Luxembourg National Research Fund

UNIVERSITÉ DU LUXEMBOURG

Emerging Contaminants, Cheminformatics, Mass Spectrometry and the Exposome

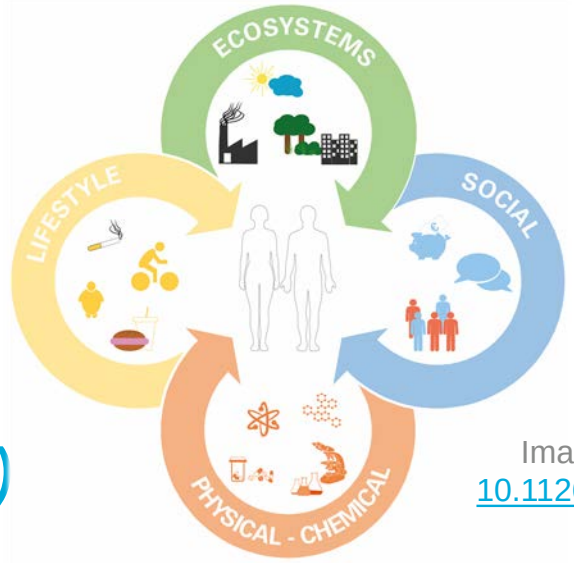


Image mod. from DOI: 10.1126/science.aay3164

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

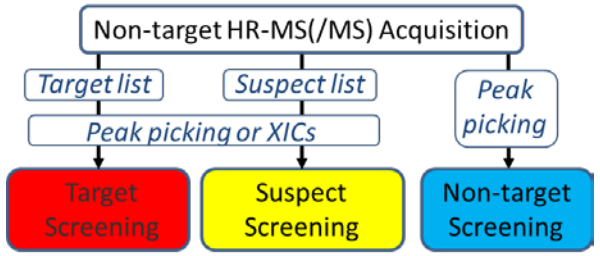
Environmental Cheminformatics Group,
Luxembourg Centre for Systems Biomedicine, University of Luxembourg
Email: emma.schymanski@uni.lu and @ESchymanski
Web: https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics/

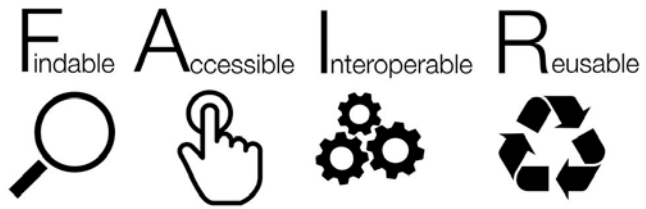
Emerging Contaminants Workshop (virtual)
Melbourne, Australia - October 26th, 2021

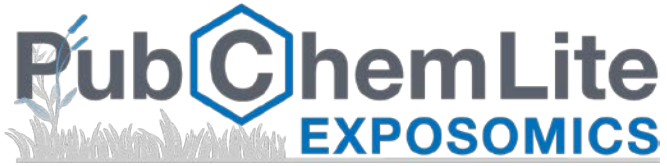
Slides available at
DOI: 10.5281/zenodo.5595630

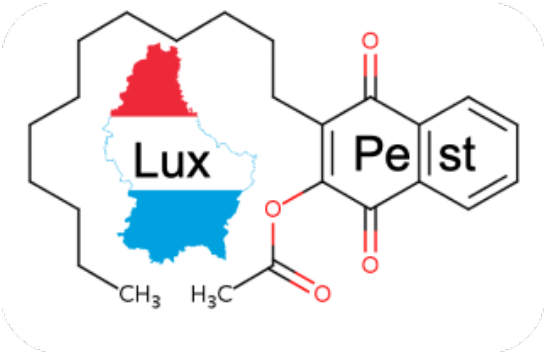
Outline of Today

- Introduction and Background
  - Definitions, Key Resources
  - Examples
- Identification & Chemical Space
  - Identification + MetFrag
  - PubChemLite for Exposomics
- Case Study: LuxPest
- Take-home messages!









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

Luxembourg, LCSB and the University

University of Luxembourg
- founded in 2003
Luxembourg Centre for Systems Biomedicine (LCSB)
- founded in 2009
Young & dynamic environment!

https://qbuild.lu/realisations/residence-rouden-eck/

Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885.

Environmental Cheminformatics @ LCSB

Adelene Lai
@adelene.lai
Given access 1 year ago

Anjana Elapavalore
@anjana.elapavalore
Given access 1 year ago

Begoña Talavera Andújar
@begona.talavera
Given access 4 months ago

Corey Griffith
@corey.griffith
Given access 1 year ago

Dagny Aurich
@dagny.aurich
Given access 3 months ago

Emma Schymanski @emma.schymanski
Given access 1 year ago

German Andres Preciat Gonzales
@german.preciat
Given access 1 year ago

Hiba Mohammed-Taha @hiba.mohammed-taha
Given access 4 months ago

Jessy Krier @jessy.krier
Given access 1 year ago

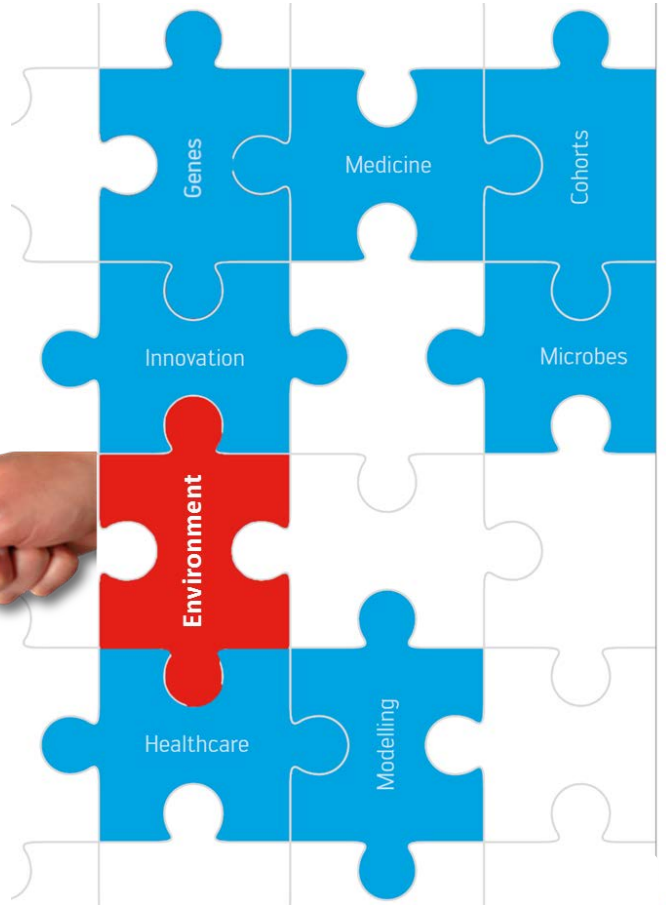
Lorenzo Favilli
@lorenzo.favilli
Given access 1 year ago

Mira Narayanan
@mira.narayanan
Given access 2 months ago

Parviel Chirsir
@parviel.chirsir
Given access 1 month ago

Todor Kondic
@todor.kondic
Given access 1 year ago

R3



INSTITUTE FOR ADVANCED STUDIES (IAS)

Luxembourg National Research Fund

… plus many National, European & International Collaborations!

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

nprman

PubChem

820 Data Sources

MassBank
High Quality Mass Spectral Database

ZerOPM
Prevent
Prioritize
Remove

WIKIDATA

Funded by the European Union

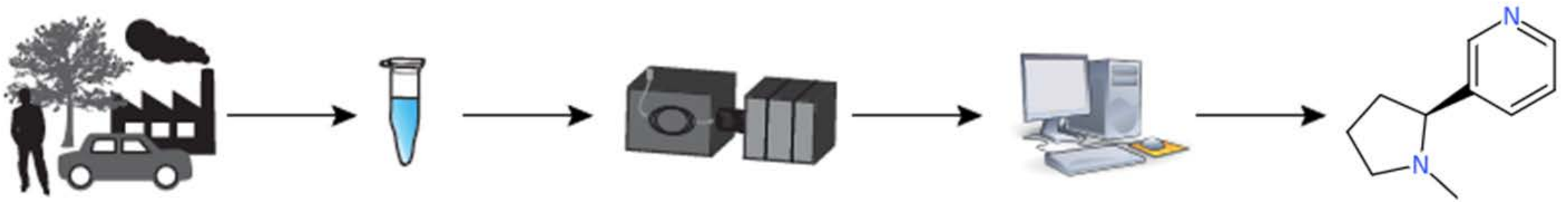
R3

Luxembourg National Research Fund

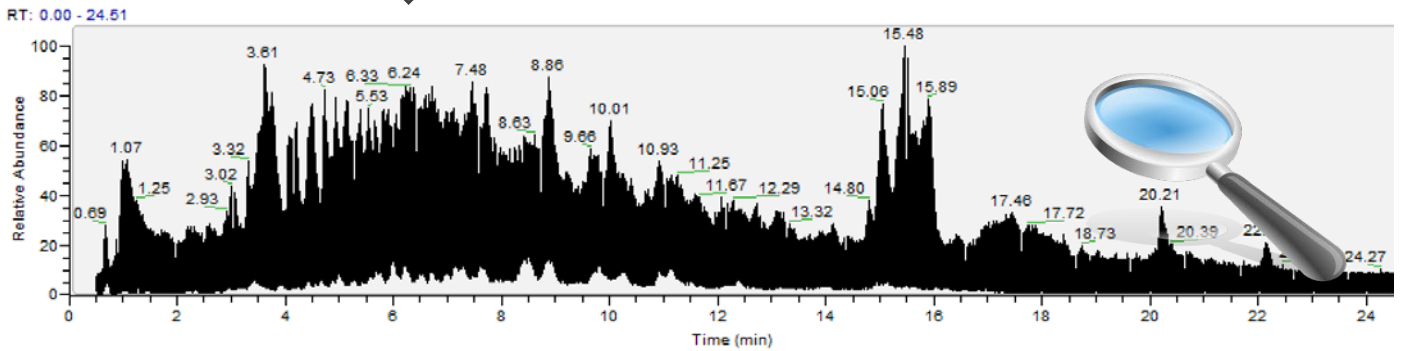
Source:
https://en.wikipedia.org/wiki/File:EU-Luxembourg.svg

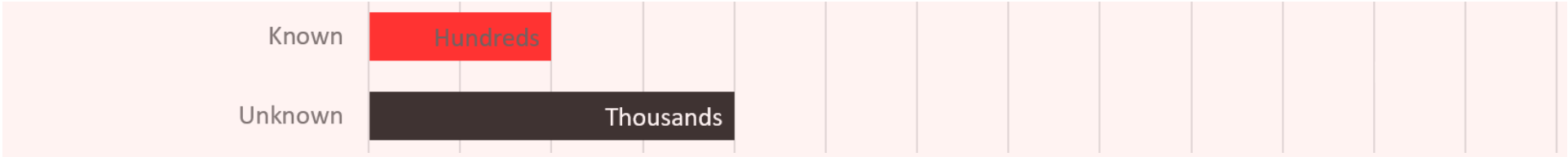
uni.lu | LCSB

Environmental Cheminformatics & HR-MS

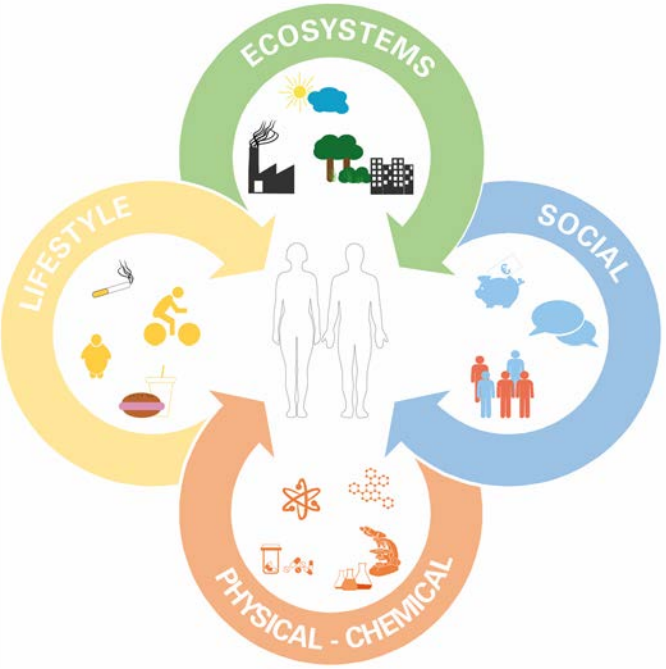


High resolution mass spectrometry





Environment … versus … The Exposome



- Christopher Wild (2005)

  "All exposures from conception onwards, including those from lifestyle, diet and the environment"

- Miller and Jones (2014) functionalized this to:

  "The cumulative measure of environmental influences and associated biological responses throughout the lifespan, including exposures from the environment, diet, behaviour and endogenous processes"

Wild (2005) DOI: 10.1158/1055-9965.EPI-05-0456; Miller & Jones (2014) DOI: 10.1093/toxsci/kft251;
Image c/o Lucie Debroux LCSB, modified from images by Utrecht Uni & Vermeulen *et al.* (2020). Science. DOI: 10.1126/science.aay3164.

The Exposome and Emerging Contaminants

- Emerging contaminants are part of the exposome …
  - But not the whole picture
- Research by necessity crosses environmental, biological, medical fields and beyond …
  - “exposomics”

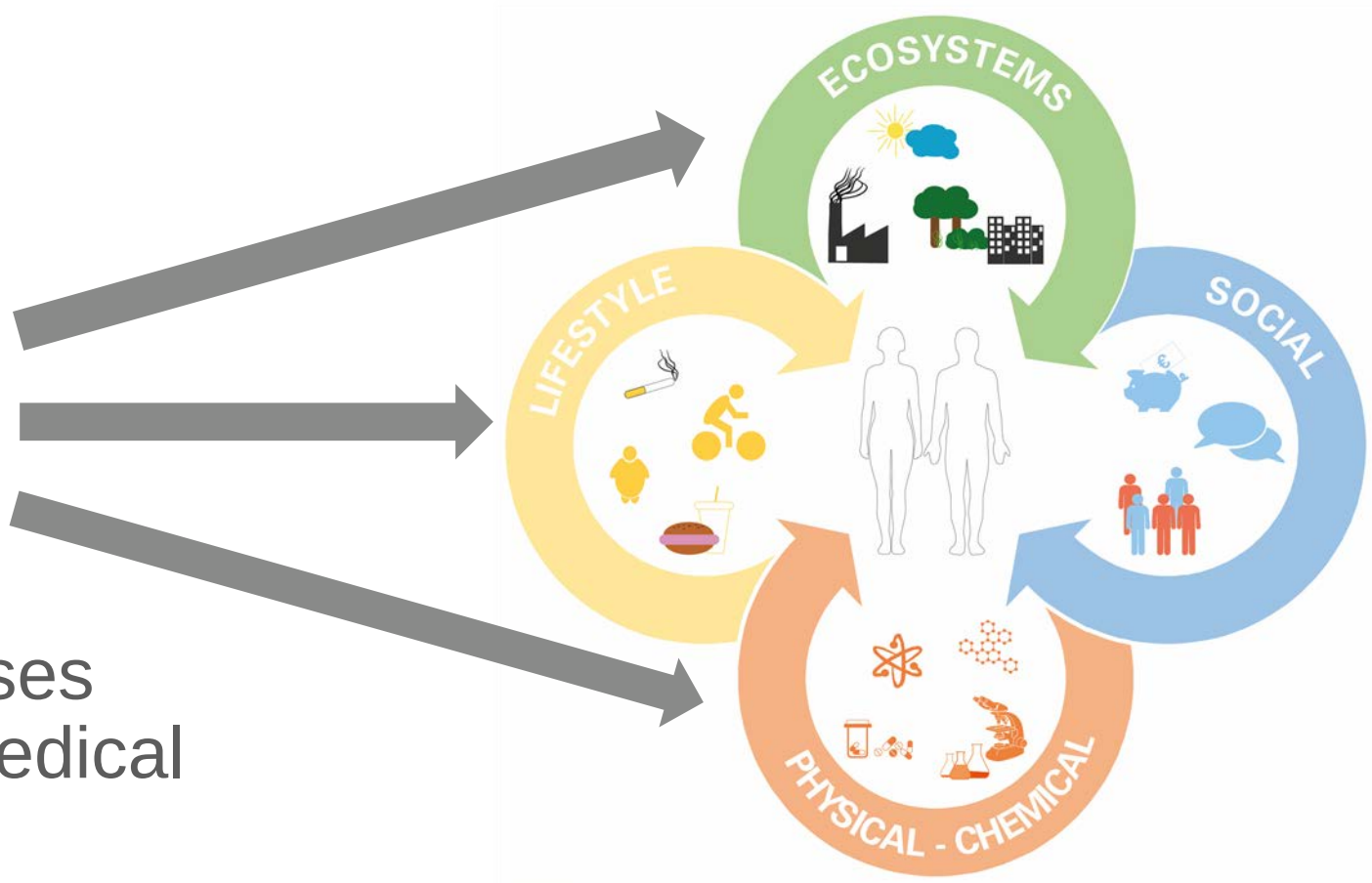
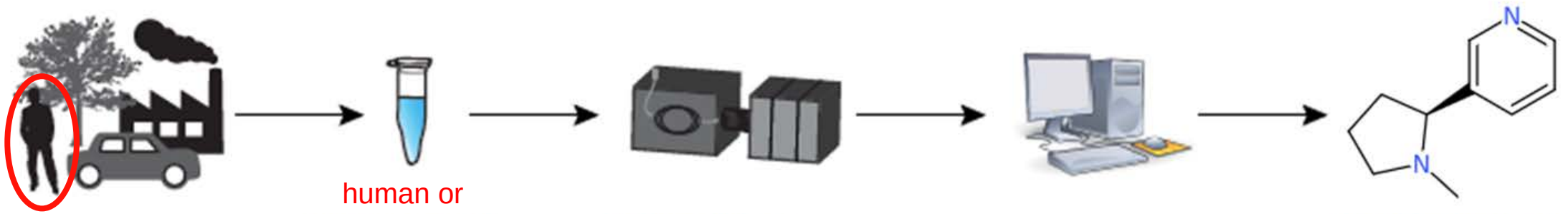
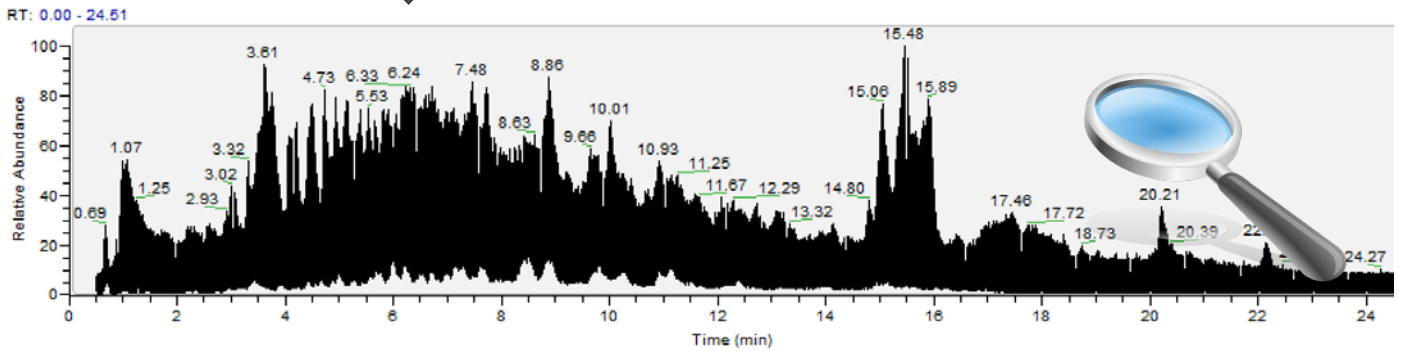


