additives and Ingredients 7,349
  - Identification 4,699
  - Information Sources 46,018,697
  - Interactions and Pathways 205,070
  - Literature 2,097,436
  - Names and Identifiers 5,292,821
  - Patents 38,442,426
  - Pharmacology and Biochemistry 113,472
  - Related Records 12,839,065
  - Safety and Hazards 169,744
  - Spectral Information 1,575,801
  - Structures 11,402,010
  - Toxicity 116,655
  - Use and Manufacturing 105,682

- Agrochemical Information 3,100
  - Agrochemical Category 1,956
  - Agrochemical Transformations 1,480
  - EU Pesticides Data 1,237
  - USDA Pesticide Data Program 670
- Associated Disorders and Diseases 30,075
  - Disease and References 13,386
- Interactions and Pathways 205,070
  - Pathways 62,538
- Use and Manufacturing 105,682
  - Uses 71,558
    - Consumer Uses 3,807
    - Household Products 3,898
    - Industry Uses 5,975
    - Use Classification 59,060
  - Consumption Patterns 1,124
  - Formulations/Preparations 4,510
  - General Manufacturing Information 50,863

Furathiocarb (Compound)

7.3 EU Pesticides Data

Active Substance	furathiocarb
Status	Not approved [Reg. (EC) No 1107/2009]
Legislation	2002/2076
ADI	0.0035 mg/kg bw/day [RMS 1999]
ARfD	0.006 mg/kg bw [RMS 1999]

EU Pesticides Database

Perfluorooctanoic acid (Compound)

8.3 U.S. Production

Production volumes for non-confidential chemicals reported under the Inventory Update Rule.

Year	Production Range (pounds)
1986	10 thousand - 500 thousand
1990	No Reports
1994	10 thousand - 500 thousand
1998	10 thousand - 500 thousand
2002	10 thousand - 500 thousand

US EPA; Non-confidential Production Volume Information Submitted by Companies for Chemicals Under the 1986-2002 Inventory Update Rule (IUR). Octanoic acid, pentadecafluoro- (335-67-1). Available from, as of November 2, 2010: https://www.epa.gov/oppt/iur/tools/data/2002-vol.html

Hazardous Substances Data Bank (HSDB)

Schymanski *et al.* (2021) DOI: 10.1186/s13321-021-00489-0

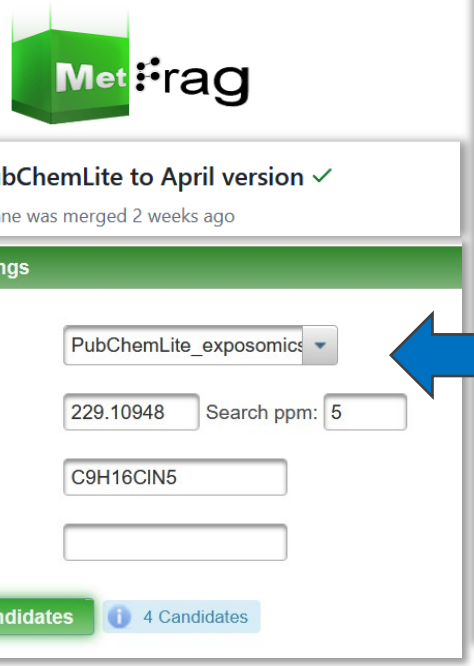
U.S. National Library of Medicine
National Center for Biotechnology Information

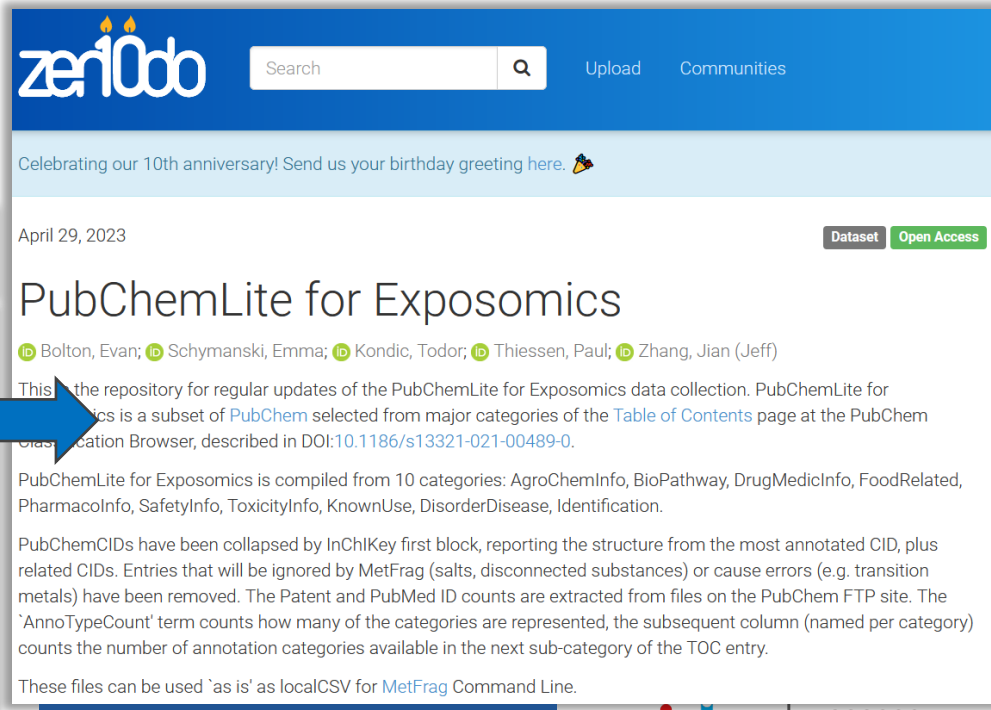
Introducing …

- PubChem Compound TOC ? 66,469,244
  - Agrochemical Information ? 3,100
  - Associated Disorders and Diseases ? 30,075
  - Biologic Description ? 2,480,712
  - Biological Test Results ? 4,514,231
  - Chemical and Physical Properties ? 267,891
  - Classification ? 22,843,372
  - Drug and Medication Information ? 19,319
  - Food Additives and Ingredients ? 7,349
  - Identification ? 4,699
  - Information Sources ? 46,018,697
  - Interactions and Pathways ? 205,070
  - Literature ? 2,097,436
  - Names and Identifiers ? 5,292,821
  - Patents ? 38,442,426
  - Pharmacology and Biochemistry ? 113,472
  - Related Records ? 12,839,065
  - Safety and Hazards ? 169,744
  - Spectral Information ? 1,575,801
  - Structures ? 11,402,010
  - Toxicity ? 116,655
  - Use and Manufacturing ? 105,682

PubChemLite EXPOSOMICS

~400,000 entries “small”





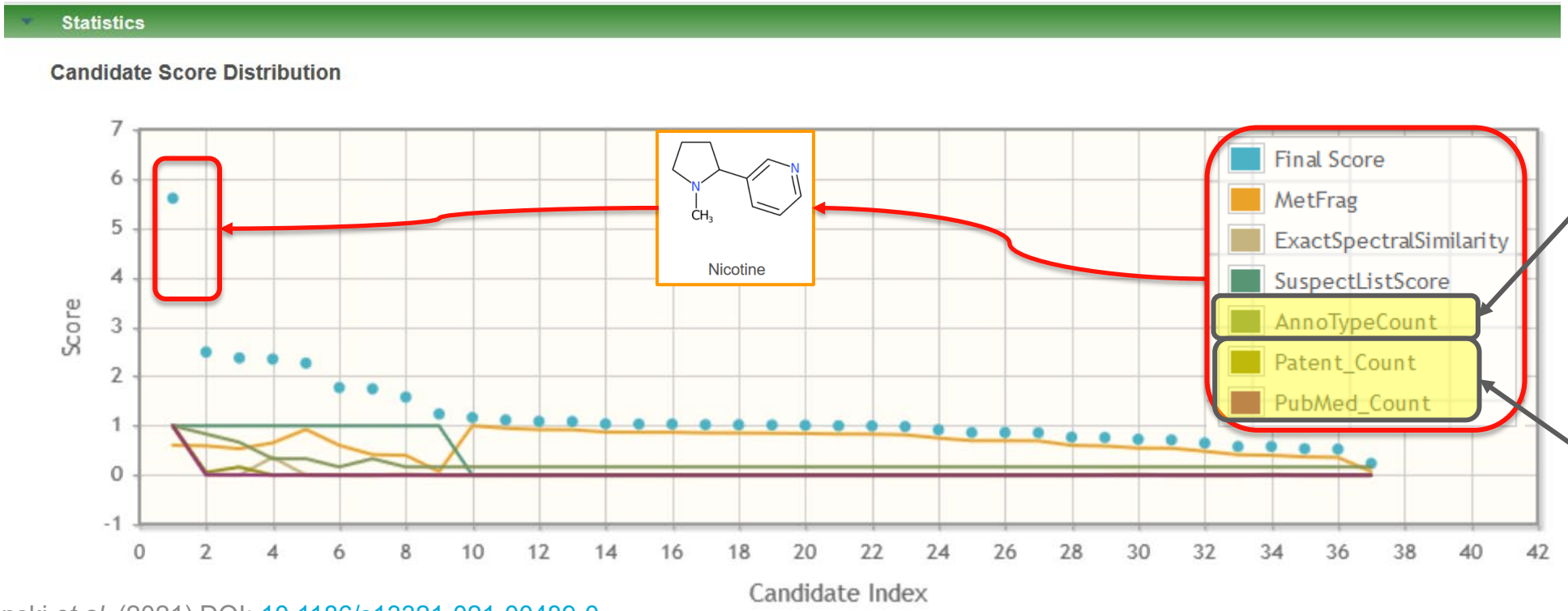
Schymanski *et al.* (2021) DOI: 10.1186/s13321-021-00489-0

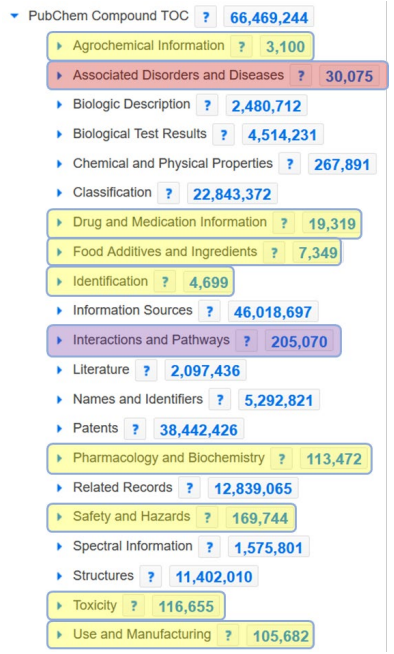
U.S. National Library of Medicine
National Center for Biotechnology Information

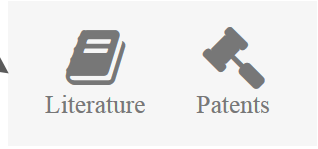
PubChemLite: Fewer and more relevant candidates – with context

MetFrag · MassBank High Quality Mass Spectral Database · norman suspects · PubChemLite EXPOSOMICS

Orders of magnitude faster, ~97% of test set in Top 3
Easier data interpretation => high throughput NTS



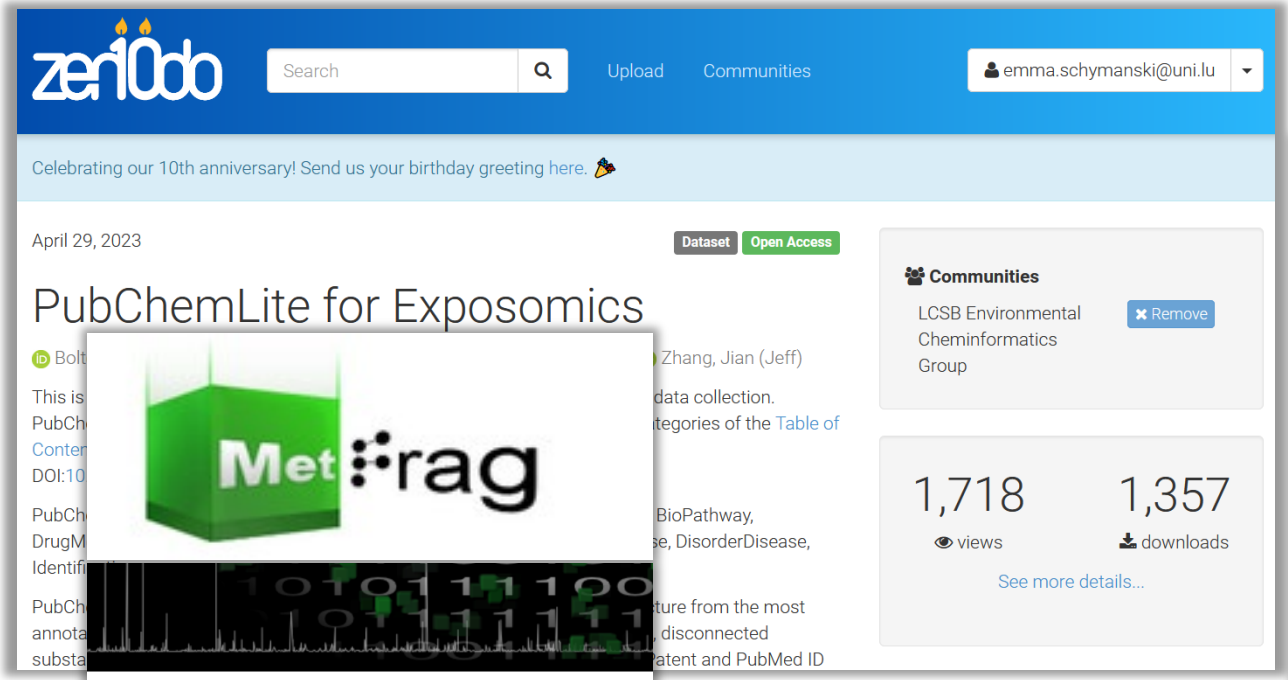




Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

PubChemLite is a Dynamic Open & FAIR Collection

- Built & evaluated every week, updated online every month
- Downloadable & fully integrated in patRoon & MetFrag



DOI: 10.5281/zenodo.5995885

https://ipb-halle.github.io/MetFrag/

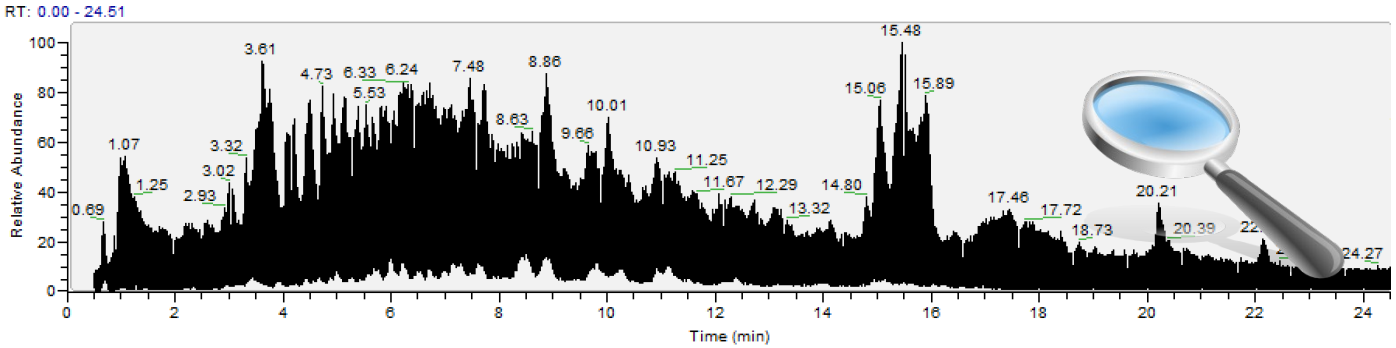
https://rickhelmus.github.io/patRoon/

Schymanski *et al.* (2021) DOI: 10.1186/s13321-021-00489-0 & Helmus *et al.* (2021) DOI: 10.1186/s13321-020-00477-w

Environmental Cheminformatics, NT HR-MS & *Transformations*

High resolution mass spectrometry

AND connecting chemical knowledge



Samples

Known — Hundreds

Unknown — Thousands

1 — 10 — 100 — 1000 — 10000 — 100K — 1M — 10M — 100M — 1B — 10B — unendless chemicals …. …. ….

Suspect Chemicals — 114,199 — norman suspects

PubChemLite Exposomics — 460,483 — PubChemLite EXPOSOMICS

Mass Spectral Libraries — 650,265 — MassBank High Quality Mass Spectral Database

CompTox Dashboard — 1.2 million — CompTox Chemicals Dashboard

PubChem Compound — 115 million — PubChem

Generated Structures — Unenumerable…

human — 2.4 M

human/env — 11 M

env — 1.2 M

humenv2 — 36 M

humenvhum — 98 M

hum2env — 86 M

Chemical space

Chromatogram: Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Resource numbers updated 16 May 2023.

Transforming PubChemLite with BioTransformer 3.0

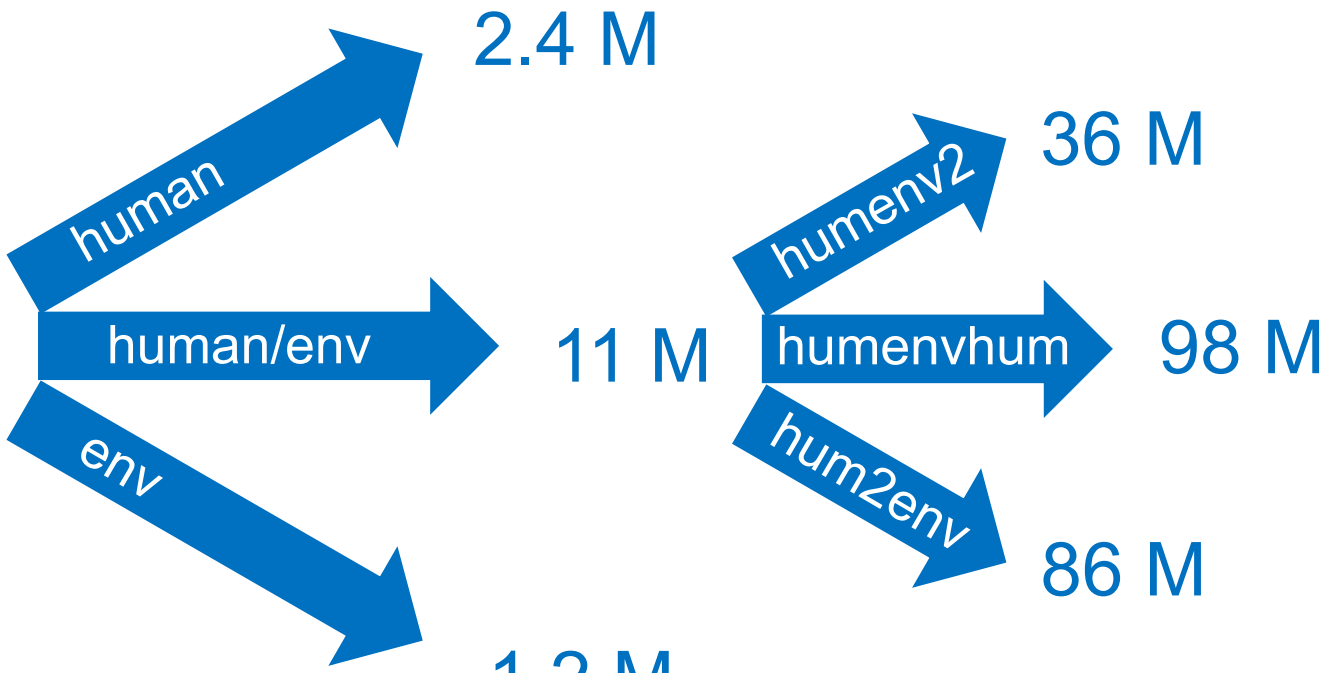
http://biotransformer.ca/

Wishart Lab
UNIVERSITY OF ALBERTA

PubChemLite
EXPOSOMICS

379,199 entries

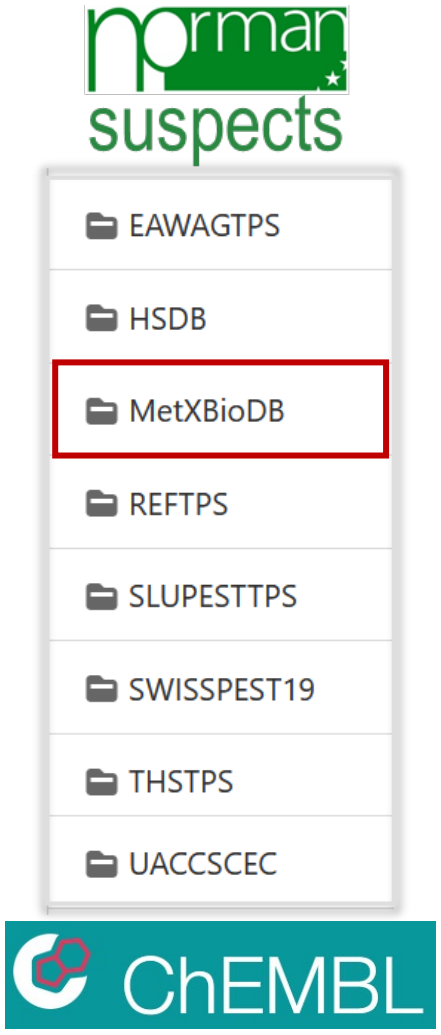
version 1.0.0; DOI: 10.5281/zenodo.5995886



Combinatorial explosion – more data needed to predict reaction likelihoods

Djoumbou Feunang *et al* (2019). BioTransformer, JCheminf. DOI: 10.1186/s13321-018-0324-5.
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Bolton *et al* (2021) *preliminary calculations;*
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