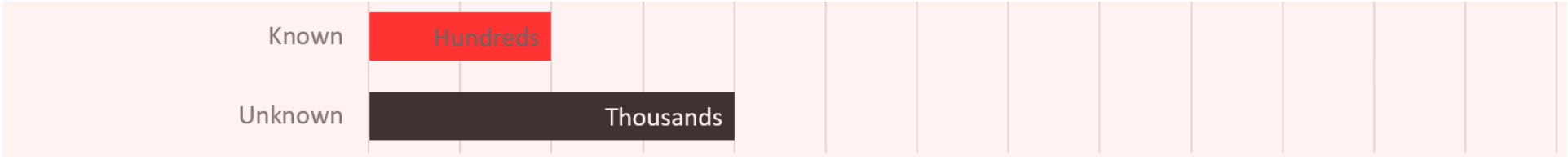
Image c/o Lucie Debroux LCSB, modified from images by Utrecht Uni & Vermeulen *et al.* (2020). Science. DOI: 10.1126/science.aay3164.

Environmental Cheminformatics, *Exposomics* & HR-MS



High resolution mass spectrometry

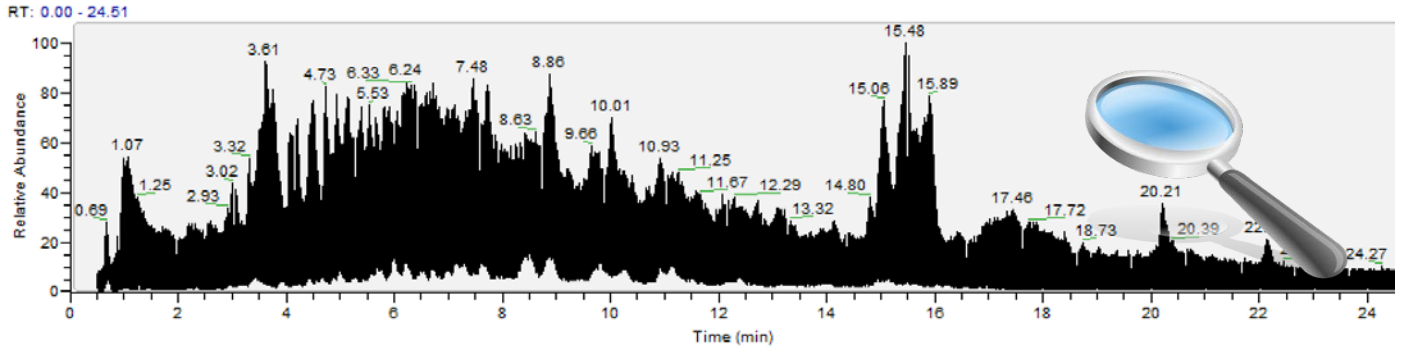


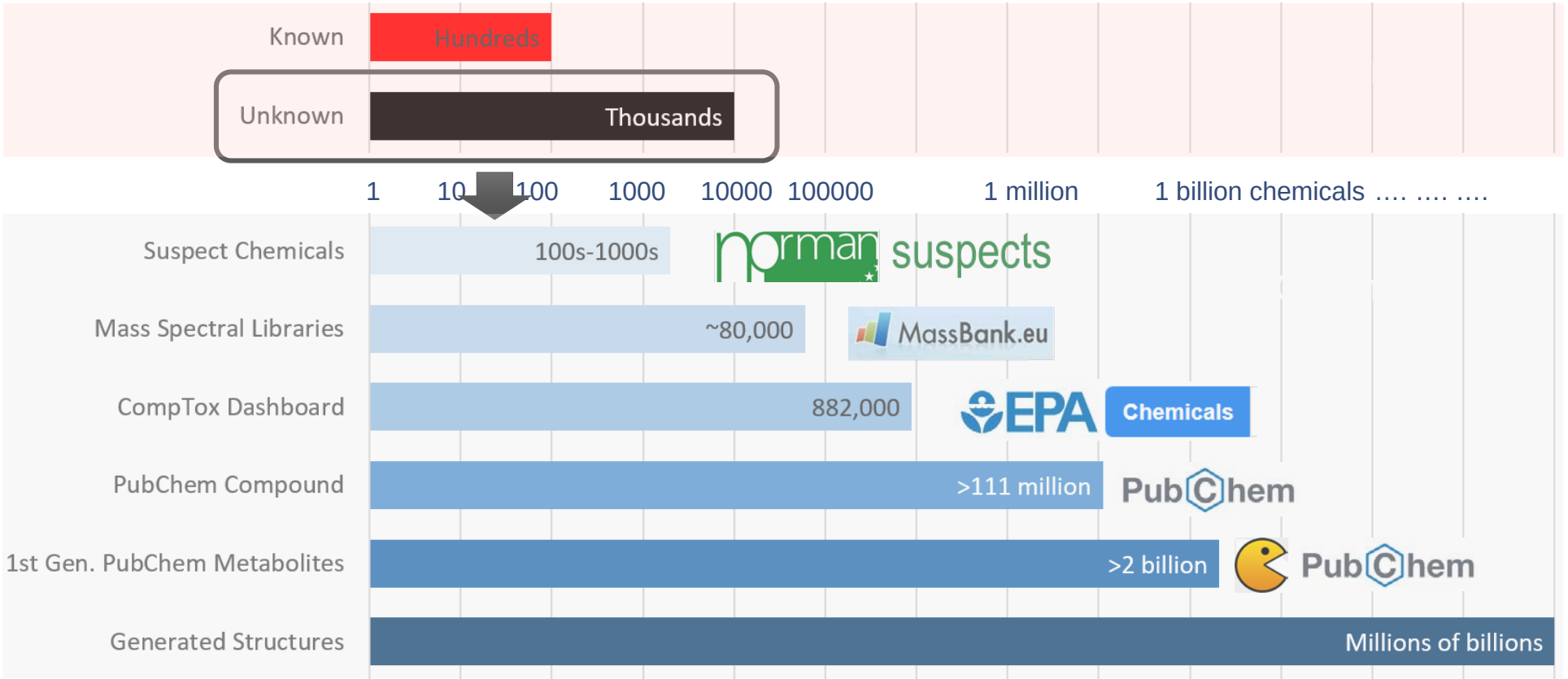


Environmental Cheminformatics, *Exposomics* & HR-MS

High resolution mass spectrometry

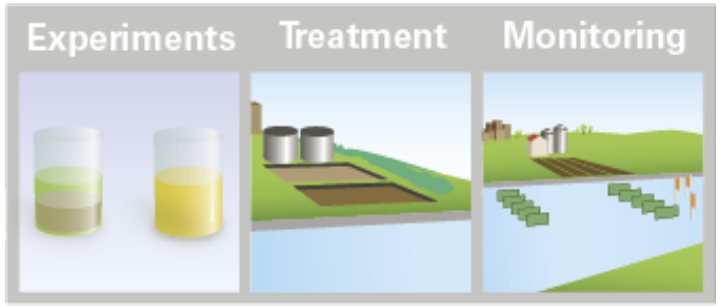
AND connecting chemical knowledge





Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Vermeulen *et al.* (2020) DOI: 10.1126/science.aay3164

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

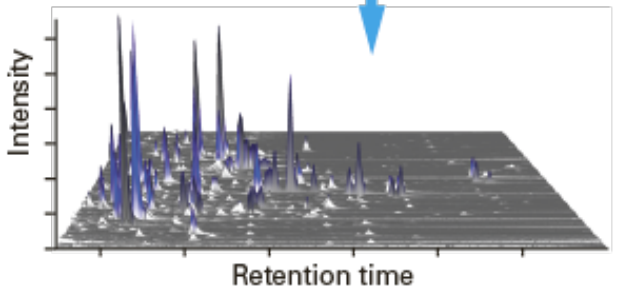


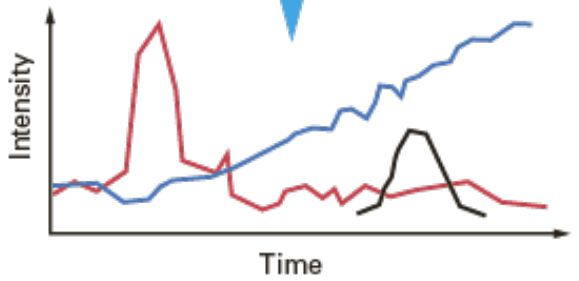
Commercial / Open Software

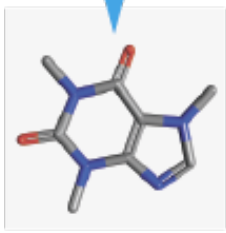
Databases

In Silico Prediction

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







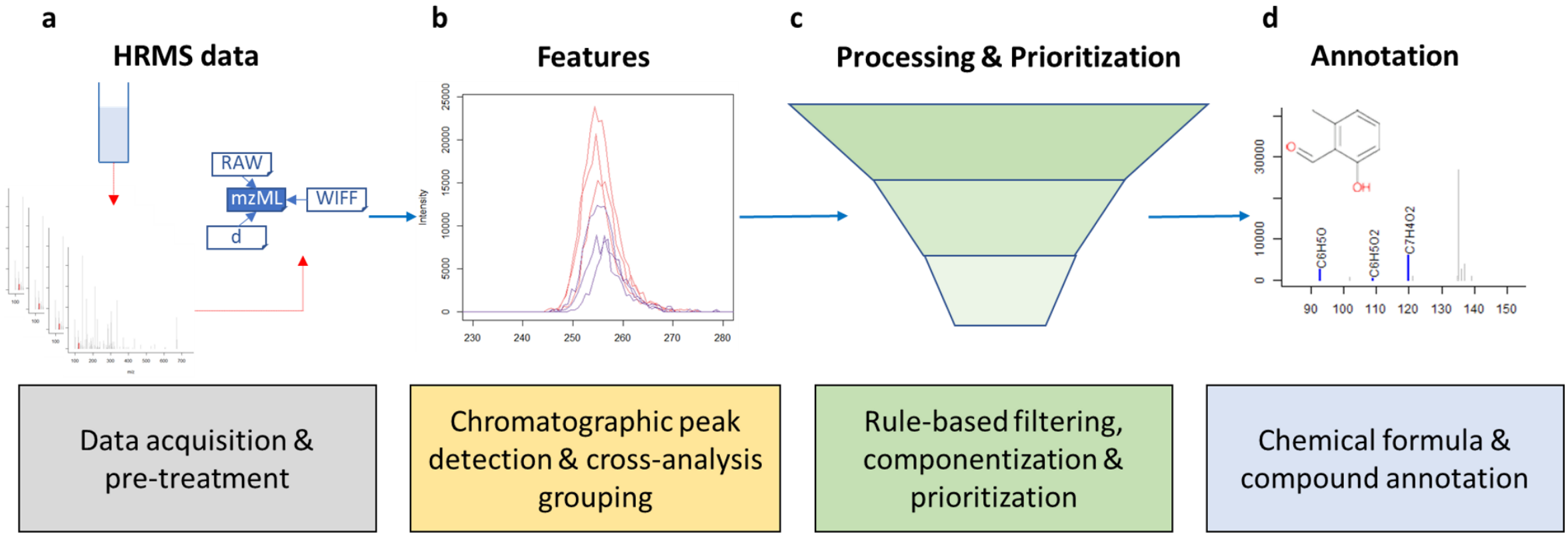
SAMPLE ARCHIVE | DIGITAL ARCHIVE | POLLUTION OVERVIEW | PATTERN | STRUCTURES

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

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

Rick Helmus



Helmus *et al.* (2021) patRoon. *J. Cheminformatics*. DOI: 10.1186/s13321-020-00477-w

University of Amsterdam

Open Source Workflows for NT-HRMS: patRoon

Rick Helmus

Software | Open Access | Published: 06 January 2021

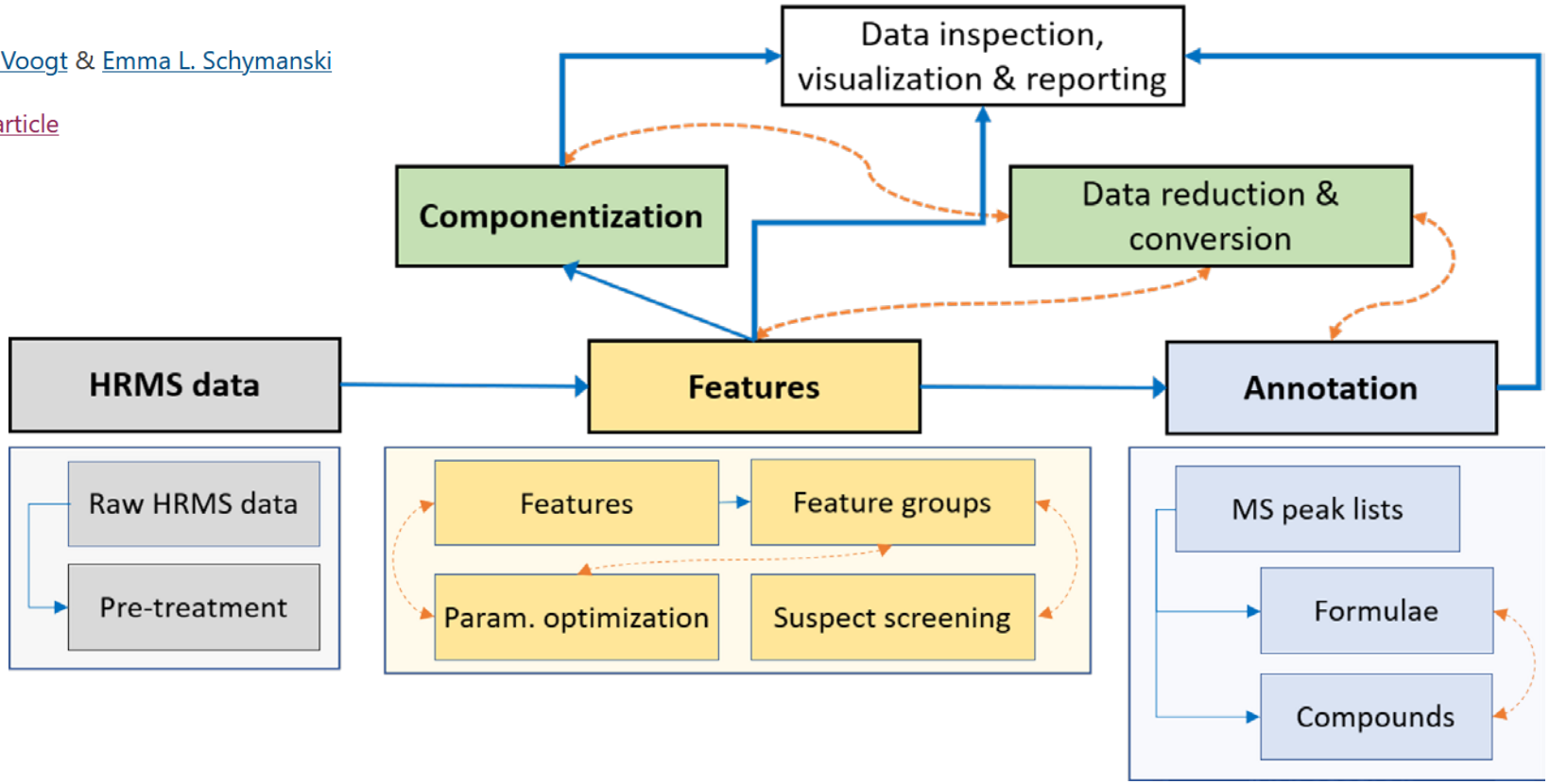
patRoon: open source software platform for environmental mass spectrometry based non-target screening