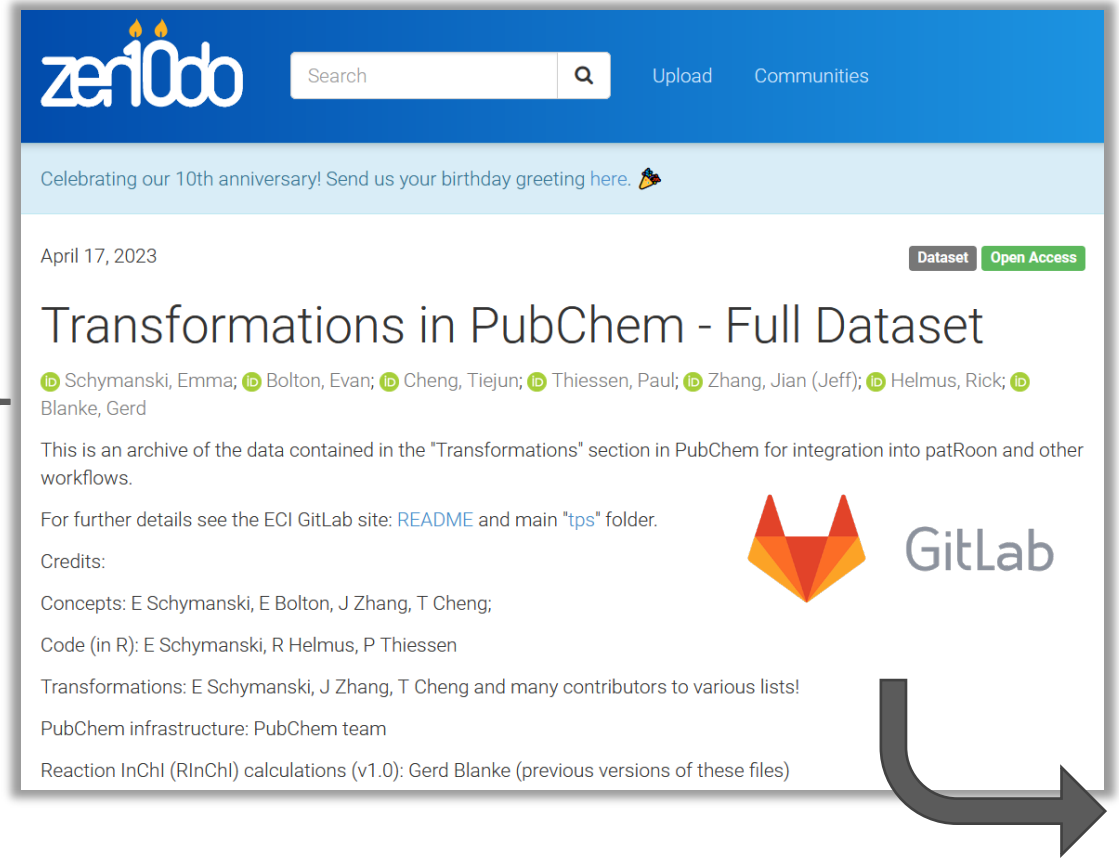
FAIR Transformations in PubChem and NORMAN-SLE



PubChem Transformations

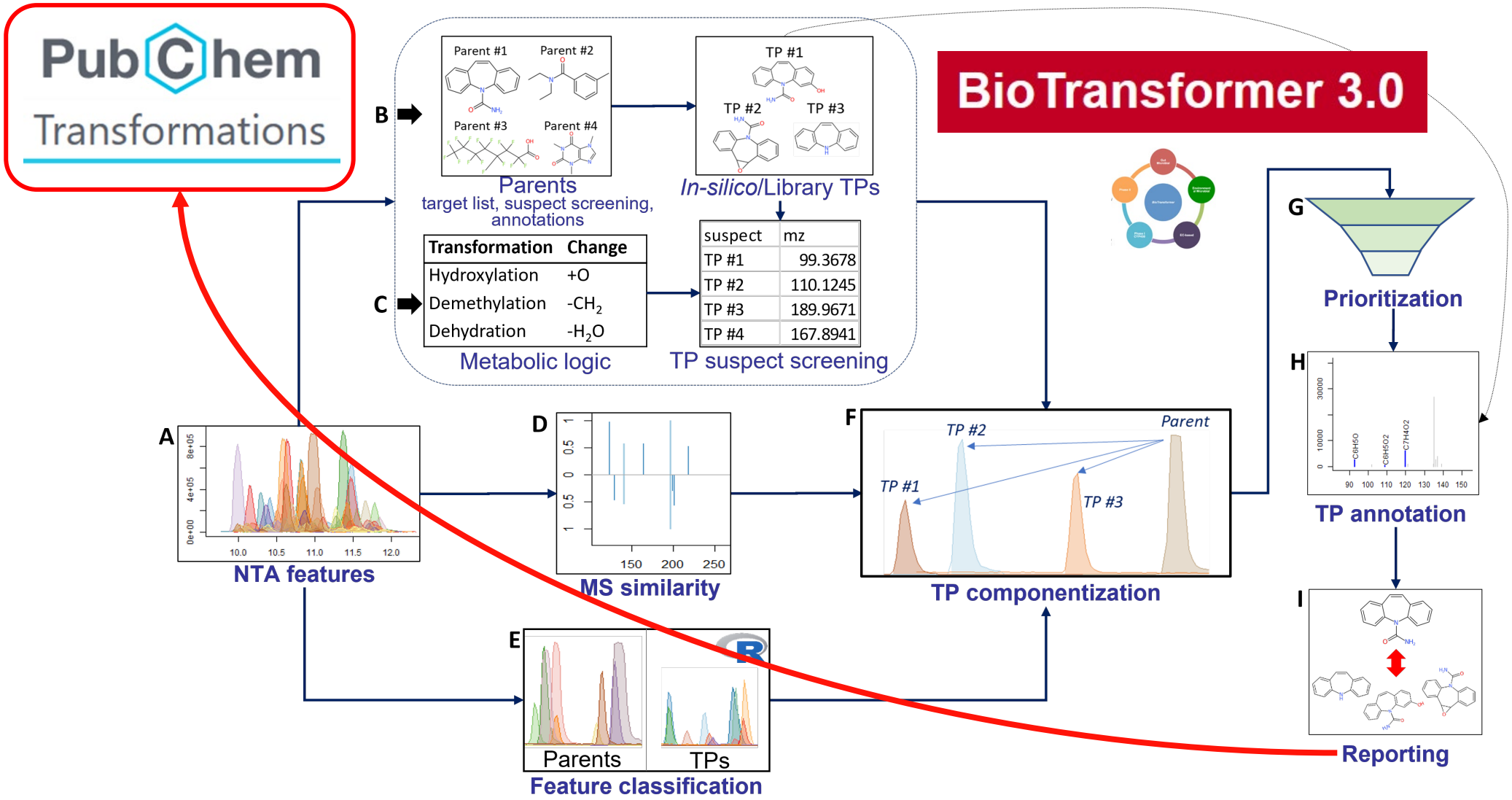
6,312 chemicals

>double the BioTransformer library



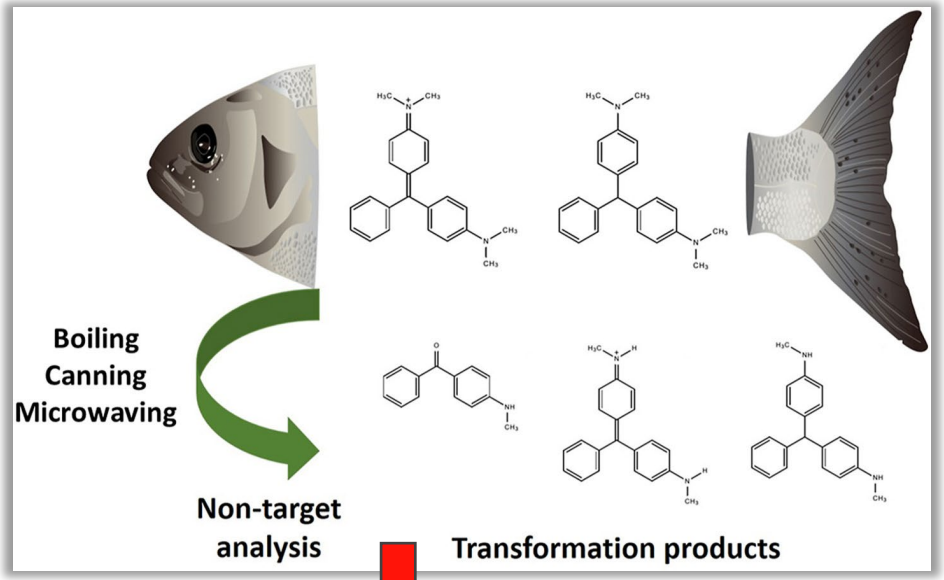
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Helmus *et al* (2022) DOI: 10.21105/joss.04029
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

Open Transformation Products Workflows in patRoon 2.0



Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Helmus *et al* (2022) DOI: 10.21105/joss.04029
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

What about *new* information? Help add it in!



zenodo

Search

Upload Communities

March 8, 2023

Dataset Open Access

S74 | REFTPS | Transformation Products and Reactions from Literature

Schymanski, Emma; Baesu, Anca; Chirsir, Parviel

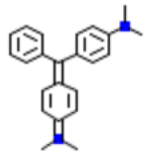
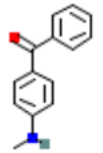
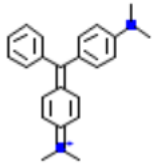
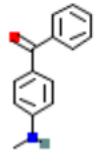
This is the collection associated with list S74 REFTPS Transformation Products and Reactions from Literature on the NORMAN Suspect List Exchange.

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

(4-(Methylamino)phenyl)(phenyl)methanone (Compound)

7.1 Transformations

2 items

Predecessor	Predecessor Name	Successor	Successor Name	Transformation
	Malachite green		4-(Methylamino)benzophenone	Thermally induced deconjugation and demethylation
	Malachite green cation		4-(Methylamino)benzophenone	Thermally induced deconjugation and demethylation

NORMAN Suspect List Exchange

Source	NORMAN Suspect List Exchange
Record Name	4-(Methylamino)benzophenone

Baesu *et al*. (2021) DOI: 10.1016/j.crfs.2021.09.010. S74 | REFTPS (2021) DOI: 10.5281/zenodo.4318838

FAIRifying PFAS

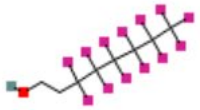
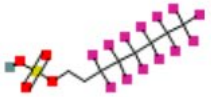
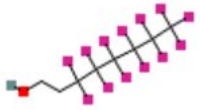
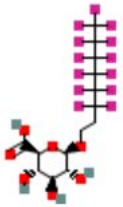
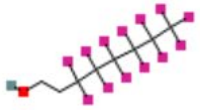
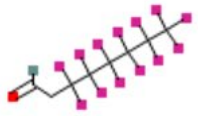
3,3,4,4,5,5,6,6,7,7,8,8,8-Tridecafluorooctan-1-ol (Compound)

7.1 Transformations

29 items

Search

SORT BY Predecessor Na

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	Evidenc
	6:2 FTOH		6:2 FTOH-Sulfate	Conjugation		10.1016
	6:2 FTOH		6:2 FTOH-Gluc	Conjugation		10.1016/j.cbi.2005.06.007
	6:2 FTOH		6:2 FTAL	Oxidation	24296	10.1016/j.cbi.2005.06.007

https://pubchem.ncbi.nlm.nih.gov/compound/69537#section=Pharmacology-and-Biochemistry

ELSEVIER

Chemico-Biological Interactions

Volume 155, Issue 3, 15 August 2005, Pages 165-180

Metabolic products and pathways of fluorotelomer alcohols in isolated rat hepatocytes

https://doi.org/10.1016/j.cbi.2005.06.007

Jonathan W. Martin [a], Scott A. Mabury [b], Peter J. O'Brien [a]

Show more

+ Add to Mendel

https://doi.org/10.10

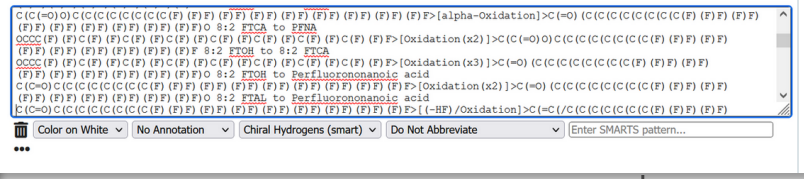
Jon Martin
@AcademicTox

Wow, this was a surprise. My postdoctoral work on metabolism of #PFAS has been FAIRified on @pubchem thanks to @ESchymanski and colleagues. Very exciting to think about how these huge efforts will improve future science for everyone.

@PChirsir

Emma Schymanski @ESchymanski · Jan 8

Spent a rainy afternoon FAIRifying #PFAS data from past @AcademicTox work, thanks to @PChirsir efforts we've committed the first set to @pubchem incl. new structures - hope to bring the 2005 PDF images to #OpenScience life soon! Again @the_cdk Depict makes it so much fun @jwmay !



https://twitter.com/AcademicTox/status/1479841982679765001

The PFAS Challenge: Updated OECD Definition in 2021

OECD

Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2021)25

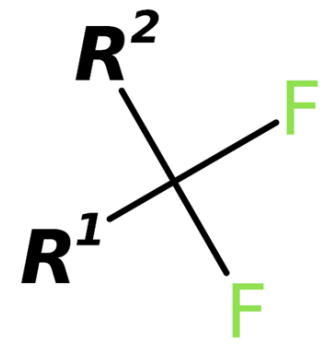
Unclassified — **English - Or. English**

9 July 2021

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Reconciling Terminology of the Universe of Per- and Polyfluoroalkyl Substances: Recommendations and Practical Guidance

Series on Risk Management
No.61

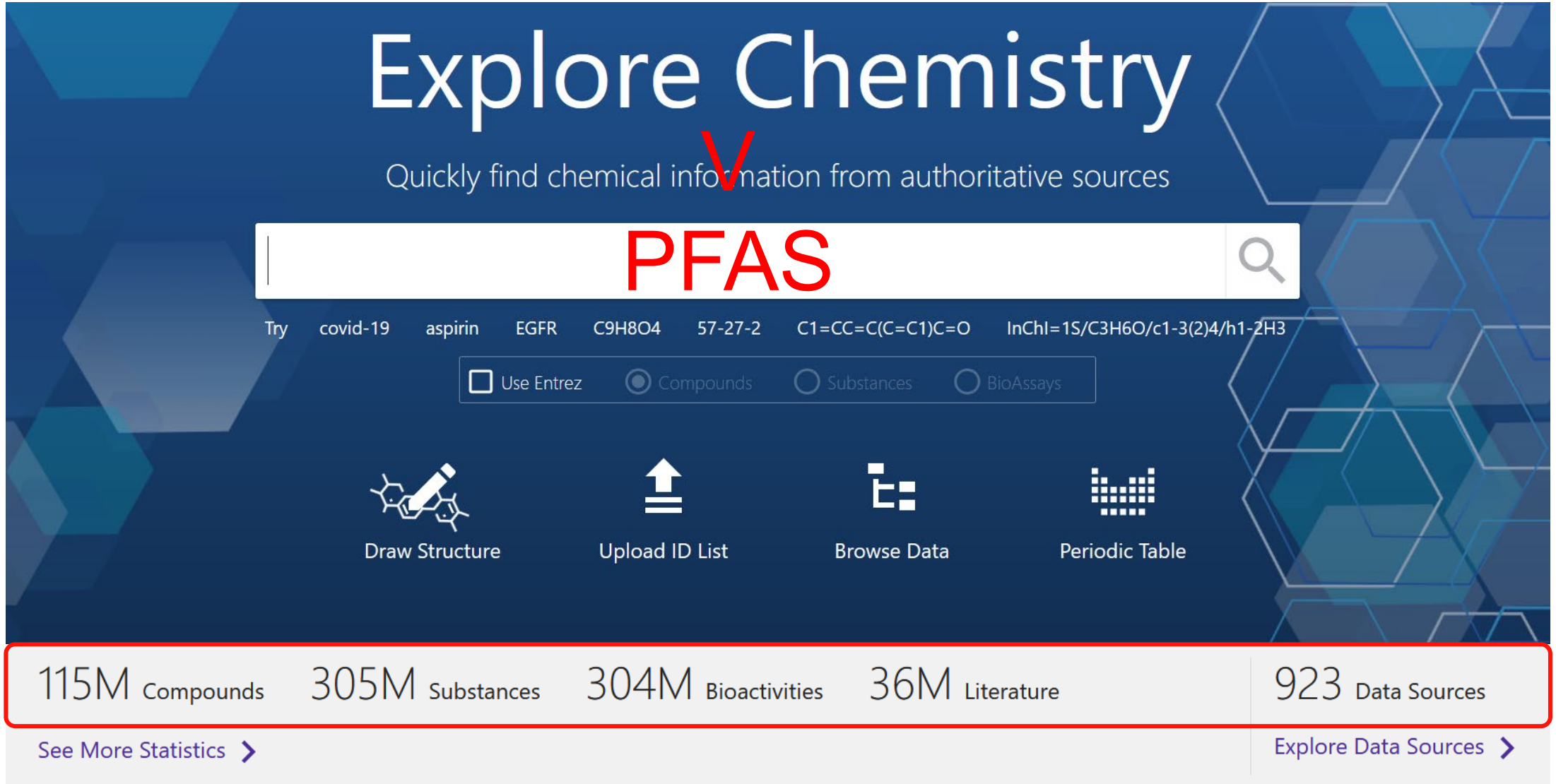


"OECD PFAS"

OECD Monograph ENV/CBC/MONO(2021)25 (9 July 2021)

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

The PFAS Challenge: How to scale this to PubChem ?



https://pubchem.ncbi.nlm.nih.gov/

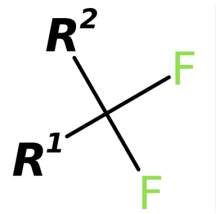
The PubChem PFAS Tree

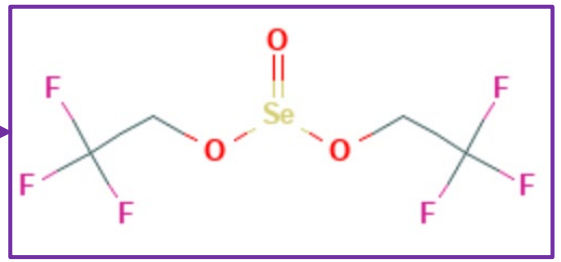
- PFAS and Fluorinated Compounds in PubChem — 21,250,352
  - OECD PFAS definition — 6,483,761
    - Molecule contains isolated CF2 — 663,762
    - Molecule contains isolated CF3 — 5,701,896
    - Molecule contains PFAS parts larger than CF2/CF3 — 228,734
  - Organofluorine compounds — 20,282,468
    - Fluorinated aliphatic substances — 898,136
    - Fluorinated aromatic substances — 19,311,523
    - Other fluorinated substances — 97,387
  - Other diverse fluorinated compounds — 122,343
  - PFAS and fluorinated compound collections — 1,789,324
  - PFAS breakdowns by chemistry — 7,415,092
  - Regulatory PFAS collections — 26,870

"OECD PFAS" (+salts/mixtures)

- other fluorinated *ALIPHATIC* substances that do not have a fully fluorinated methyl or methylene carbon atom
- fluorinated ***ALIPHATIC*** substances that have a fully fluorinated methyl or methylene carbon atom → Per- and Polyfluoroalkyl Substances (PFASs) e.g. R- C_nF_{2n+1} or R- C_nF_{2n}- R', n ≥1 (for examples, see Figure 5 above)
- non-fluorinated ***AROMATIC*** ring(s) with fluorinated ***ALIPHATIC*** side-chain(s) that have a fully fluorinated methyl or methylene carbon atom
- fluorinated ***AROMATIC*** ring(s) with fluorinated ***ALIPHATIC*** side chain(s) that have a fully fluorinated methyl or methylene carbon atom
- (non-)fluorinated *AROMATIC* ring(s) with fluorinated *ALIPHATIC* side chain(s) that do NOT have a fully fluorinated methyl or methylene carbon atom
- fluorinated *AROMATIC* ring(s) with non-fluorinated *ALIPHATIC* side chain(s)
- fluorinated *AROMATIC* substances without a side chain

"OECD PFAS"



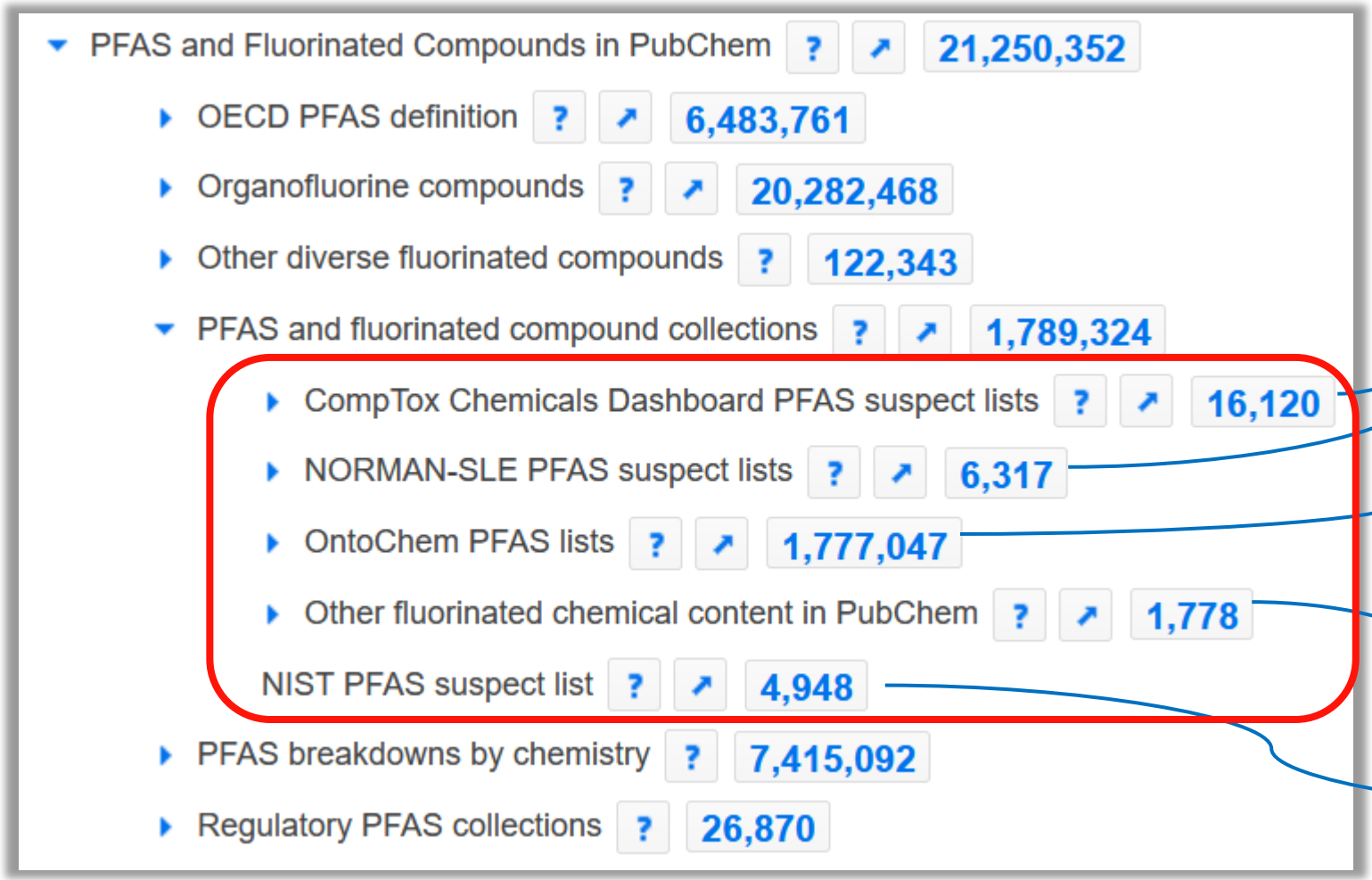


OECD Monograph ENV/CBC/MONO(2021)25 (9 July 2021)