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

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

1394 Accesses | **2** Citations | **20** Altmetric | Metrics

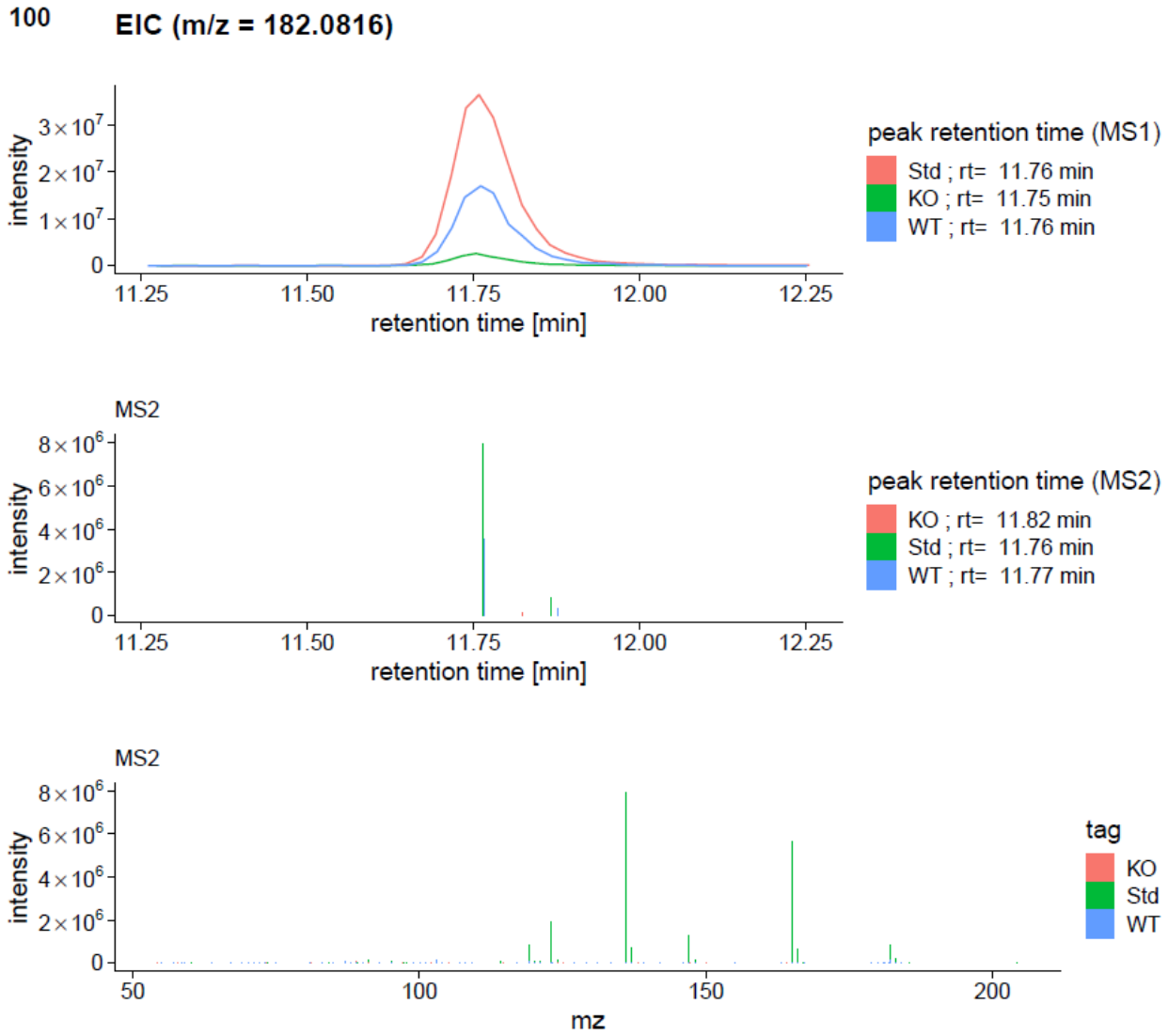
Journal of Cheminformatics



UNIVERSITY OF AMSTERDAM

Helmus *et al.* (2021) patRoon *J Cheminform* **13,** 1. DOI: 10.1186/s13321-020-00477-w

Open Source Workflows for NT-HRMS: Shinyscreen



https://git-r3lab.uni.lu/eci/shinyscreen

Anjana Elapavalore, Mira Narayanan, Todor Kondic, Jessy Krier, Hiba Mohammed Taha.

Kondić *et al.* (in prep) https://gitlab.lcsb.uni.lu/eci/shinyscreen

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 | Peak Differences

Compound Information

Compound name

Exact Mass

AND

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

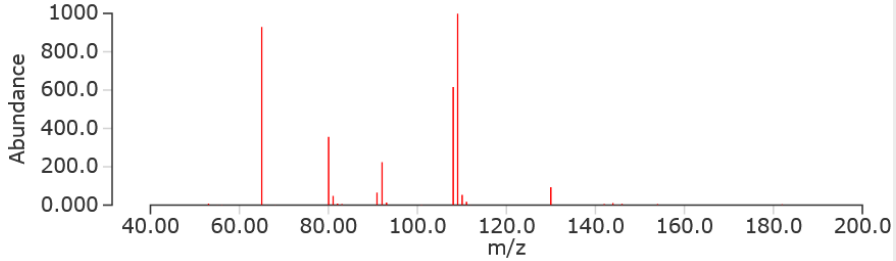
AND

Search

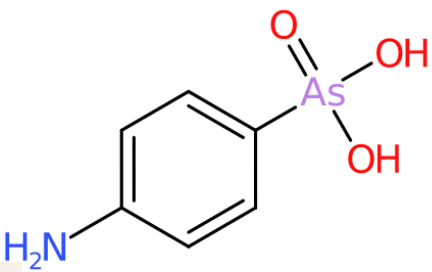
MassBank Record: LU040605

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

Mass Spectrum

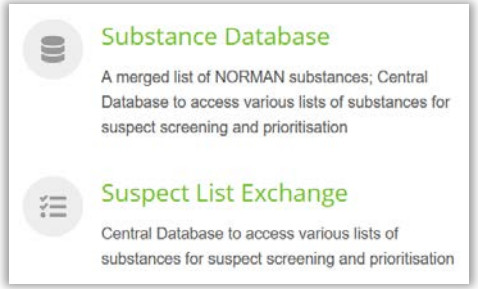


Chemical Structure



Expert Knowledge: NORMAN Database System

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



NORMAN Database System

NORMAN organises the development and maintenance of various web-based databases for the collection & evaluation of data / information on emerging substances in the environment

SEARCH All Databases

Searching for individual substance or group(s) of substances in all databases

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

SARS-CoV-2 in sewage

A database with the latest information on SARS-CoV-2 in sewage across Europe and internationally; including a common protocol for sample collection, storage, extraction, analysis and data sharing to support the development of an international comparable data set.

Substance Database

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

Chemical Occurrence Data

A database of geo-referenced monitoring data on emerging substances

Ecotoxicology

A platform for systematic collection and evaluation of ecotoxicity studies for harmonised derivation of environmental quality standards

Suspect List Exchange

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

Antibiotic Resistance Bacteria/Genes

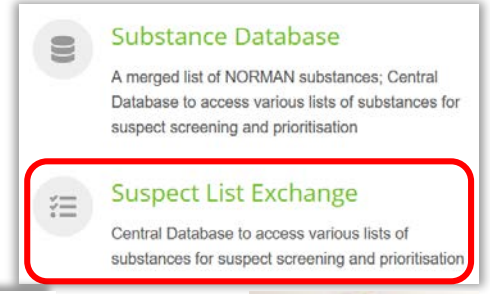
A database of ARBs/ARGs in environmental matrices

MassBank Europe

A database of mass spectra of emerging substances to support identification of unknown substances

Expert Knowledge: NORMAN-SLE

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



Contact us:
normansle@uni.lu

NORMAN WEBSITE | NORMAN DATABASE SYSTEM | HOME | LOGIN

NORMAN SUBSTANCE DATABASE

NORMAN Suspect List Exchange – NORMAN SLE

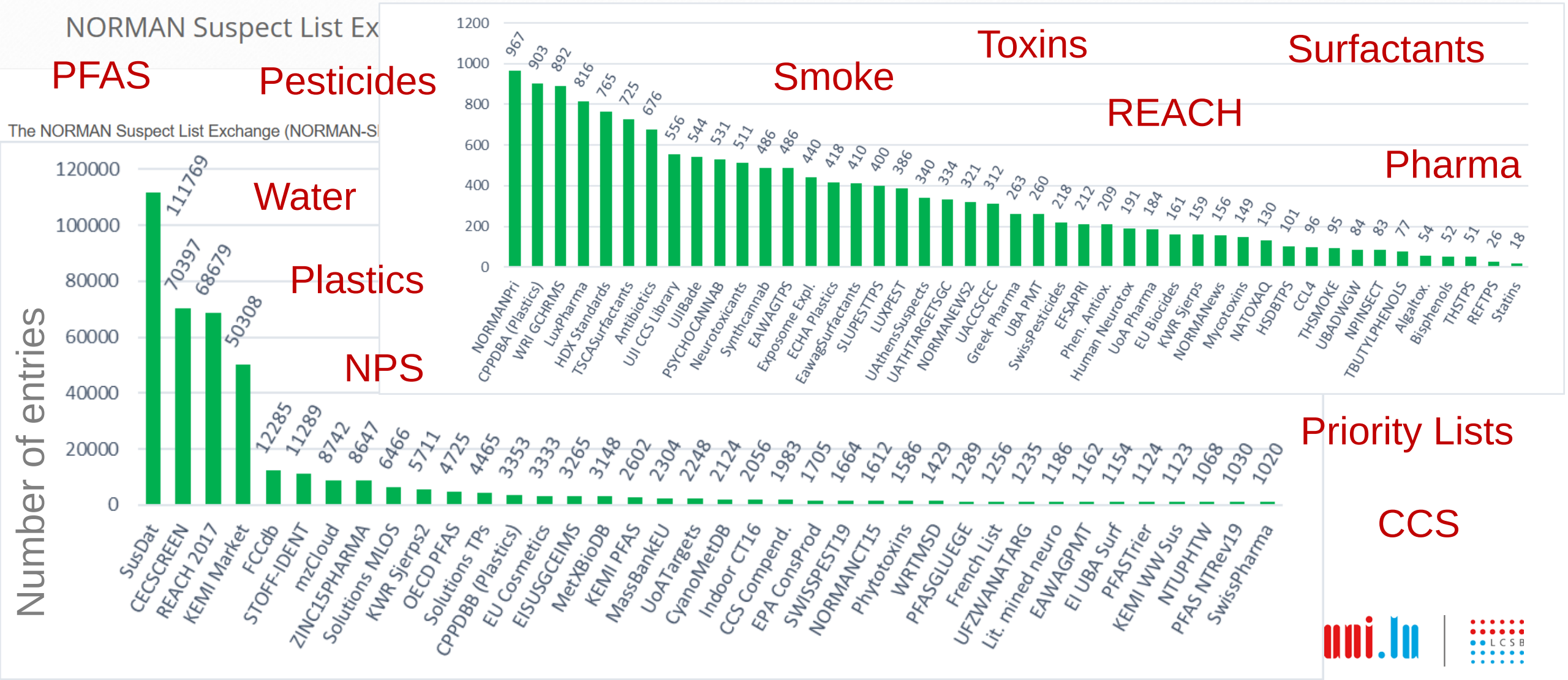
The NORMAN Suspect List Exchange (NORMAN-SLE) was established in 2015 as a central access point for NORMAN members (and others) to find suspect lists relevant for their environmental monitoring question. This Exchange documents all individual collections that form a part of NORMAN SusDat, the merged NORMAN Substance Database (DOI: 10.5281/zenodo.2664077).

UPDATE: Dec 2020: Check out updated Transformations Tables and the NORMAN-SLE Classification Tree in PubChem!

No.	Abbreviation	Description	Link to full list	Link to InChIKey list	References
S0	SUSDAT	**Merged NORMAN Suspect List: SusDat**	Interactive Data table SusDat with Haz and Expo scores as XLSX, CSV (06/11/2020) MetFrag CSV (03/03/2020) CompTox SUSDAT List	SusDat InChIKeys: All, MS-ready (18/06 /2020)	A merged list of >111,000 structures from SLE suspect lists. See interactive version. Compiled by Reza Aalizadeh, Nikiforos Alygizakis and Lubos Cirka, University of Athens/EI, including RTI and toxicity values, with Hazard and Exposure values provided by Stellan Fischer, KEMI, documented here. *Work in progress ... please report any issues!* DOI: 10.5281/zenodo.2664077
S1	MASSBANK	**NORMAN Compounds in MassBank**	CSV, XLSX with Fragments (3/10/2017) CompTox MassBank EU Reference List	MassBankEUInChIKeys (17/06/2019)	www.massbank.eu Stravs *et al.* 2013. DOI: 10.1002/jms.3131

NORMAN Suspect List Exchange (now >80 lists!)