https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120

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

The PubChem PFAS Tree – Collection of Suspect Lists



EPA CompTox Chemicals Dashboard

norman suspects

Google ontochem

PubChem

NIST PUBLIC DATA REPOSITORY

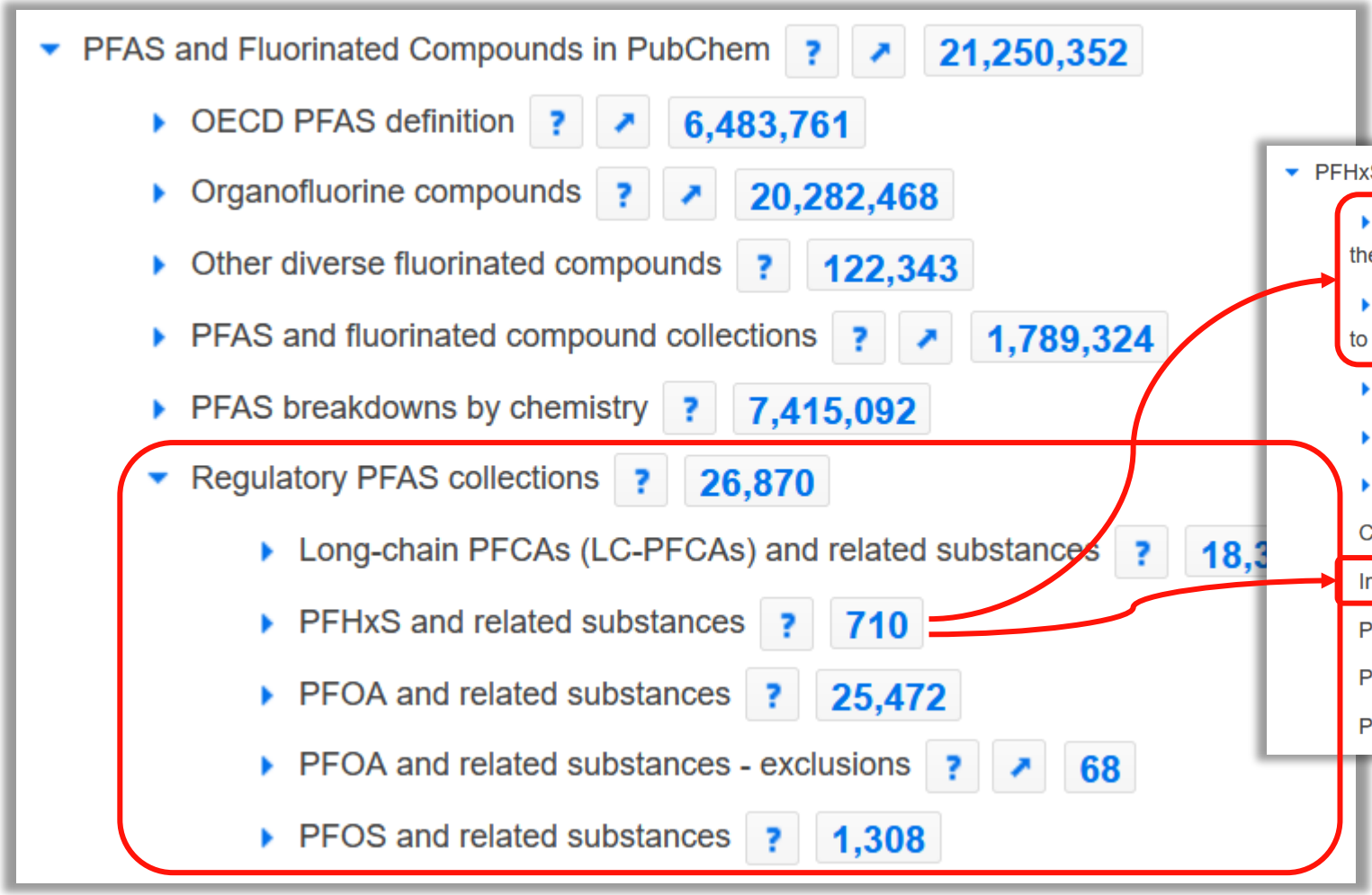
https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120

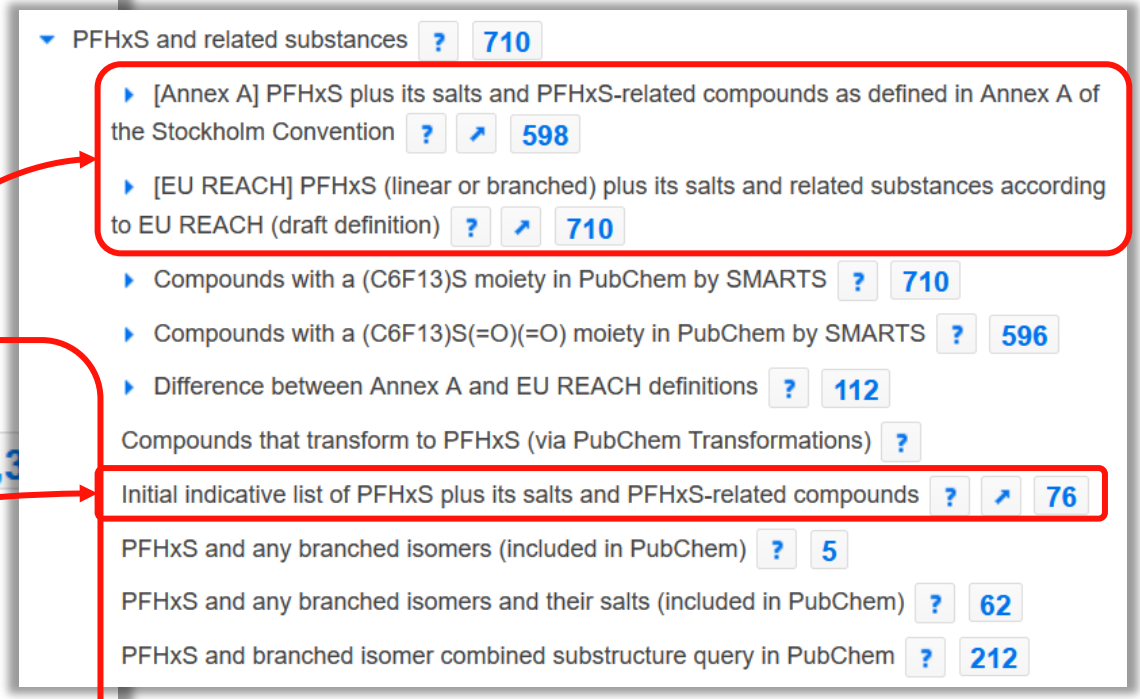
POPRC Meets PubChem PFAS Tree …

- Adding regulatory collections to the PFAS Tree

Today's slides zenodo

POPRC slides available at
DOI: 10.5281/zenodo.7118551



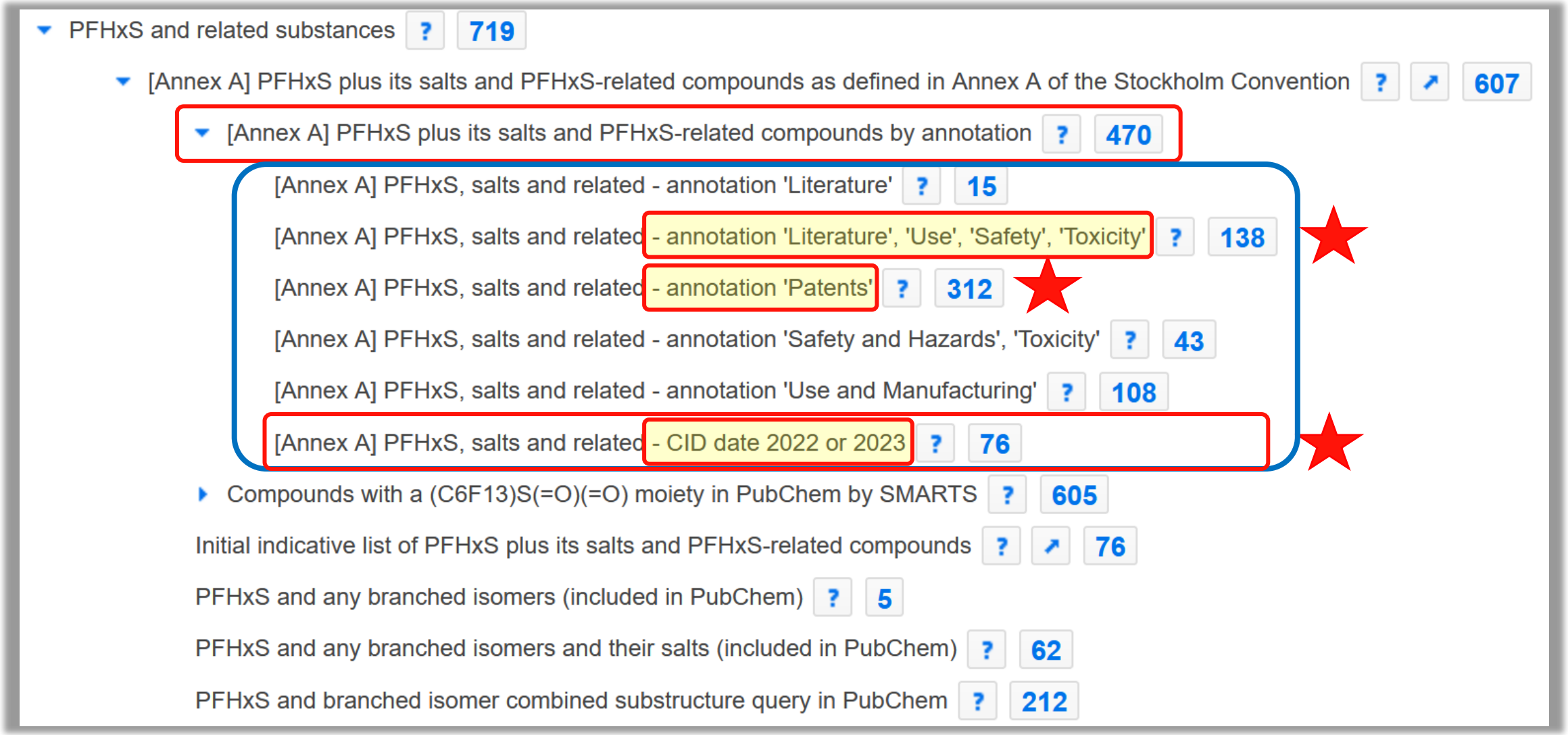


PubChem PFAS Tree: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120

Thanks to Andreas Buser, FOEN, CH

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

Example of PFHxS – Which PFAS are relevant? Which are new?



Who does all this PFAS data come from?

https://tarheels.live/bakerlab/

An overview of the uses of per- and polyfluoroalkyl substances (PFAS)†

Juliane Glüge, *[a] Martin Scheringer, [a] Ian T. Cousins, [b] Jamie C. DeWitt, [c] Gretta Goldenman, [d] Dorte Herzke, [ef] Rainer Lohmann, [g] Carla A. Ng, [h] Xenia Trier [i] and Zhanyun Wang [j]

Glüge *et al.* (2020) ESPI, DOI: 10.1039/d0em00291g

PubChem 13C3-PFHxS (Compound)

3.2.1 Collision Cross Section

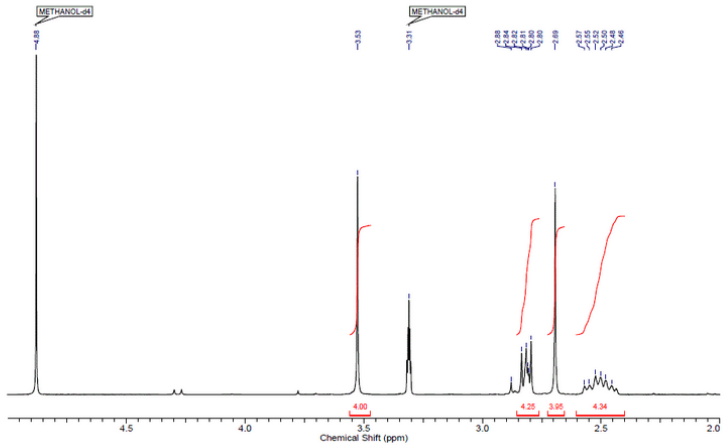
150.51 Å^2 [M-H]$^-$ [CCS Type: DT; Buffer gas: N2; Dataset: PFAS]

DOI:10.1021/acs.est.2c00201

Baker Lab, Chemistry Department, The University of North Carolina at Chapel Hill

PubChem 6:2 FTMAP (Compound)

Reference for Source	B Bugsel, R Bauer, F Herrmann, ME Maier, C Zwiener (2022) Analytical and Bioanalytical Chemistry, 414, 1217-1225 doi:10.1007/s00216-021-03463-9



Reference for Dataset	S74 | REFTPS | Transformation Products and Reactions from Literature doi:10.5281/zenodo.4318838

NORMAN Suspect List Exchange

…and many more!