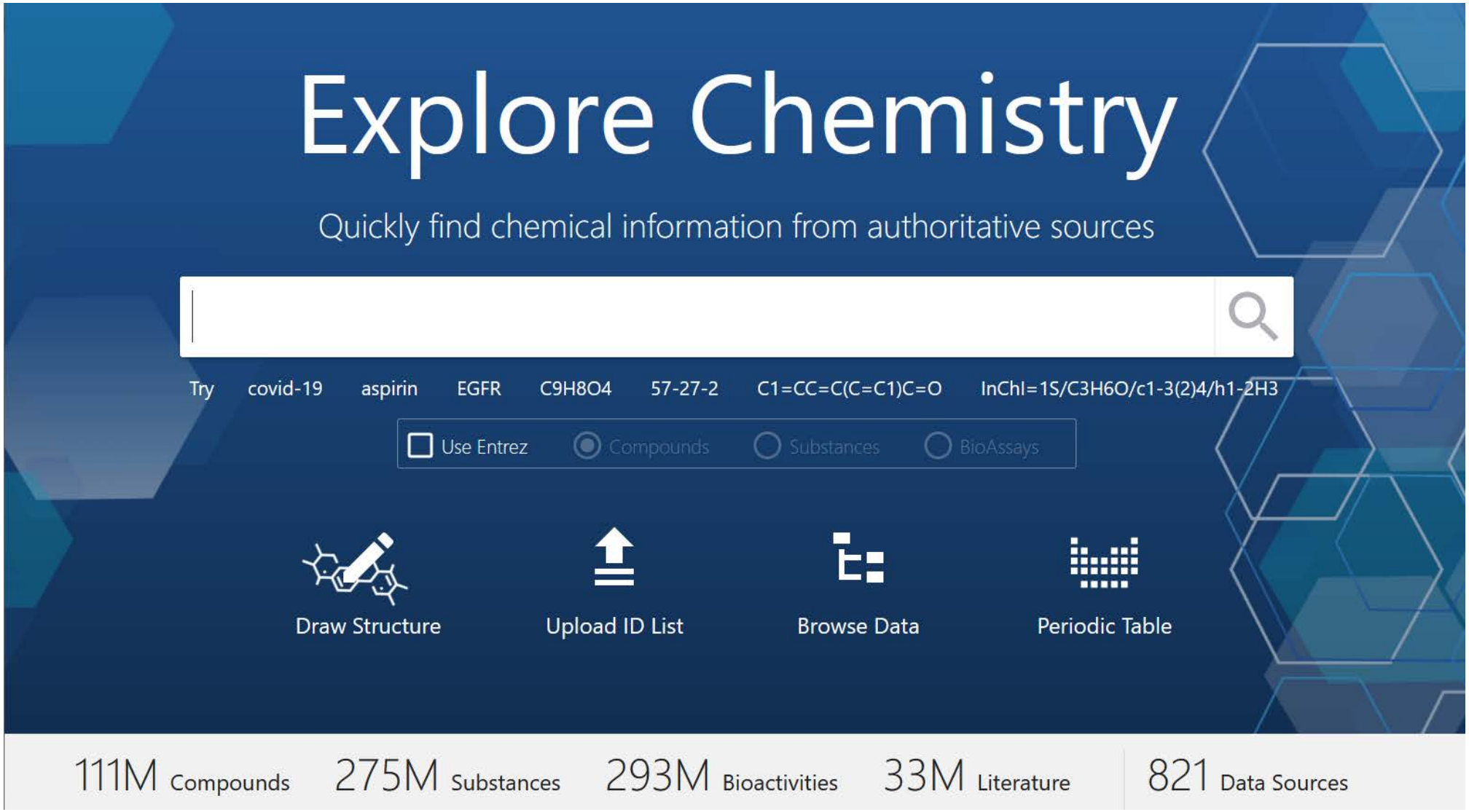
https://www.norman-network.com/nds/SLE/ https://zenodo.org/communities/norman-sle



Open Chemistry Knowledge bases: PubChem

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

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



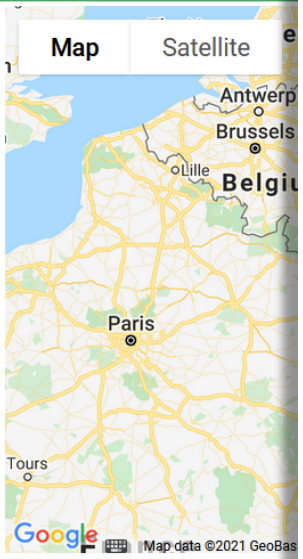
NORMAN-SLE in PubChem

DATA SOURCES

NORMAN Suspect List Exchange

The NORMAN network enhances the exchange of information on emerging environmental substances, an harmonisation of common measurement methods and monitoring tools so that the requirements of risk a met. It specifically seeks both to promote and to benefit from the synergies between research teams from emerging substances.

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	115,138 Live Substances 16,752 Annotations 1 Classification
Last Updated	2021/10/09

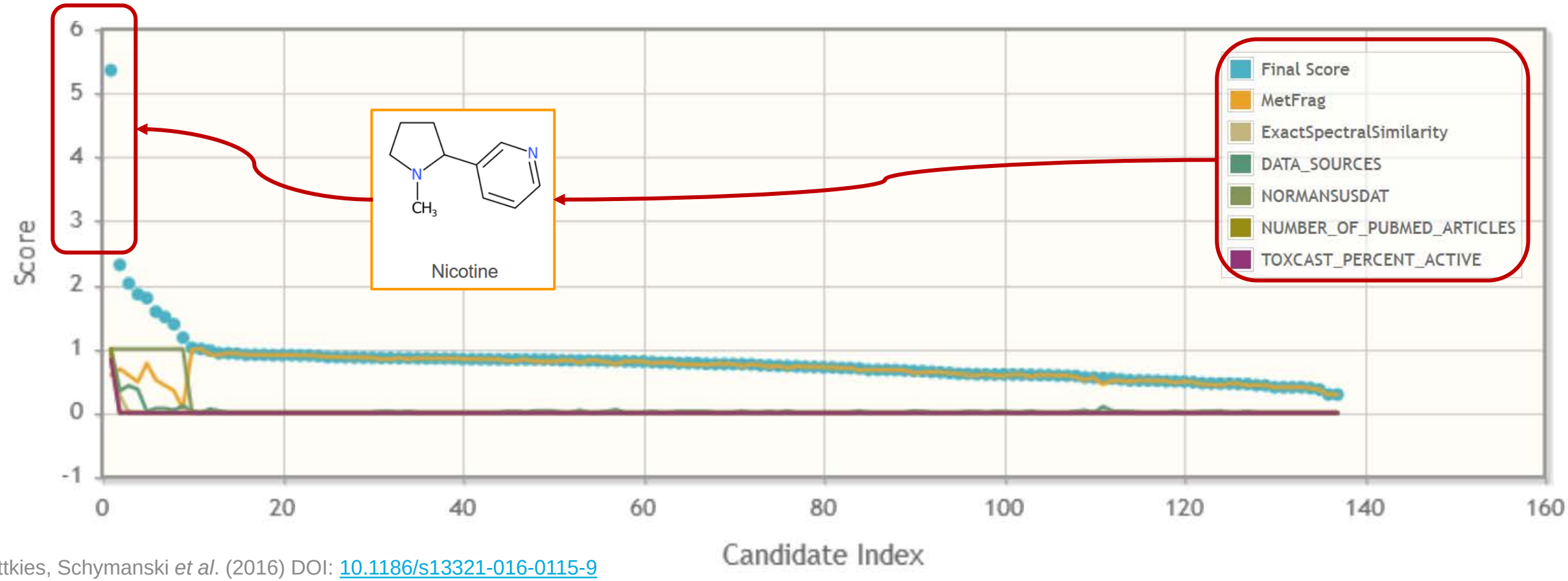


- NORMAN Suspect List Exchange Classification ? 113,715
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 3,935
  - S25 | OECDPFAS | List of PFAS from the OECD ? 3,677
  - S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances ? 254
  - S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 885
  - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ? 1,344
  - S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574
  - S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258
  - S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem ? 102
  - S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386
  - S72 | NTUPHTW | Pharmaceutically Active Substances from National Taiwan University ? 1,068
  - S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria ? 2,088
  - S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp ? 148
  - S00 | SUSDAT | Merged NORMAN Suspect List: SusDat ? 99,130
  - S01 | MASSBANK | NORMAN Compounds in MassBank EU ? 7,166
  - S02 | STOFFIDENT | HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances ? 11,261

Data source: https://pubchem.ncbi.nlm.nih.gov/source/23819
Classification tree: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101

Connecting Knowledge for Chemical Identification: MetFrag

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

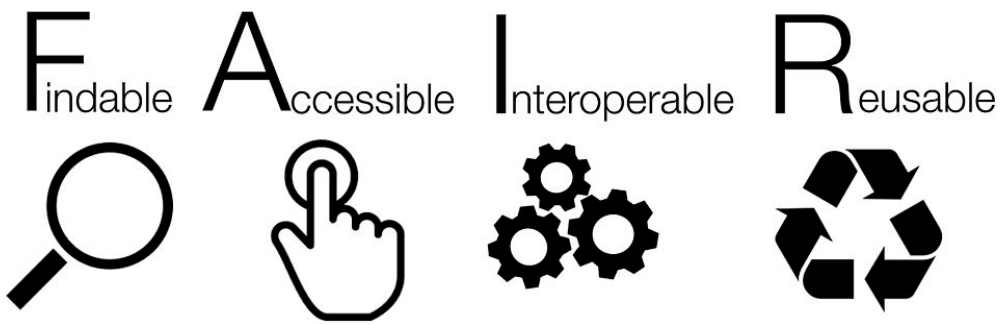


Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9

How? Community Efforts … through *Open* and *FAIR* Science

- Open access (OA) is a subset of *Open Science*:
  - Research outputs are distributed online,
  - Free of cost or other access barriers

- The *FAIR* principles – how to make data:
  - Findable,
  - Accessible,
  - Interoperable and
  - Reusable

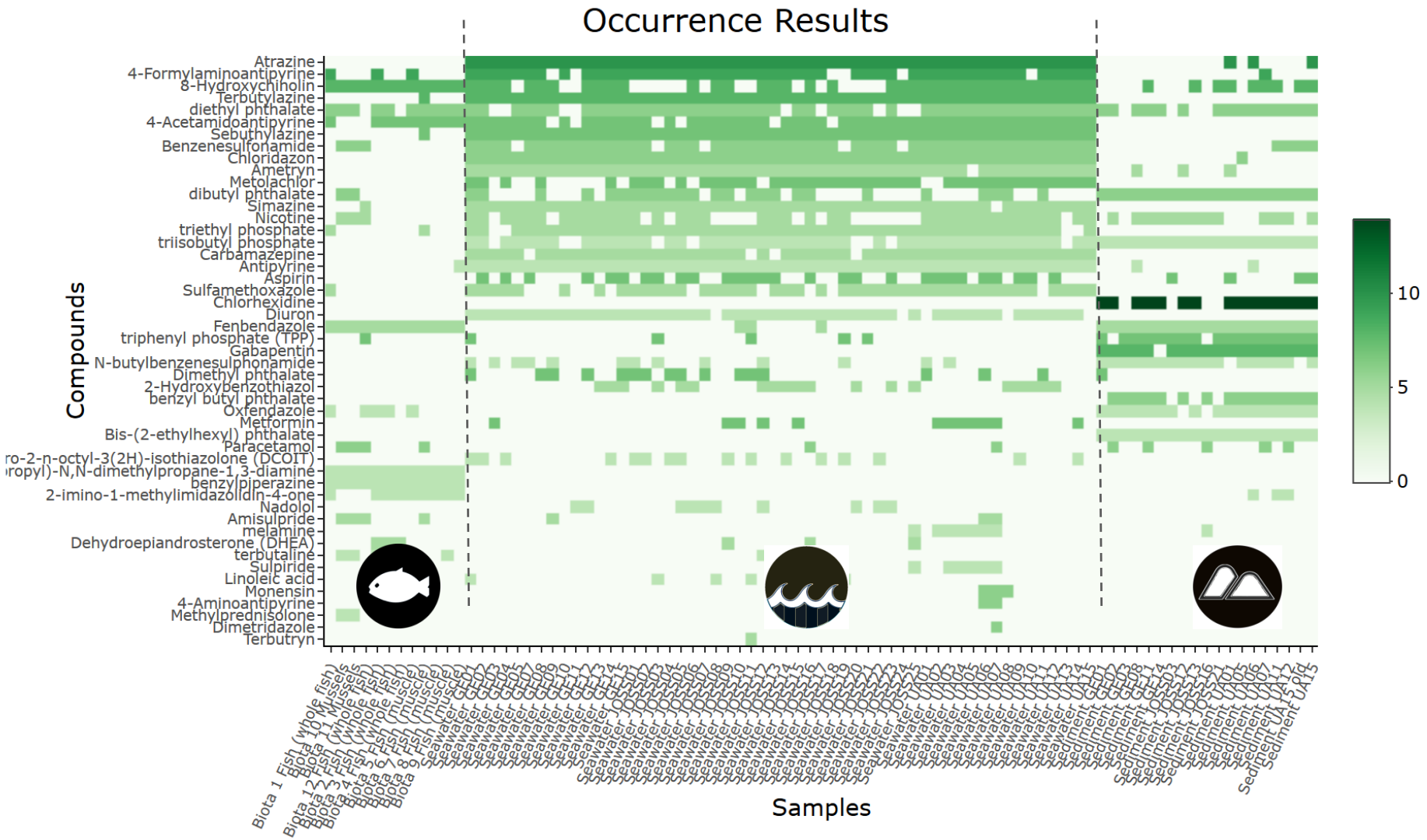


https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

P Suber: http://legacy.earlham.edu/~peters/fos/overview.htm & GO FAIR: https://www.go-fair.org/fair-principles/ &
MD Wilkinson et al. DOI:10.1038/sdata.2016.18 & EL Schymanski & EE Bolton Preprint DOI: 10.5281/zenodo.5495109

Examples: Distribution of Chemicals in Various Matrices

Retrospective screening of REACH chemicals in Black Sea samples

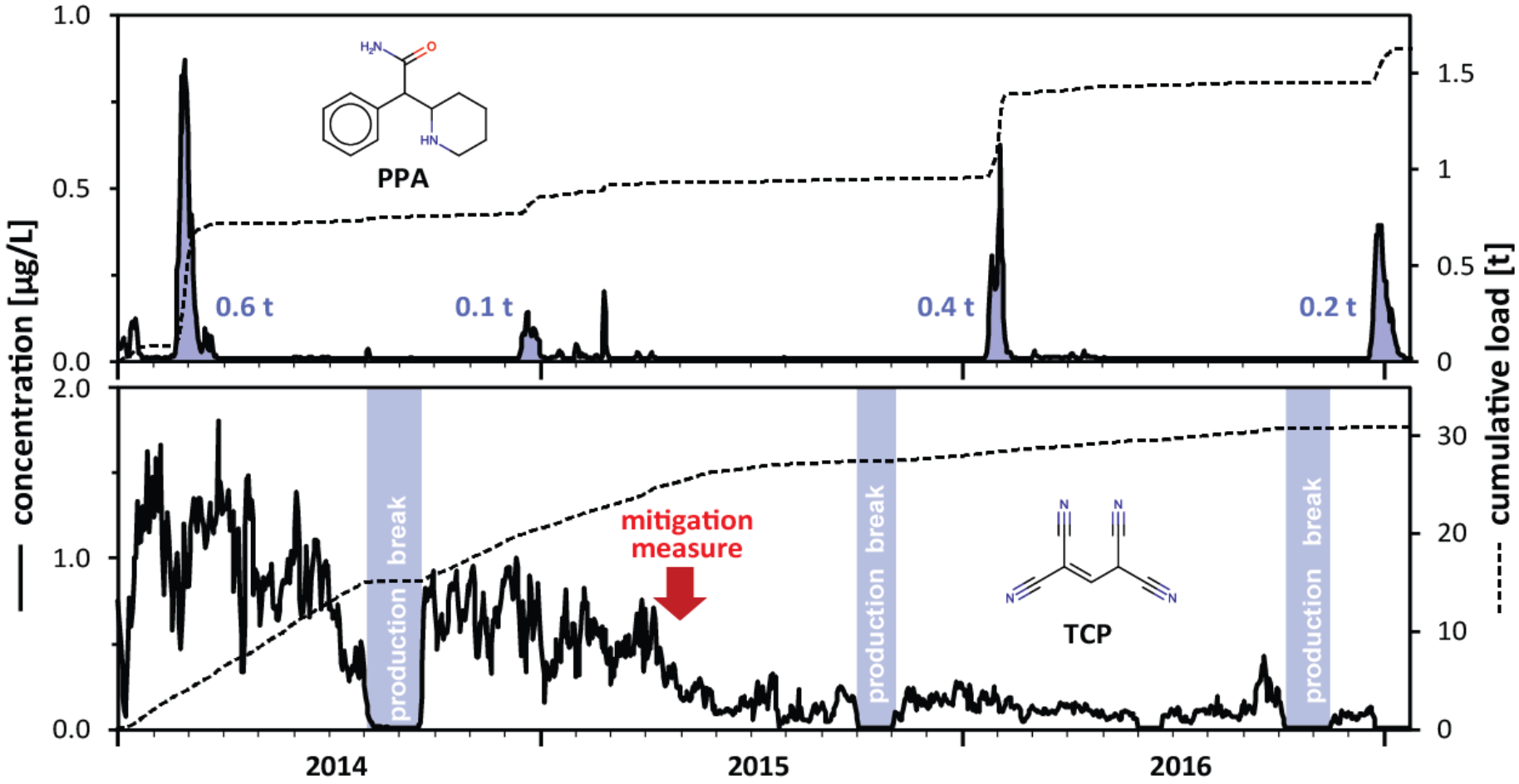


Alygizakis NA *et al* (2019) DOI: [10.1016/j.trac.2019.04.008](https://doi.org/10.1016/j.trac.2019.04.008)

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

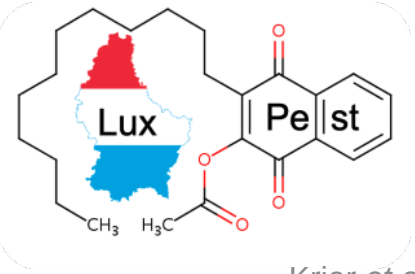
Examples: Real Time Monitoring of the Rhine River


Previously unknown chemicals detected due to “stand-out” patterns

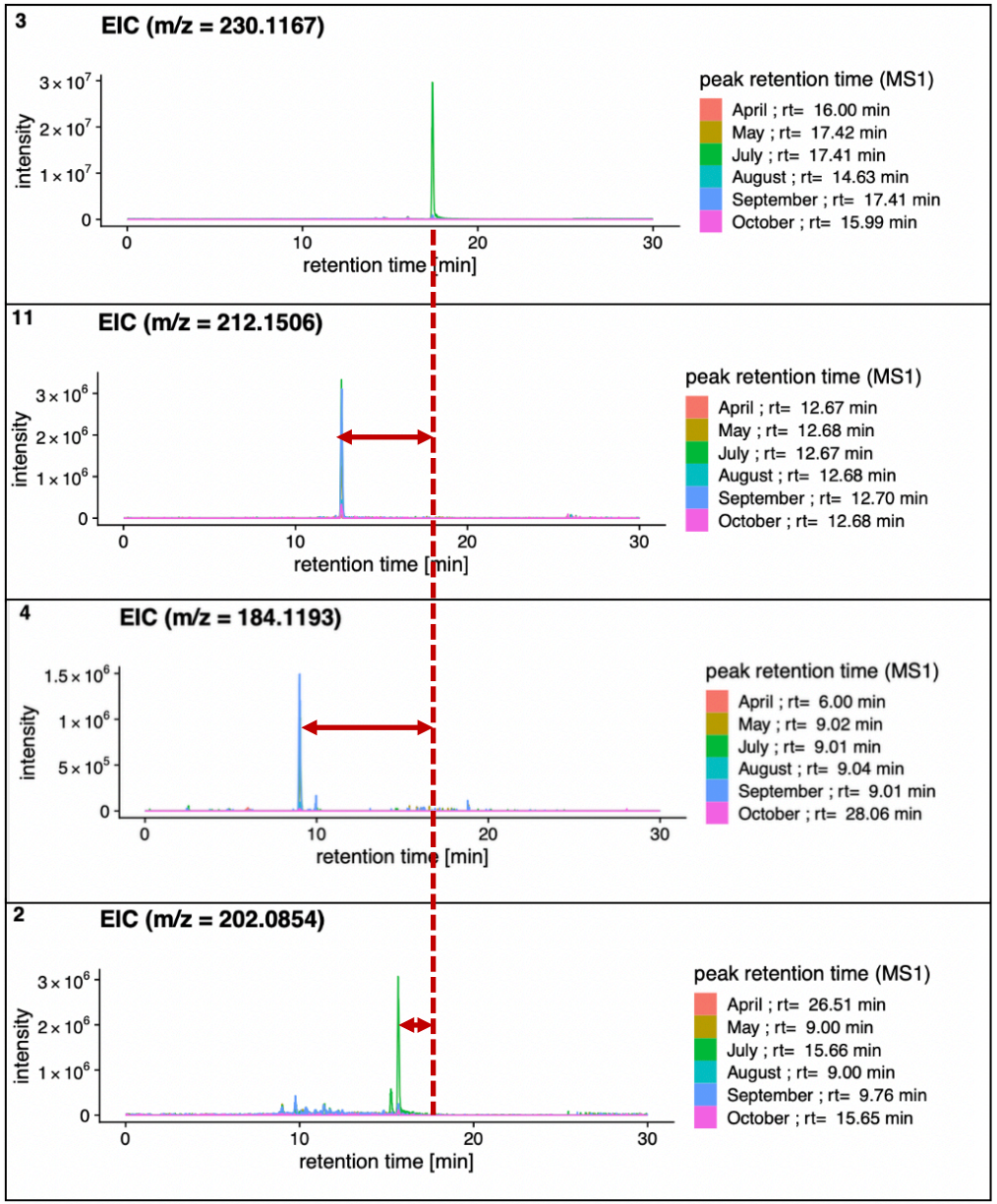


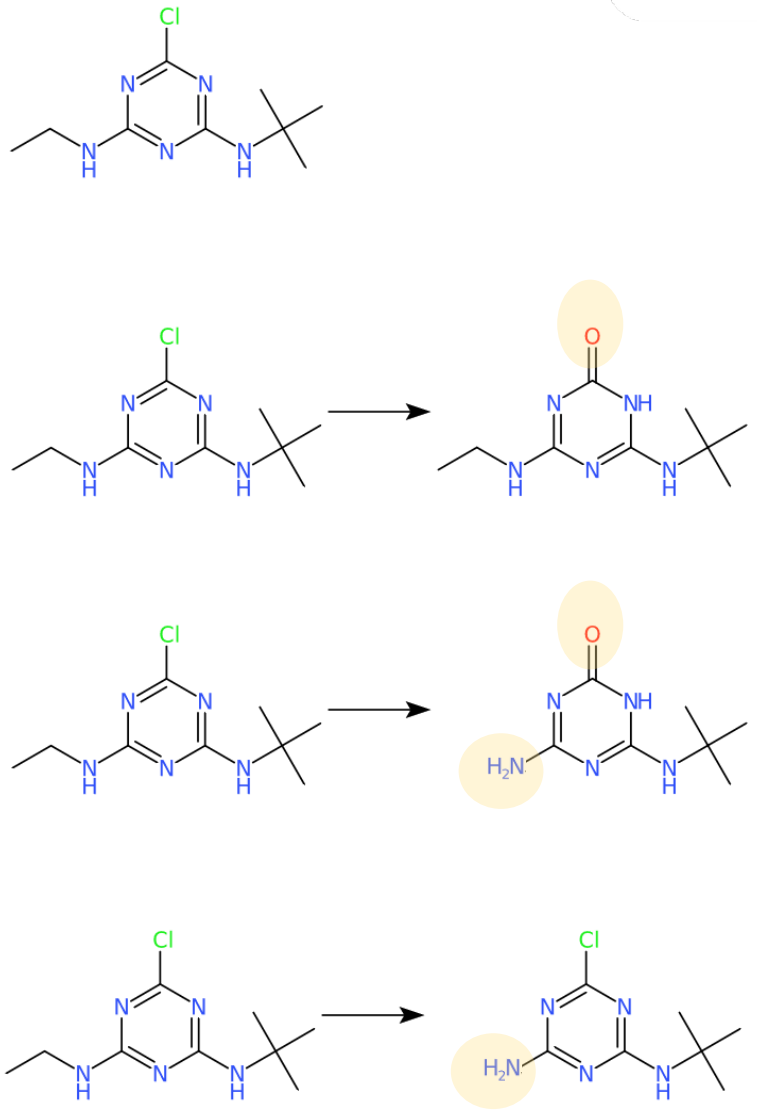
Hollender, Schymanski, Singer & Ferguson (2018) ES&T Feature, 51:20, 11505-11512. DOI: 10.1021/acs.est.7b02184

Examples: LuxPest and Transformation Products

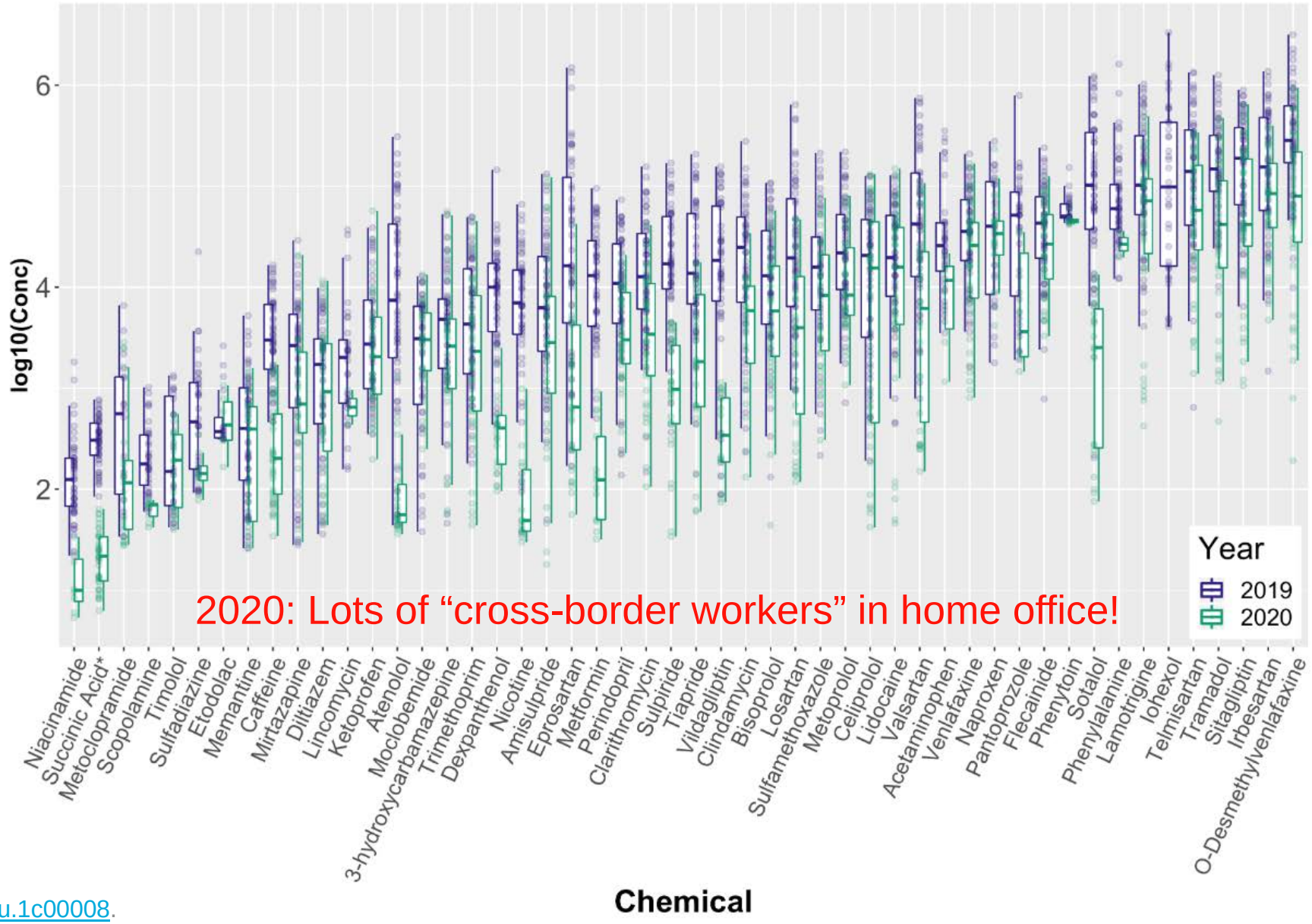


Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885





Examples: LuxPharma – 2019 versus 2020 & COVID?



Singh *et al* (2021).
DOI: [10.1021/acsenvironau.1c00008](https://doi.org/10.1021/acsenvironau.1c00008).

The Problem: Which chemicals are relevant? How to find them?