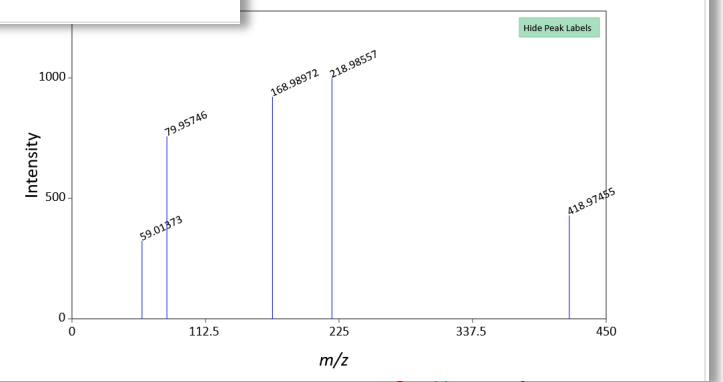
Bugsel *et al.* (2022) ABC, 414, 1217-1225. DOI:10.1007/s00216-021-03463-9

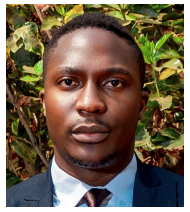
https://pubchem.ncbi.nlm.nih.gov/compound/156620404#section=1H-NMR-Spectra

PubChem Perfluorononanoic acid (Compound)

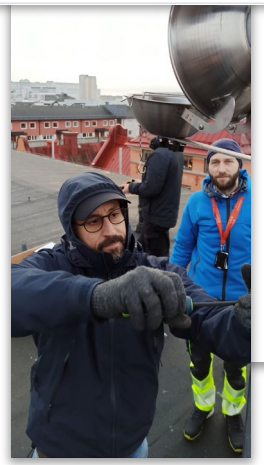
Accession ID	MSBNK-ACES_SU-AS000012
Authors	ACESx, Martin Group
Instrument	QExactive Orbitrap HF-X (Thermo Scientific)
Instrument Type	LC-ESI-QFT

391ed2b0ea221bd037





Parviel Chirsir @PChirsir



https://twitter.com/AcademicTox/status/1605997310726443009

Help Expand “Chemical Space” with Expert Knowledge

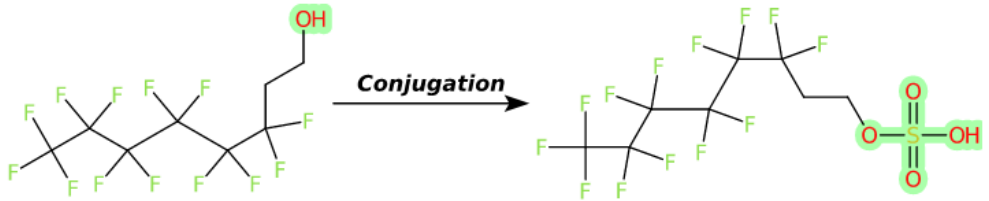
- Help us help you! Add data with FAIR templates & open license

Chemical Structures

PubChem_CID	Name	SMILES	InChIKey
2256	Atrazine	CCNC1=NC(=NC(=N1)Cl)NC(C)C	MXWJVTOOROXGIU-UHFFFAOYSA-N
2328	Bentazone	CC(C)N1C(=O)C2=CC=CC=C2NS1(=O)=O	ZOMSMJKLGFBRBS-UHFFFAOYSA-N
3030	Dicamba	COC1=C(C=CC(=C1C(=O)O)Cl)Cl	IWEDIXLBFLAXBO-UHFFFAOYSA-N
3120	Diuron	CN(C)C(=O)NC1=CC(=C(C=C1)Cl)Cl	XMTQQYYKAHVGBJ-UHFFFAOYSA-N

Transformations

Predecessor_CID	Predecessor_Name	Transformation	Successor_CID	Successor_Name	Biosystem
13101	6PPD	Ozone	154926030	6PPD-quinone	Environment
2256	Atrazine	Environmental	13878	Deisopropyl-atrazine	Soil
2256	Atrazine	Mammalian metabolism	135408770	Ammeline	Mammal
2256	Atrazine	Fungal metabolism	22563	Desethyl-atrazine	Fungus



6:2 FTOH to 6:2 FTOH-Sulfate

FAIR chemical structures in the Journal of Cheminformatics

Emma L. Schymanski and Evan E. Bolton

Letter to the Editor | 7 July 2021

The Letter Response to this article has been published in *Journal of Cheminformatics* 2021 **13**:49

https://ftp.ncbi.nlm.nih.gov/pubchem/Other/Submissions/

Reply to "FAIR chemical structure in the Journal of Cheminformatics"

Rajarshi Guha, Nina Jeliazkova, Egon Willighagen and Barbara Zdrazil

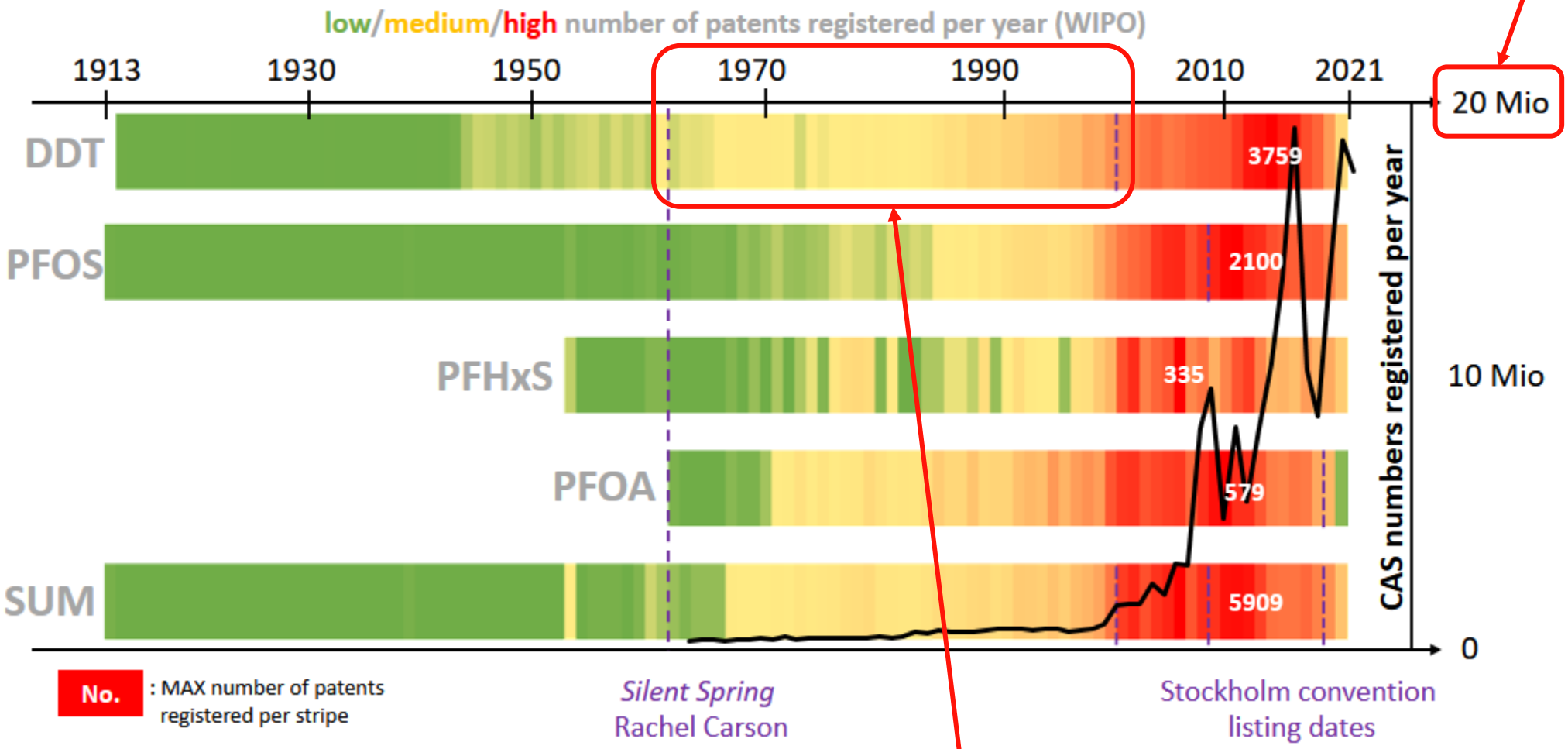
Letter Response | 7 July 2021

The Letter to the Editor to this article has been published in *Journal of Cheminformatics* 2021 **13**:50

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

The Grand Challenge of Chemical Numbers

…and timescale for regulatory actions …

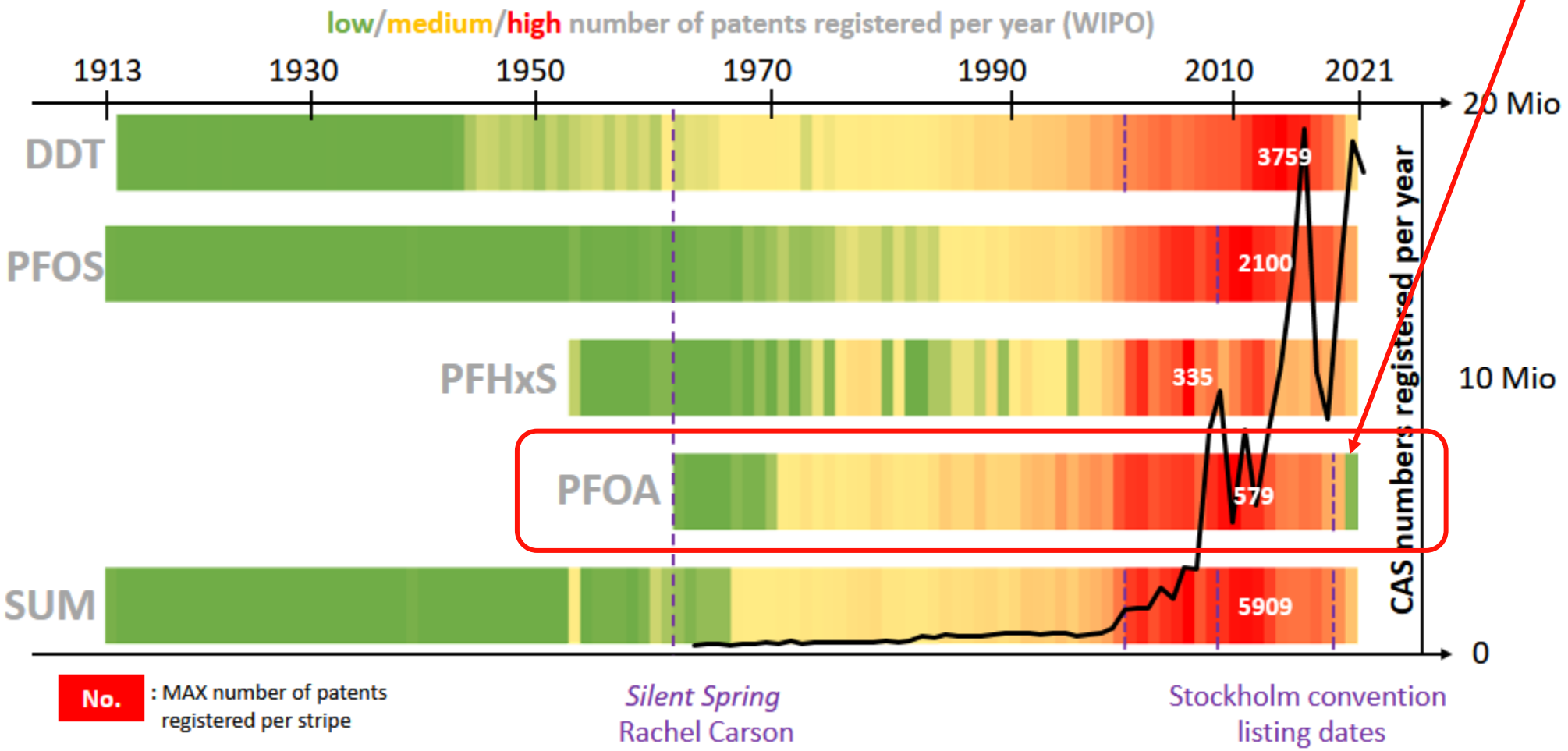


Arp *et al.* (2023) ES&T. DOI: 10.1021/acs.est.3c01735

The Next Grand Challenge – Tackling Patent Data

Can patent data reveal the next big trends?