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

180 million

PubChem

111 million

ChemSpider
Search and share chemistry

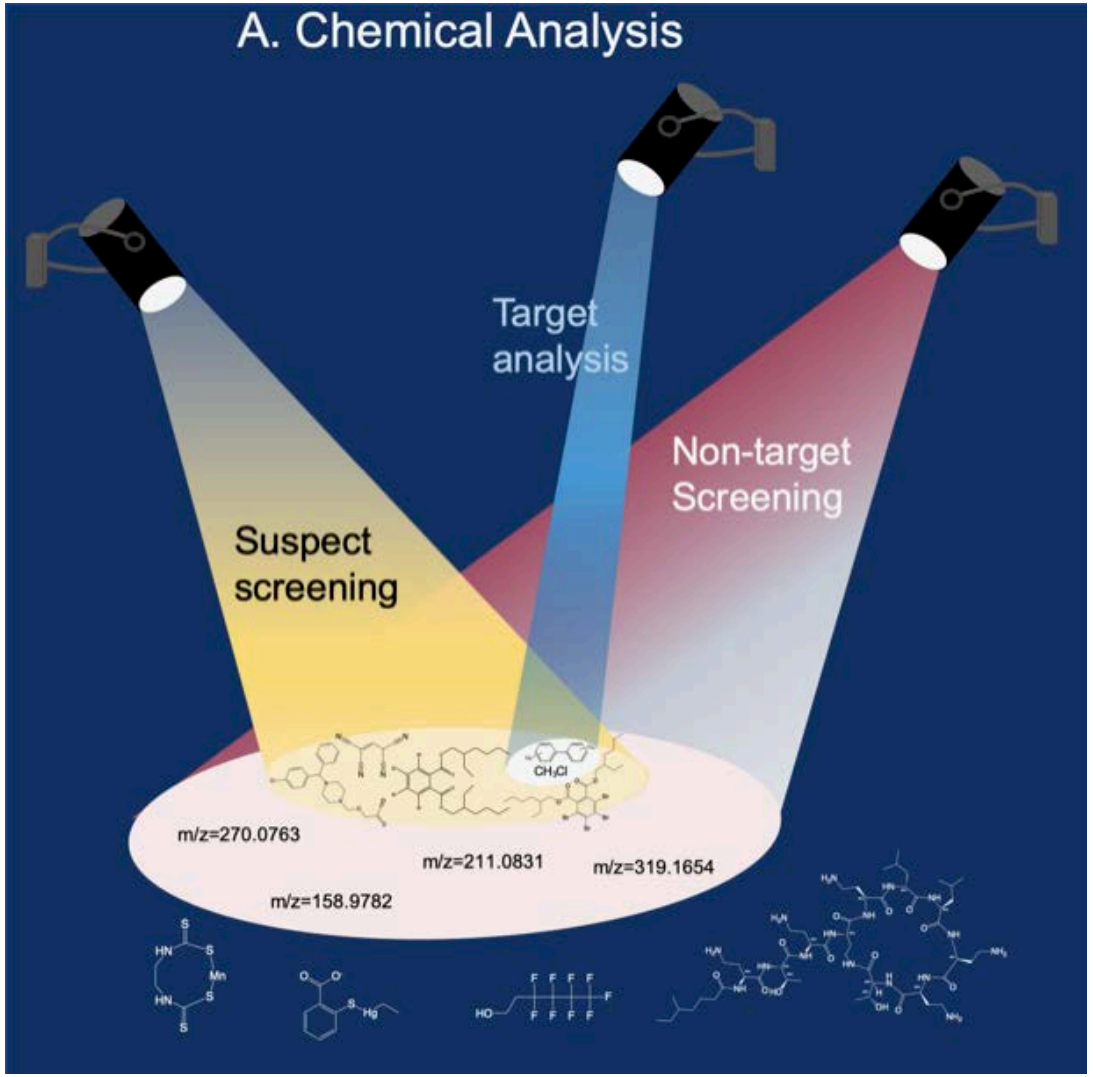
114 million

EPA Chemicals

883,000

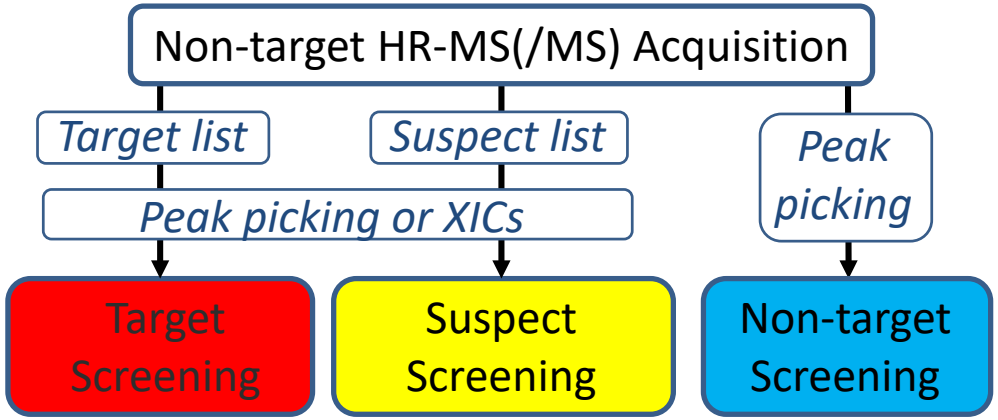
norman suspects

113,000



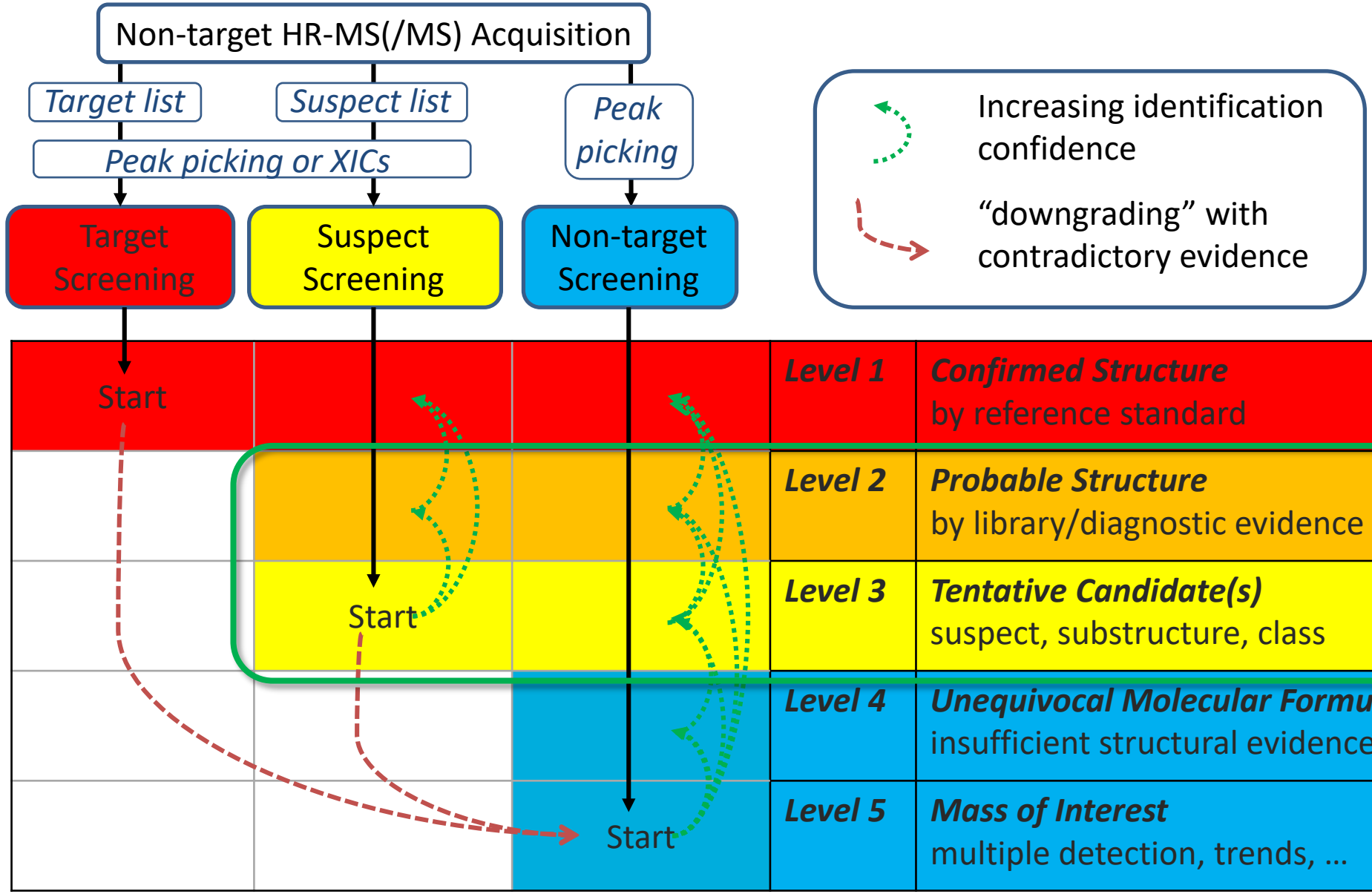
Source: ESO/IDA/Danish 1.5 m http://www.eso.org/public/images/potw1015a/ Mod. from Escher *et al* (2020). Science. DOI: 10.1126/science.aay6636

Identification Strategies and Confidence in NT-HRMS(/MS)



Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

Identification Strategies and Confidence in NT-HRMS(/MS)



MassBank
High Quality Mass Spectral Database

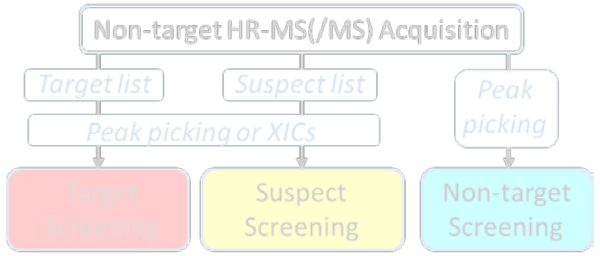
MetFrag

eawag
aquatic research

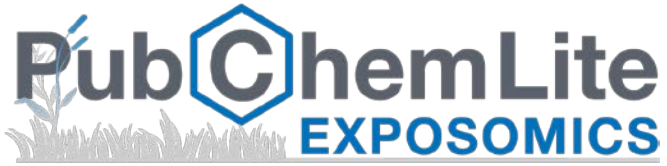
Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

Outline of Today

- Introduction and Background
  - Definitions, Key Resources
  - Examples



- Identification & Chemical Space
  - Identification + MetFrag
  - PubChemLite for Exposomics



- Case Study: LuxPest

- Take-home messages!

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

Introduction to MetFrag

https://msbi.ipb-halle.de/MetFrag/

mz [M-H]⁻
213.9637
± 5 ppm

5 ppm
0.001 Da

PubChem

Ranked Candidates

MS/MS

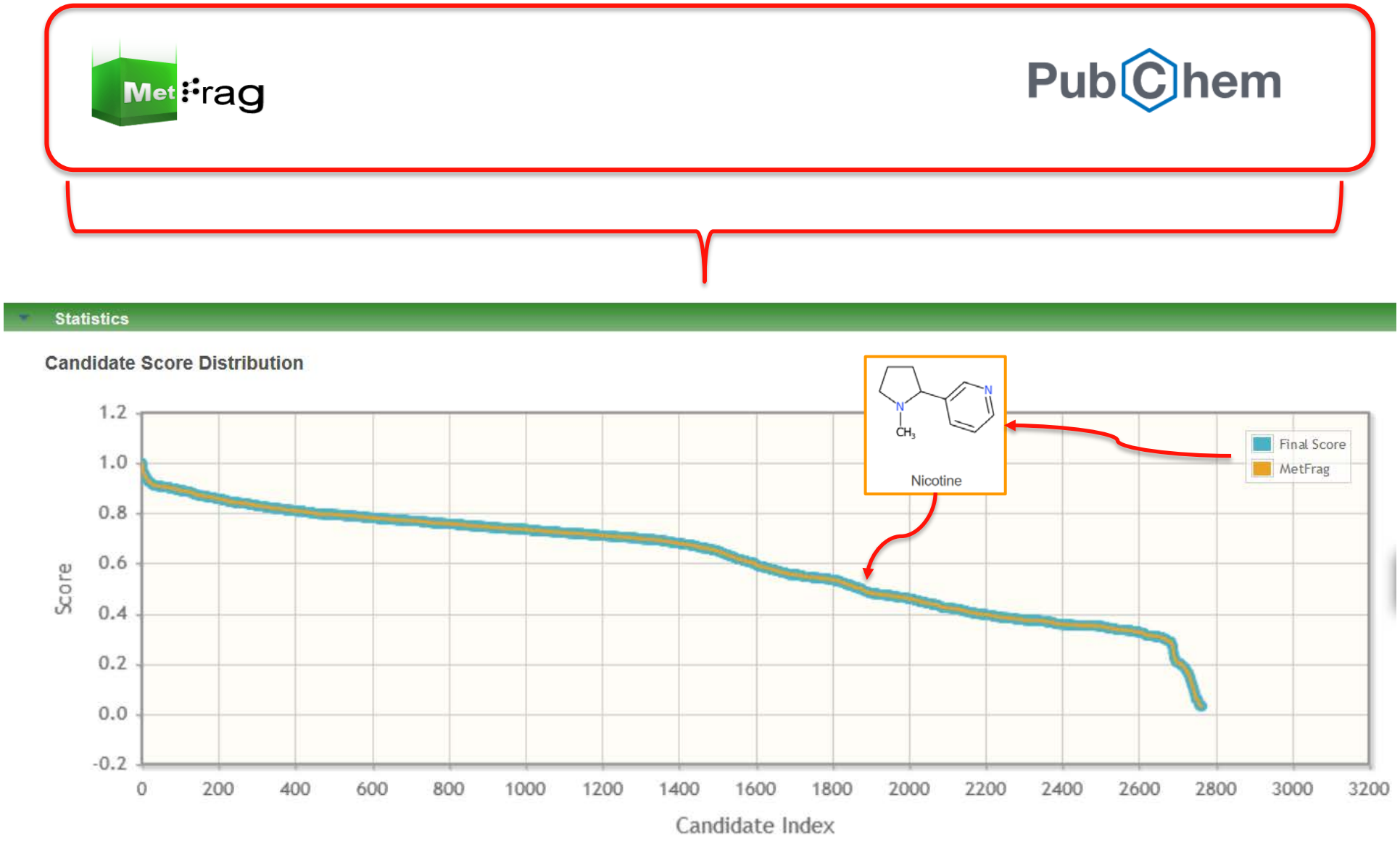
134.0054	339689
150.0001	77271
213.9607	632466

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

Key Challenge: MS and MS/MS alone is not enough!

Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9 Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Key Challenge: MS and MS/MS alone is not enough!

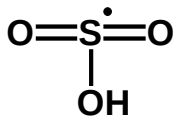


MetFrag + PubChem + Formula Search + https://massbank.eu/MassBank/RecordDisplay?id=EQ300804

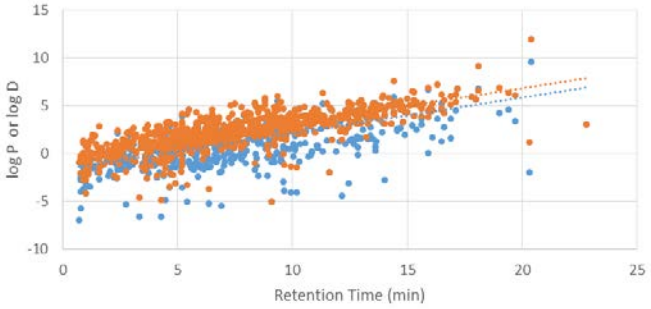
Status Quo in 2016: MetFrag Relaunched …

mz [M-H]⁻
213.9637
± 5 ppm

Elements: C,N,S
5 ppm
0.001 Da



RT: 4.54 min
355 InChI/RTs



EPA Chemicals
or
PubChem

References
Tox. Data
Data Sources
Exposure Info
MS-ready links

Literature Patents

Suspect Lists

TOFF IDENT
norman suspects
EPA Chemicals

MoNA
MassBank of North America
MassBank
High Quality Mass Spectral Database

MS/MS

134.0054	339689
150.0001	77271
213.9607	632466



IPB

eawag aquatic research

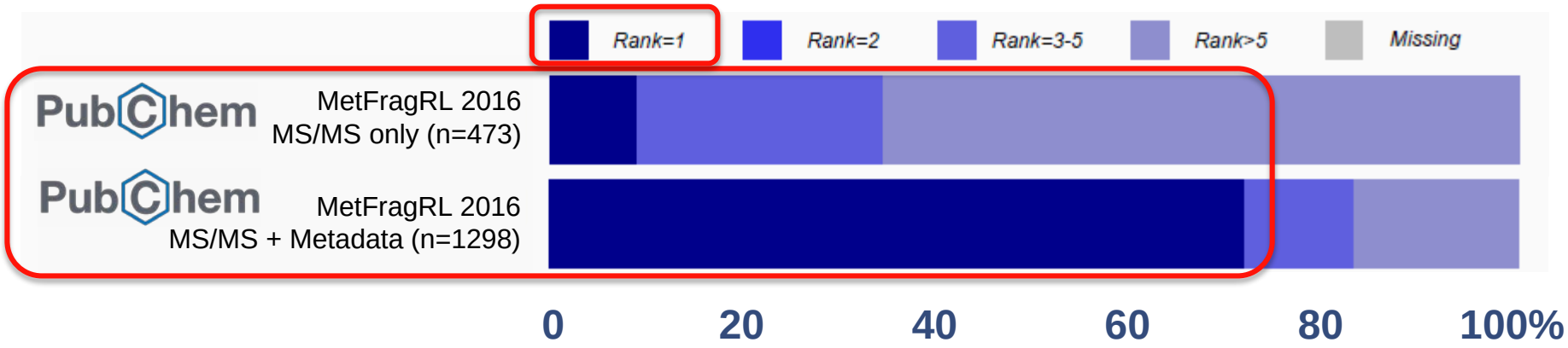
solutions

uni.lu LCSB

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

MetFragRL + PubChem + MS/MS + Metadata

- Adding literature, references & RT boosts to ~71 % rank 1!

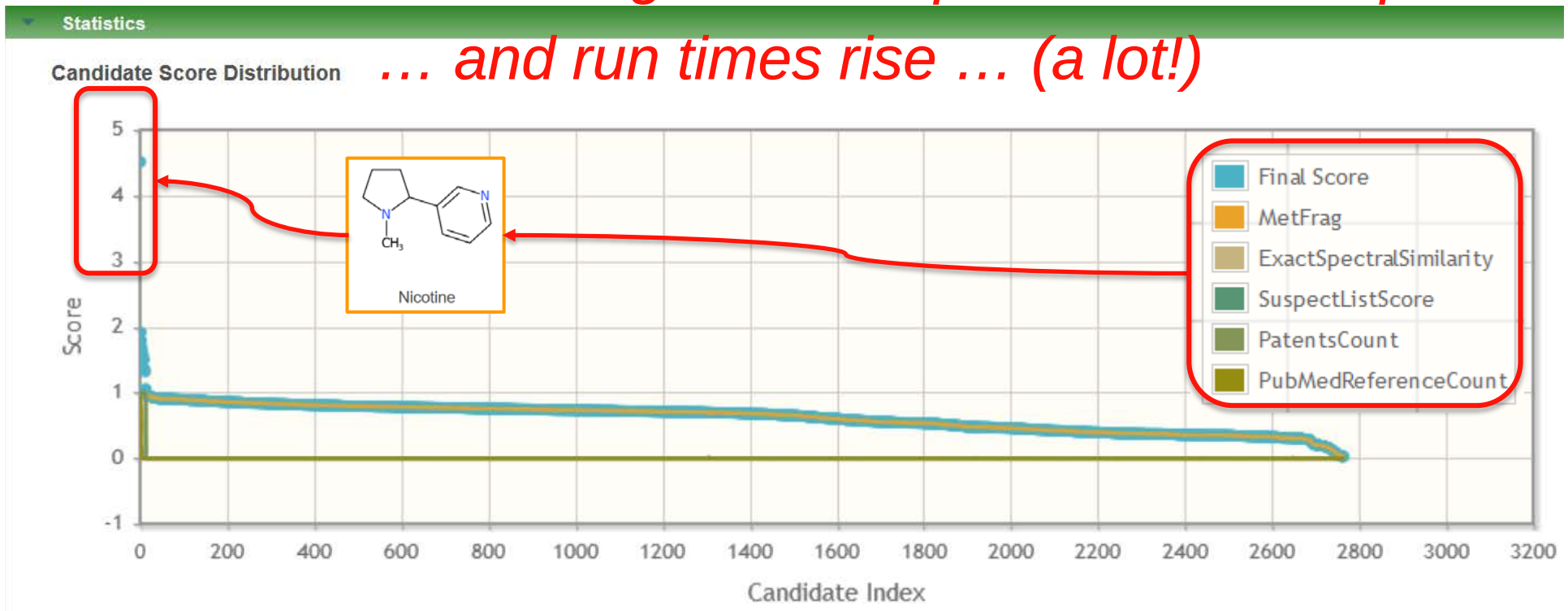


Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9 Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

MetFragRL + PubChem + MS/MS + Metadata

BUT …databases grow … ID performance drops

… and run times rise … (a lot!)



MetFrag + PubChem + Formula + MoNA + SusDat + Pat + Refs + https://massbank.eu/MassBank/RecordDisplay?id=EQ300804

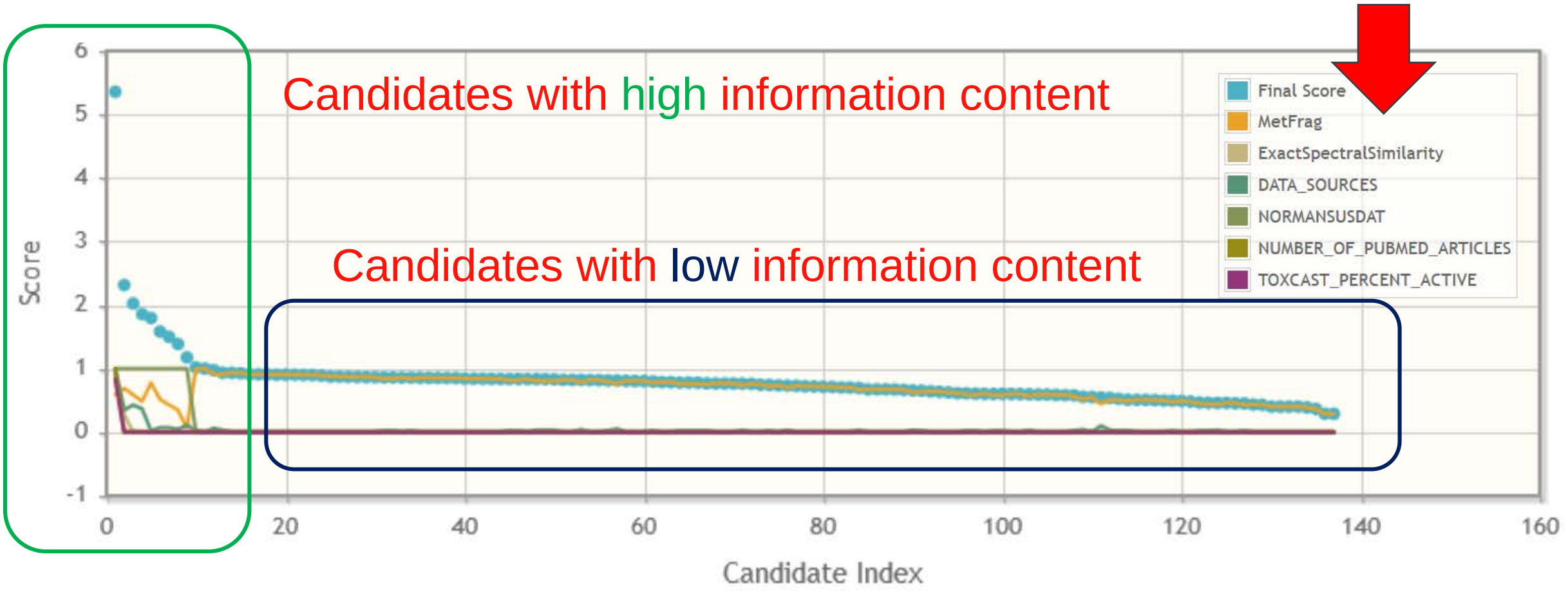
Problem: Exposomics “Chemical Space” is too big!

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY — 180 million

PubChem — 111 million

ChemSpider Search and share chemistry — 114 million

EPA Chemicals — 883,000



Can we break down PubChem into useful bits?

- PubChem Compound TOC 49,493,641
  - Agrochemical Information 3,045
  - Associated Disorders and Diseases 20,847
  - Biologic Description 2,056,521
  - Biological Test Results 3,622,920
  - Biomolecular Interactions and Pathways 125,253
  - Chemical and Physical Properties 263,015
  - Classification 1,454,824
  - Drug and Medication Information 17,922
  - Food Additives and Ingredients 8,414
  - Identification 4,968
  - Information Sources 20,271,277
  - Literature 1,833,941
  - Names and Identifiers 1,275,170
  - Patents 36,351,418
  - Pharmacology and Biochemistry 110,628
  - Related Records 9,224,590
  - Safety and Hazards 149,319
  - Spectral Information 480,730
  - Structures 9,117,635
  - Toxicity 114,012
  - Use and Manufacturing 115,321
  - Chemical Safety 147,023

- Agrochemical Information 3,045
  - Agrochemical Category 1,903
  - Agrochemical Transformations 1,383
  - EU Pesticides Data 1,213
  - USDA Pesticide Data Program 579
- Use and Manufacturing 115,321
  - Uses 27,314
  - Consumption Patterns 1,168
  - Formulations/Preparations 4,742
  - General Manufacturing Information 52,353
  - Impurities 870
  - Methods of Manufacturing 6,192
  - Overview 1,179
  - Sampling Procedures 579
  - U.S. Exports 1,038
  - U.S. Imports 1,547
  - U.S. Production 5,500
  - Use Classification 63,558

PubChem Furathiocarb (Compound) CONTENTS

7 Agrochemical Information

7.1 Agrochemical Category

Insecticides

- EU Pesticides Database

7.2 Agrochemical Transformations

Furathiocarb has known environmental transformation products that include carbofuran.

S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 | DOI:10.5281/zenodo.3544759

- NORMAN Suspect List Exchange

7.3 EU Pesticides Data

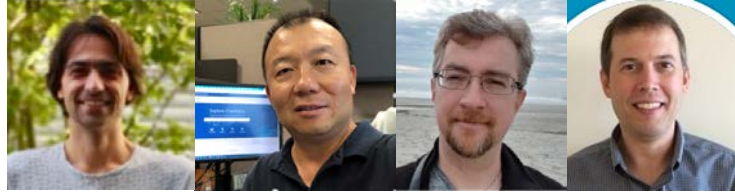
Active Substance	furathiocarb
Status	Not Approved [Reg. (EC) No 1107/2009]
Categories	Insecticides

https://pubchem.ncbi.nlm.nih.gov/compound/Furathiocarb#section=Agrochemical-Information

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

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

Introducing …

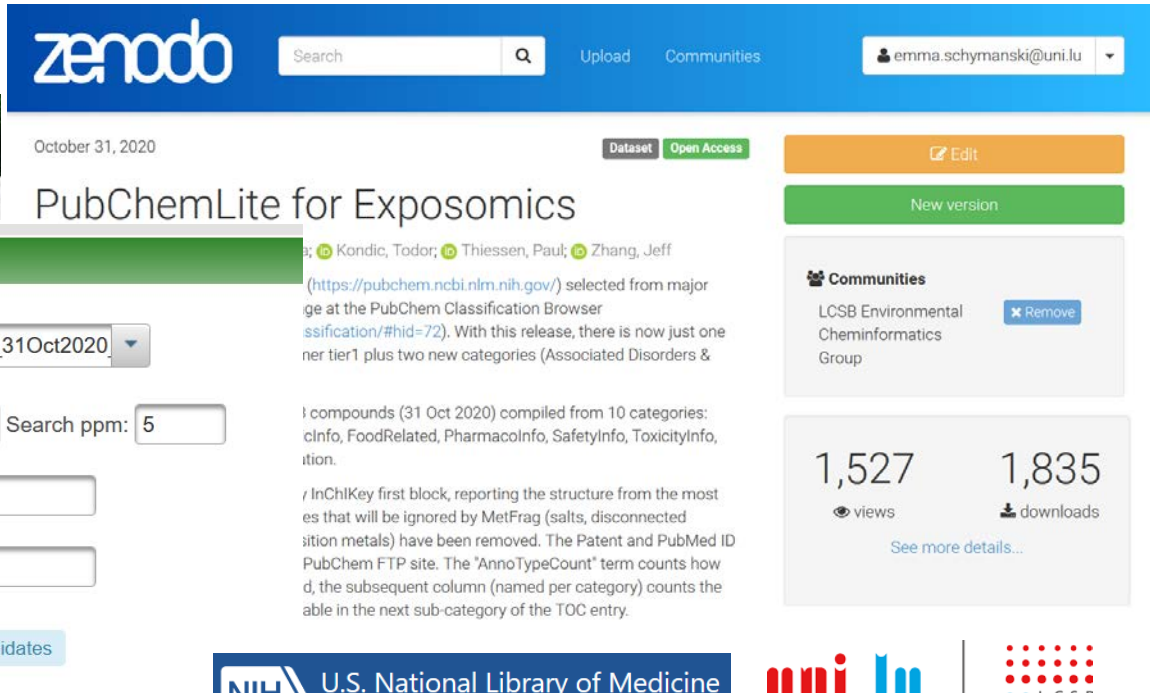


- PubChem Compound TOC ? 49,493,641
  - Agrochemical Information ? 3,045
  - Associated Disorders and Diseases ? 20,847
  - Biologic Description ? 2,056,521
  - Biological Test Results ? 3,622,920
  - Biomolecular Interactions and Pathways ? 125,253
  - Chemical and Physical Properties ? 263,015
  - Classification ? 1,454,824
  - Drug and Medication Information ? 17,922
  - Food Additives and Ingredients ? 8,414
  - Identification ? 4,968
  - Information Sources ? 20,271,277
  - Literature ? 1,833,941
  - Names and Identifiers ? 1,275,170
  - Patents ? 36,351,418
  - Pharmacology and Biochemistry ? 110,628
  - Related Records ? 9,224,590
  - Safety and Hazards ? 149,319
  - Spectral Information ? 480,730
  - Structures ? 9,117,635
  - Toxicity ? 114,012
  - Use and Manufacturing ? 115,321

Chemical Safety ? 147,023

PubChemLite EXPOSOMICS

~370,000 entries “small”



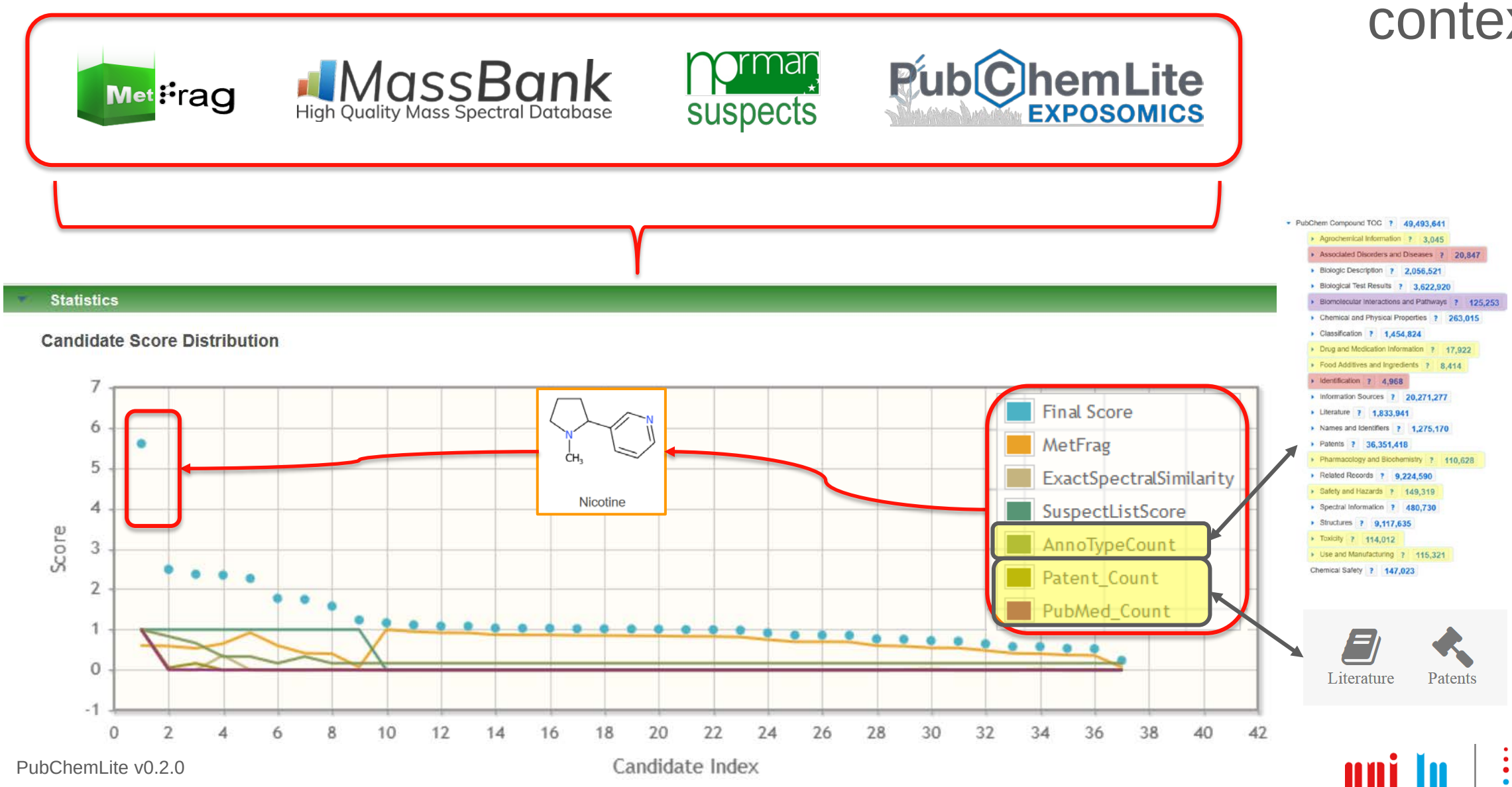
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

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

uni.lu

LCSB

PubChemLite: Fewer and more relevant candidates – with context

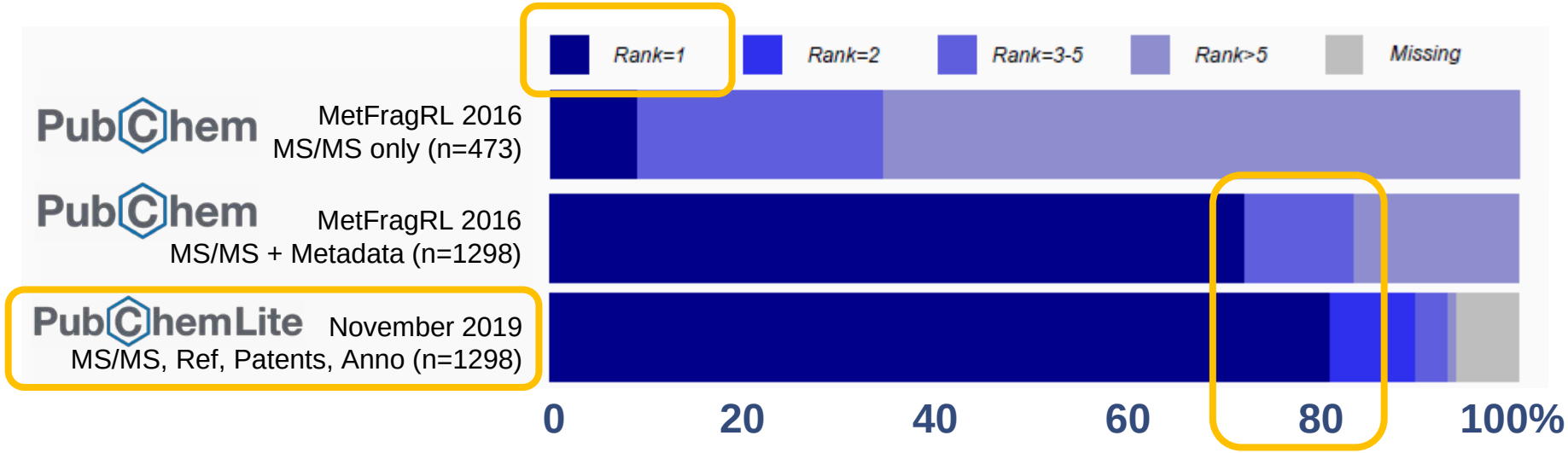


PubChemLite v0.2.0

MetFrag+PubChemLite+Formula+MoNA+SusDat+Pat+Refs+Anno + https://massbank.eu/MassBank/RecordDisplay?id=EQ300804

How does PubChemLite perform?

- ~110 M => ~370 K … how does this influence performance?

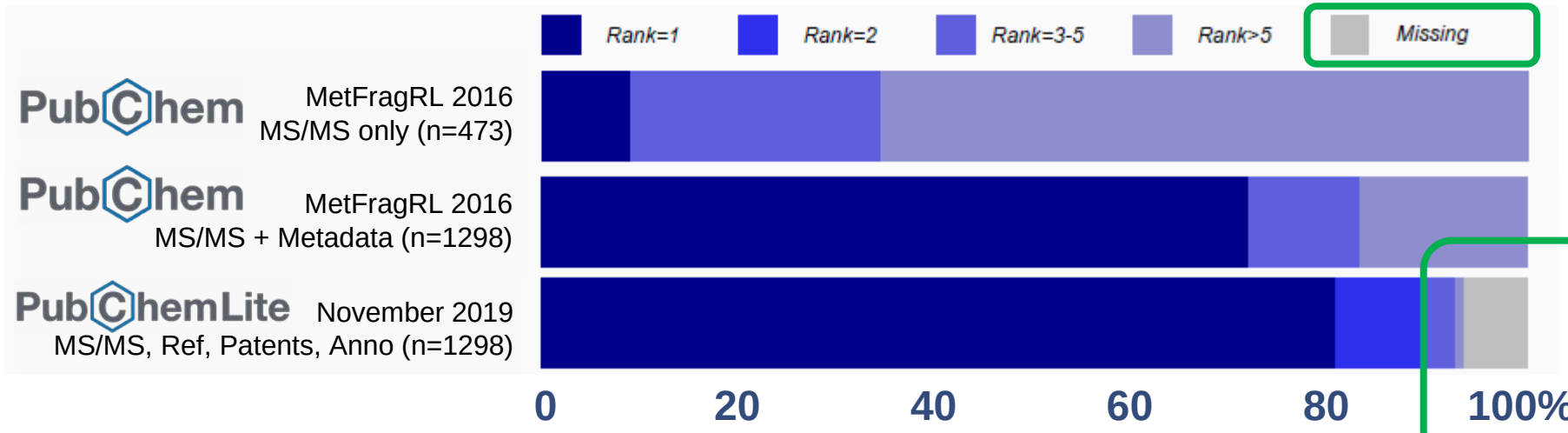


>80 % ranked in first place
~90 % ranked first or second!

Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9 Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

How does PubChemLite perform?