Are we observing action and reaction with PFOA?



Arp *et al.* (2023) ES&T. DOI: 10.1021/acs.est.3c01735

The Chemical Stripes and Patent Data

Environmental Cheminformatics › chemicalstripes

chemicalstripes

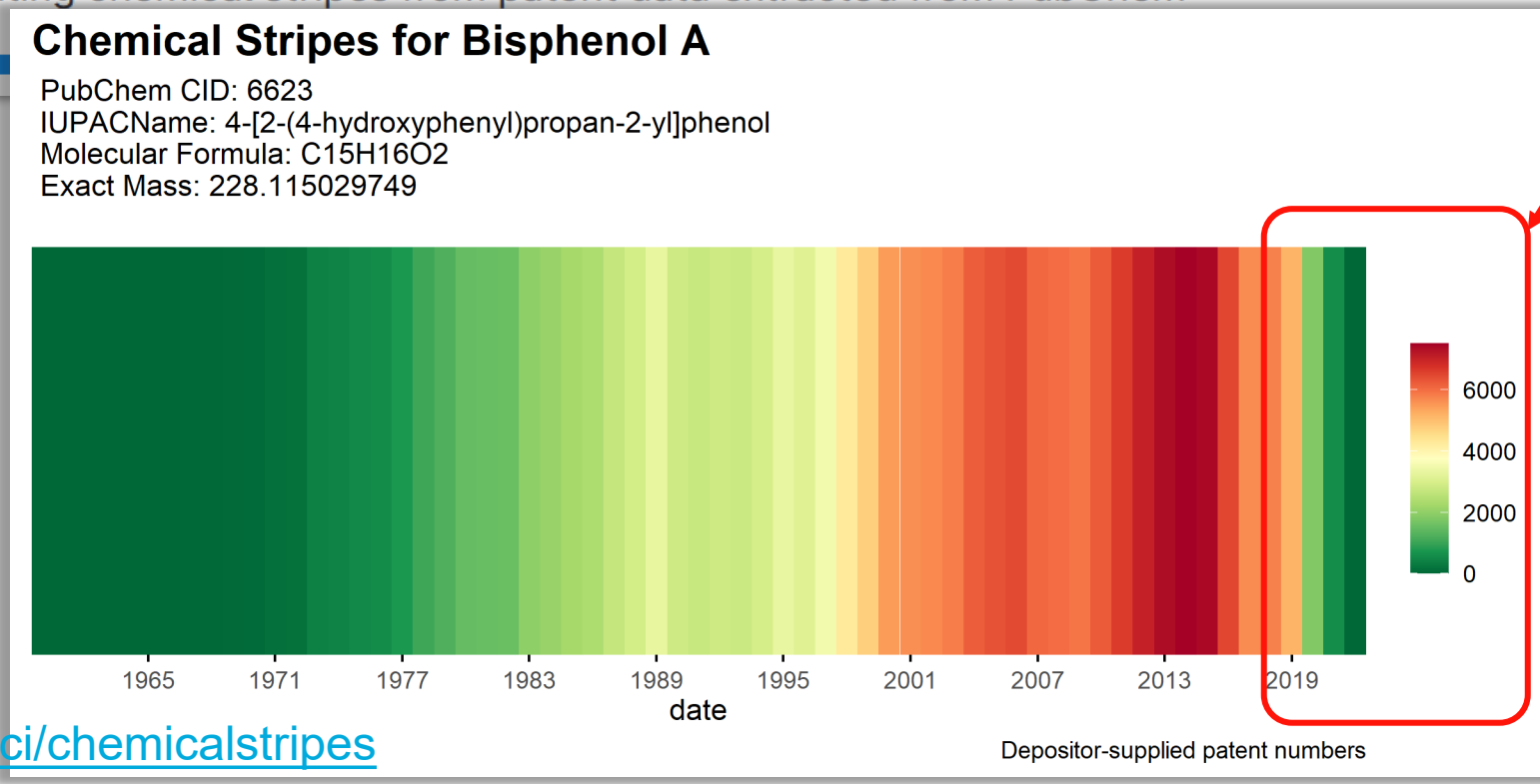
Project ID: 2985

Star 2 | Fork 0

20 Commits **1** Branch **0** Tags **485 KB** Project Storage

An R package for generating chemical stripes from patent data extracted from PubChem

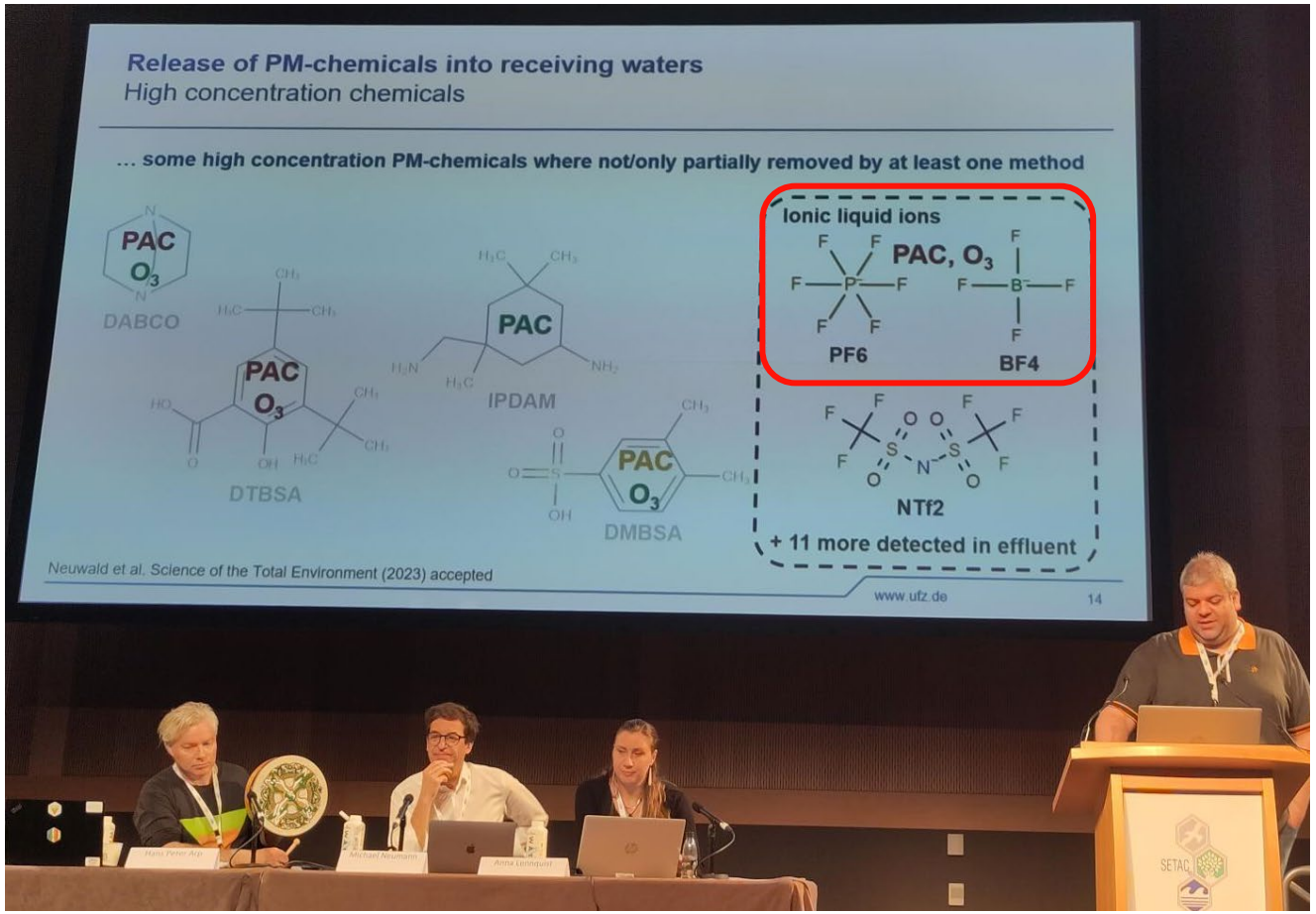
Also very low patent numbers post 2019…

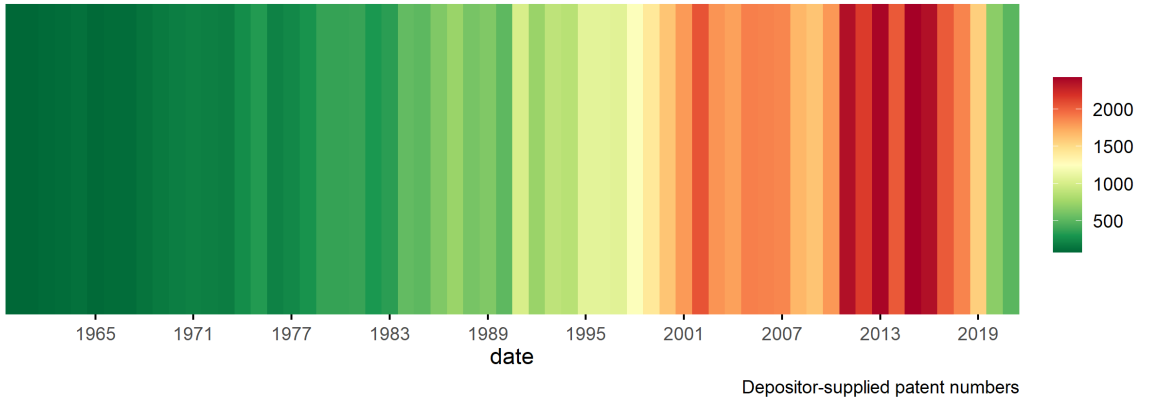


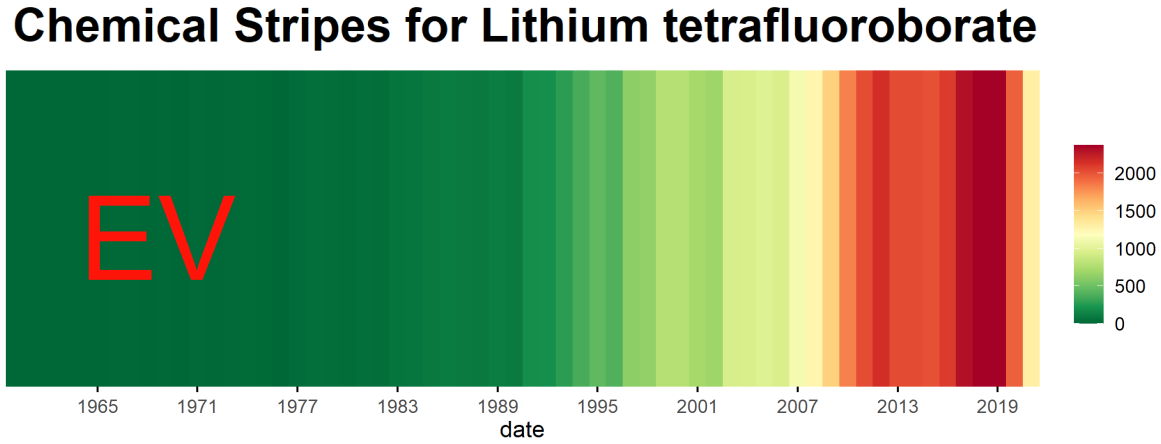
Patent data is huge
Data-rich stripes
(the nicest ones!)
> 10 minutes !!

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

The Chemical Stripes and Patent Data



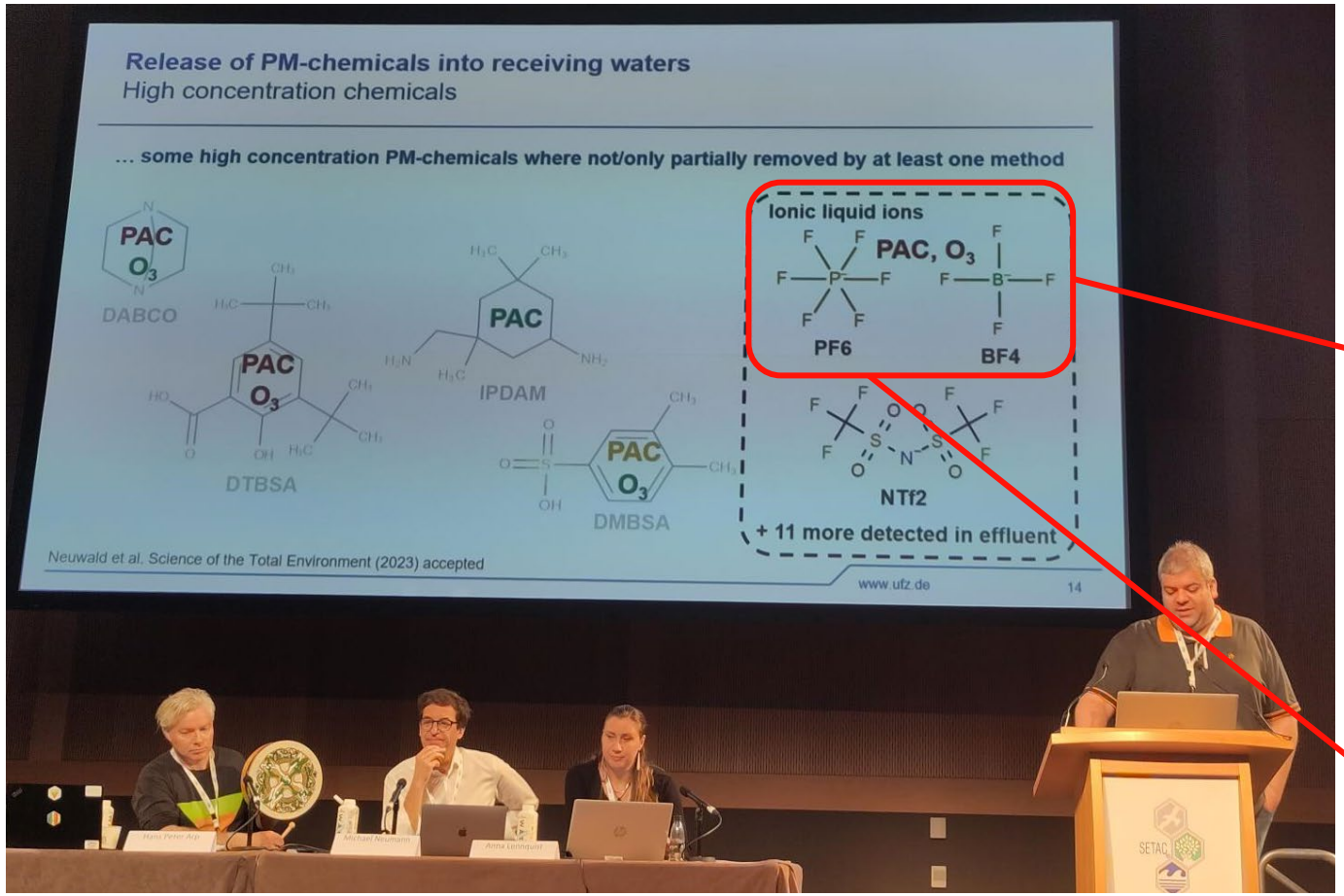




Left: Neuwald *et al, STOTEN*, accepted.
Photo of Daniel Zahn, UFZ at SETAC Europe, 30 April – 4 May, 2023. Image reused with permission

The Chemical Stripes and Patent Data

- Really big patent data…



PubChem Tetrafluoroborate (Compound)

5.2 Related Compounds

Same Connectivity Count	4
Mixtures, Components, and Neutralized Forms Count	19660
Similar Compounds Count	11

PubChem Hexafluorophosphate (Compound)

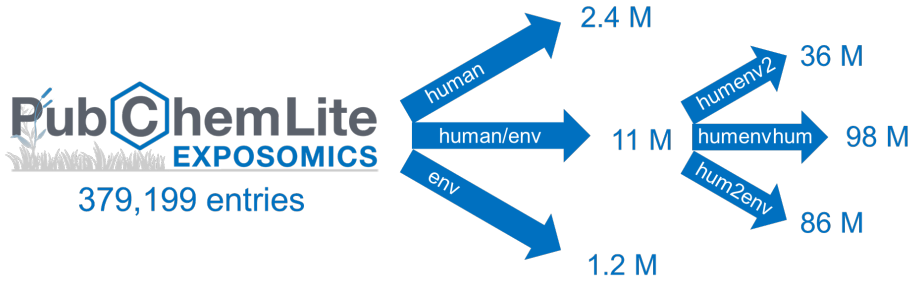
5.2 Related Compounds

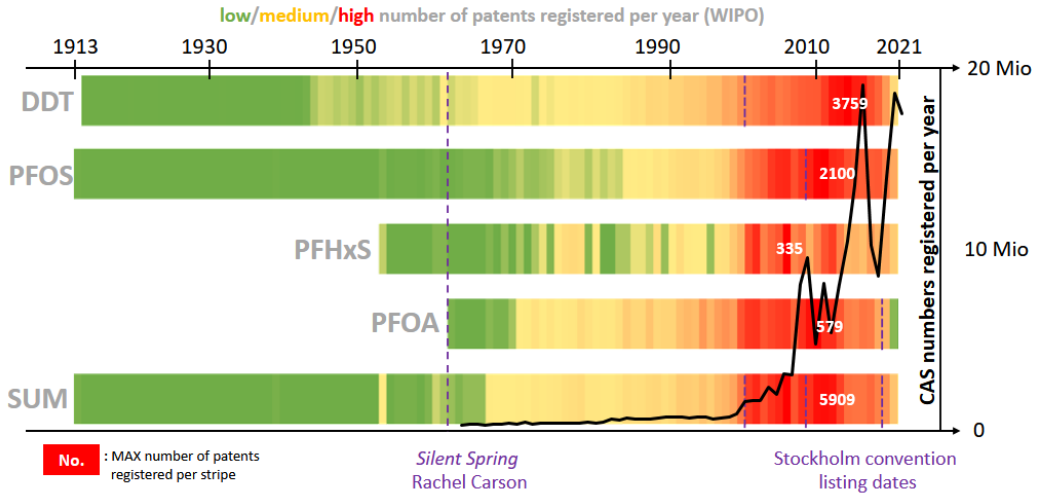
Mixtures, Components, and Neutralized Forms Count	15408
Similar Compounds Count	7

Left: Neuwald *et al, STOTEN*, accepted.
Photo of Daniel Zahn, UFZ at SETAC Europe, 30 April – 4 May, 2023. Image reused with permission

“Take Home” Messages

- Dynamic, Open Collection: PubChemLite EXPOSOMICS
  - Expert knowledge to fill the gaps - Annotation content is extremely powerful
- Transformations and Metabolism
  - More data needed to improve predictions
- Open, FAIR workflows available in patRoon & MetFrag
- PFAS & Chemical numbers are a huge challenge
- Patent data may help
  - …but is an even bigger challenge!





Acknowledgements!

Today's slides:

DOI: 10.5281/zenodo.7944066

Email: emma.schymanski@uni.lu

Twitter: @ESchymanski

Today's slides zenodo

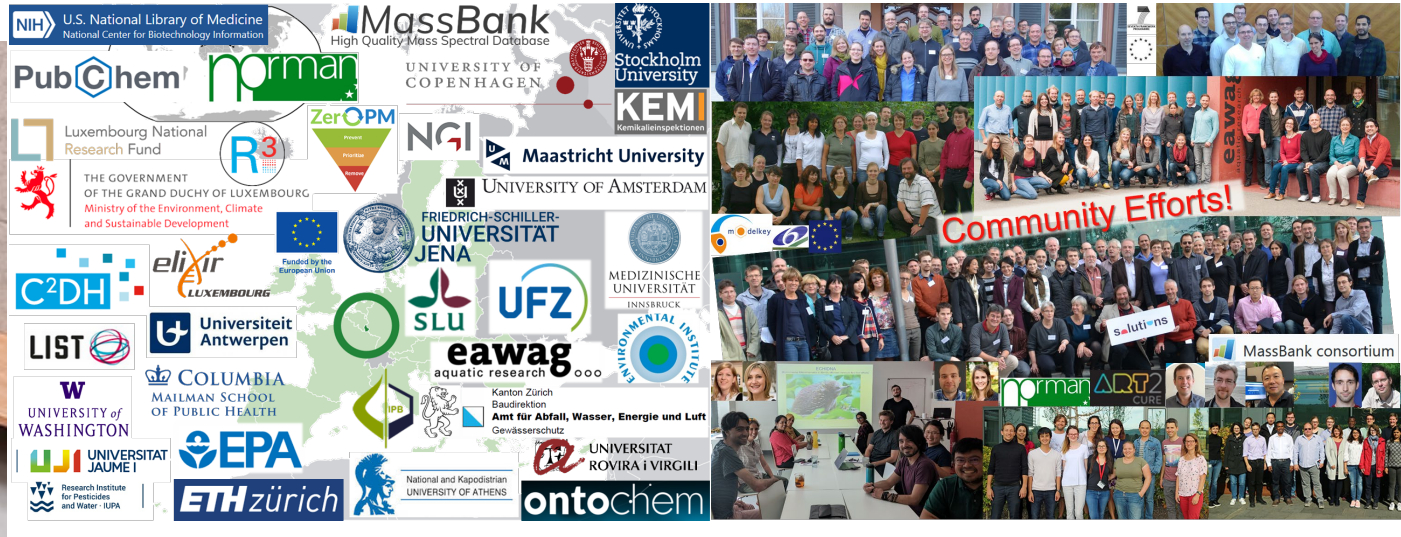
PubChem

@EvanBolton
@pubchem

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

pubchem-help@ncbi.nlm.nih.gov

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



This work was supported in part by the National Center for National Library of Institutes of Health

Funded by the European Union

H2020: 101036756