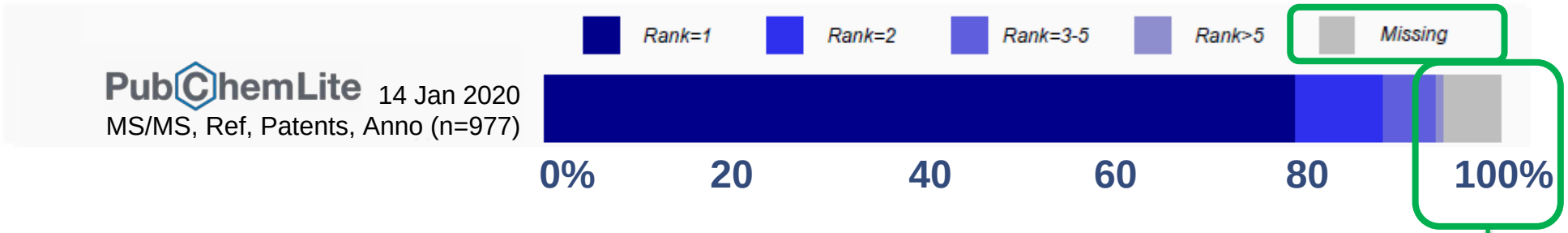
- ~110 M => ~370 K … how does this influence performance?



"Transformation products are missing from databases"
[General complaint; environmental community]

Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9 Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Assessing the Missing Entries in PubChemLite



- NORMAN Suspect List Exchange Classification ? 117,037
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and
  - S25 | OECDPFAS | List of PFAS from the OECD ? 3,680
  - S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 647
  - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites ? 1,358
  - S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574
  - S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258
  - S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem ? 97

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

Transformation Products: Filling the Data Gaps!

PubChem NORMAN Suspect List Exchange

- NORMAN Suspect List Exchange Classification ? 113,080
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 3,85
  - S25 | OECDPFAS | List of PFAS from the OECD ? 3,677
  - S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances ? 254
  - S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 885
  - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ? 1,343
  - S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574
  - S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258
  - S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem ? 102
  - S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386
  - S72 | NTUPHTW | Pharmaceutically Active Substances from National Taiwan University ? 1,068
  - S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria ? 2,088
  - S00 | SUSDAT | Merged NORMAN Suspect List: SusDat ? 99,130
  - S01 | MASSBANK | NORMAN Compounds in MassBank EU ? 7,164
  - S02 | STOFFIDENT | HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances ? 11,261
  - S03 | NORMANCT15 | NORMAN Collaborative Trial Targets and Suspects ? 624
  - S04 | UJIBADE | Target List from UJI used in Bade et al 2015 ? 542

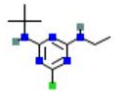
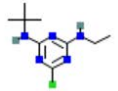
- Pharmacology and Biochemistry ? 112,039
  - Human Metabolite Information ? 64,199
  - Metabolism/Metabolites ? 8,204
  - Transformations ? 5,857

PubChem Terbuthylazine (Compound)

8.5 Transformations

Page 3 of 25 items View More Rows & Details Download

SORT BY Please Choose One

Predecessor Image	Predecessor Name	Transformation	Successor Image	Successor Name	Evidence DOI
	Terbuthylazine	Mammalian metabolism		6-Chloro-1,3,5-triazine-2,4-diamine	10.5281/zenodo.382
	Terbutylazine	Deethylation		Terbutylazine-desethyl	10.1007/s13361-017-

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

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

Transformation Products: Filling the Data Gaps!

PubChem Terbuthylazine (Compound)

7 Agrochemical Information

7.1 Agrochemical Category

Pesticides -> Herbicides -> Triazine herbicides -> Chlorotriazine herbicides

S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag | DOI:10.5281/zenodo.3754448

NORMAN Suspect List Exchange

7.2 Agrochemical Transformations

Terbutylazine has known environmental transformation products that include Terbutylazine-2-hydroxy, Terbutylazine-desethyl, and Terbutylazine-desethyl-2-hydroxy.

S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag | DOI:10.5281/zenodo.3754448

NORMAN Suspect List Exchange

Terbutylazine has known environmental transformation products that include CSAA036479, CSAA04949, CSCD648241, CSCD692760, GS31398, MT1, GS 26379, MT13, GS 23158, Terbutylazine metabolite MT14, Terbutylazine metabolite MT23, and Terbutylazine metabolite MT24.

S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 | DOI:10.5281/zenodo.3544759

NORMAN Suspect List Exchange

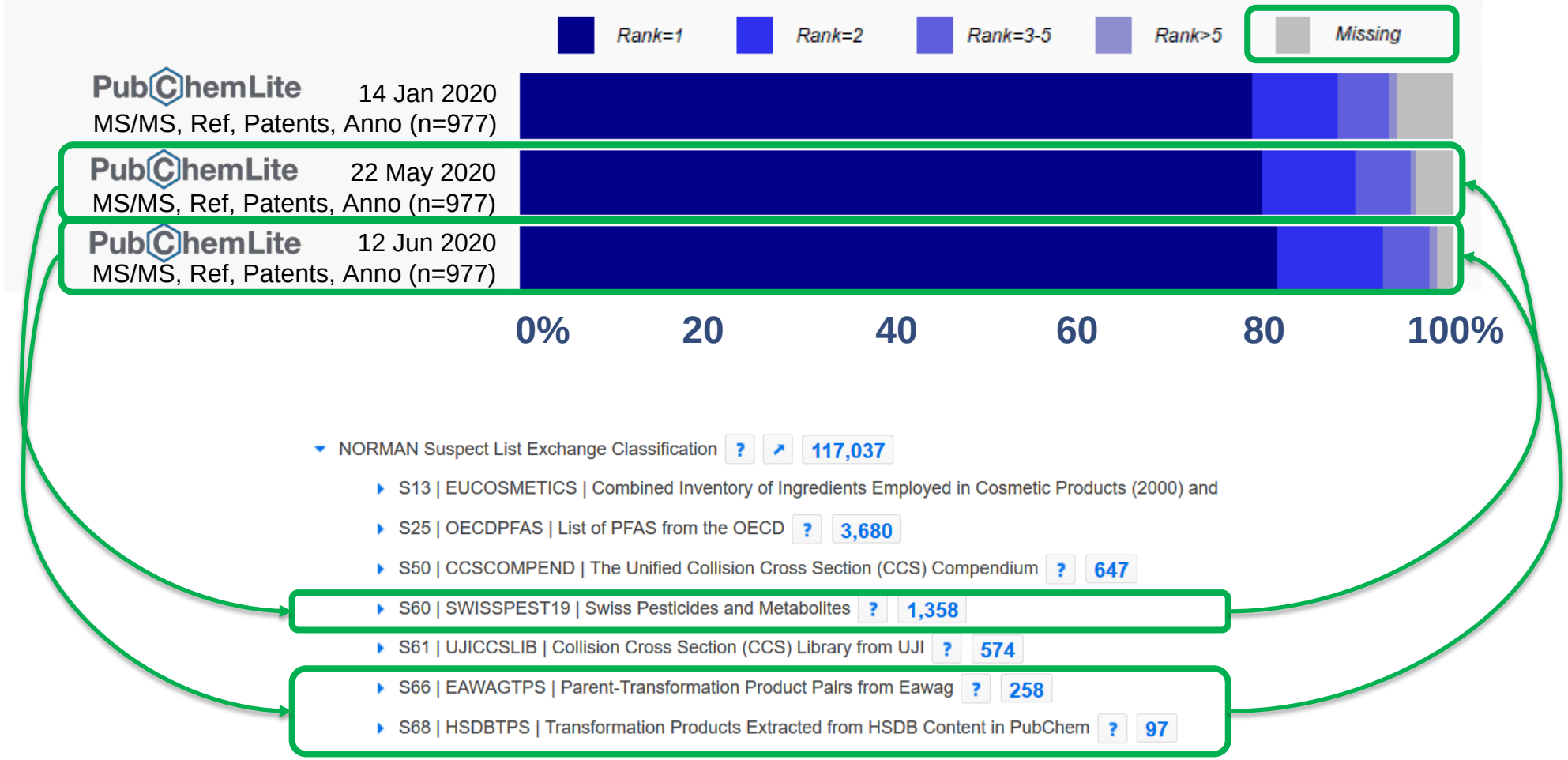
- Agrochemical Information ? 3,045
  - Agrochemical Category ? 1,903
  - Agrochemical Transformations ? 1,383
  - EU Pesticides Data ? 1,213
  - USDA Pesticide Data Program ? 579

PubChemLite EXPOSOMICS

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

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

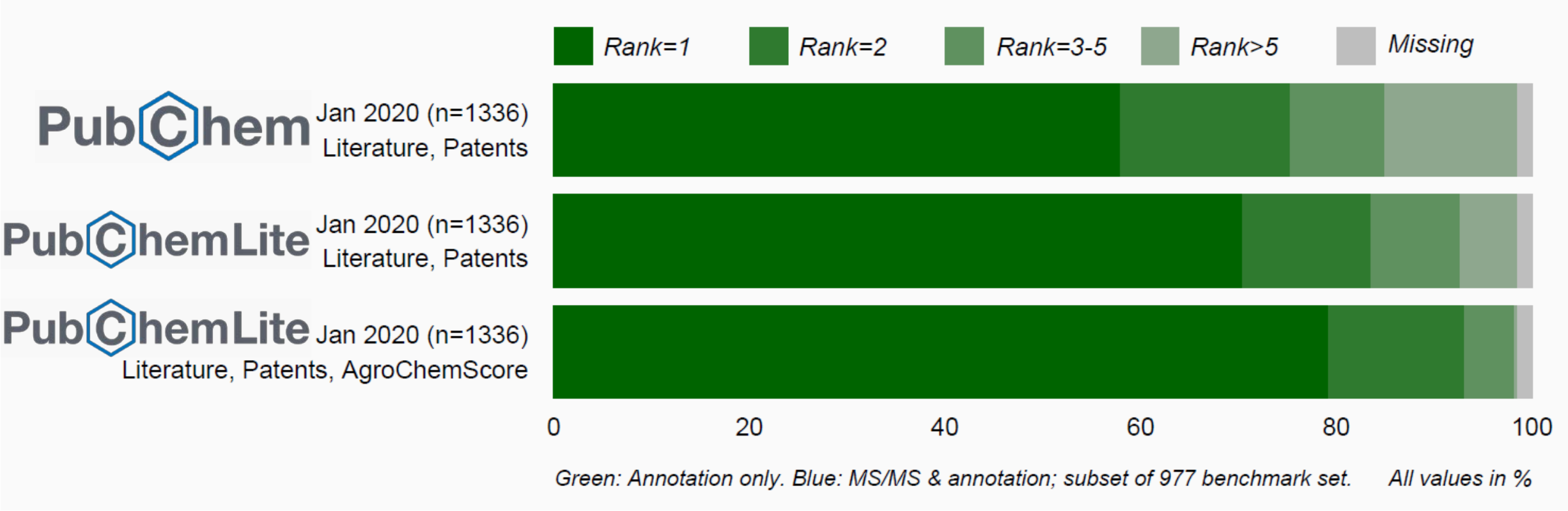
Assessing the Missing Entries in PubChemLite



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

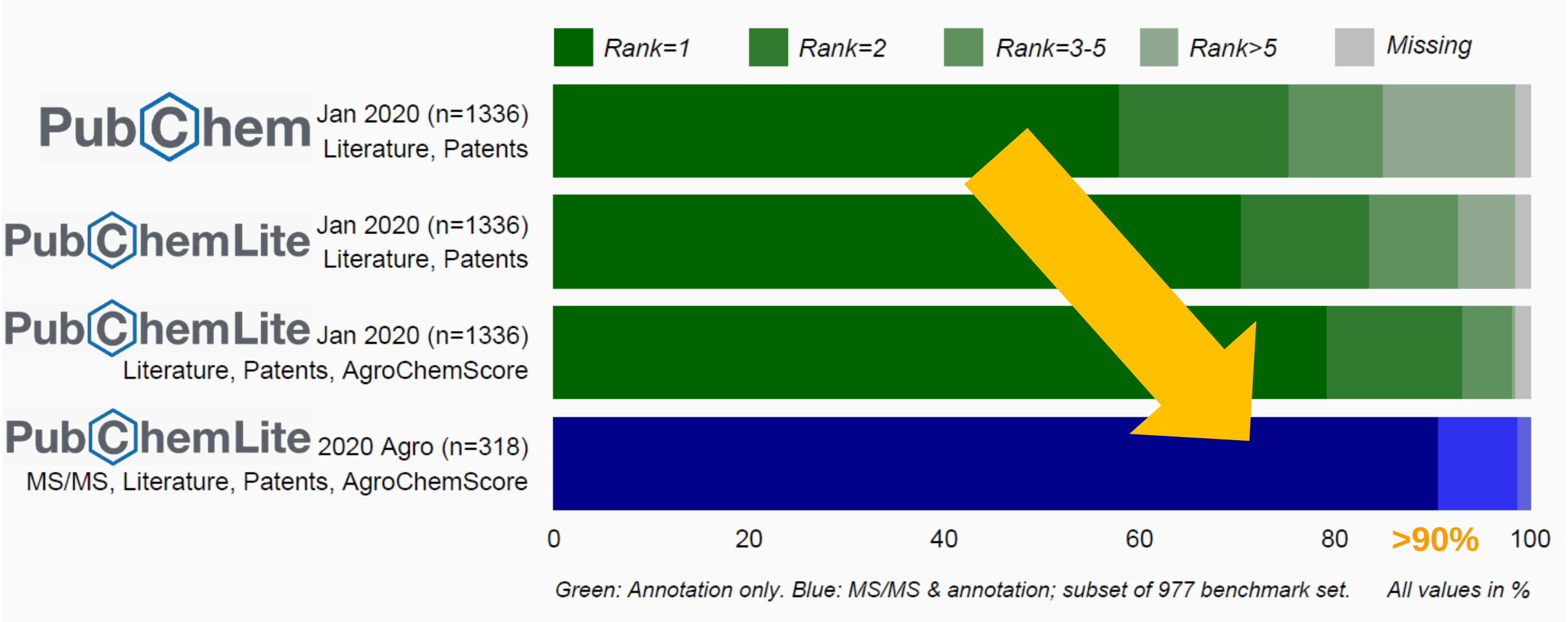
NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Influence of the Annotation Content in PubChemLite



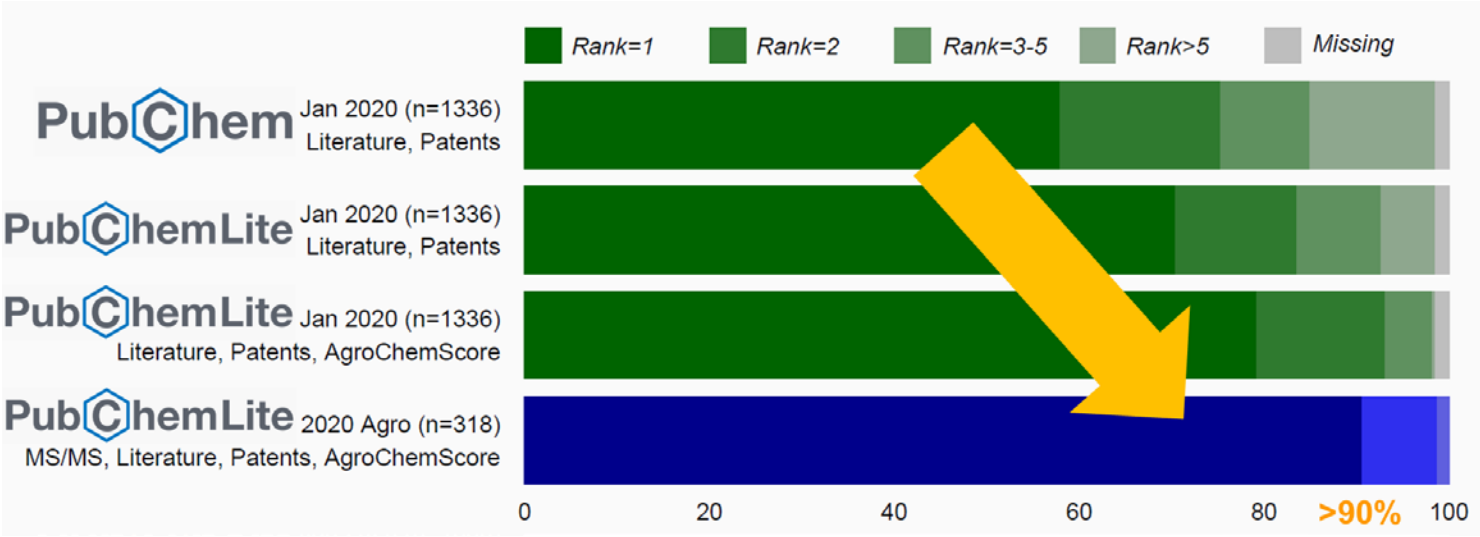
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Influence of the Annotation Content in PubChemLite



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

Influence of the Annotation Content in PubChemLite



>90% of (well known) entries are Top 1 rank

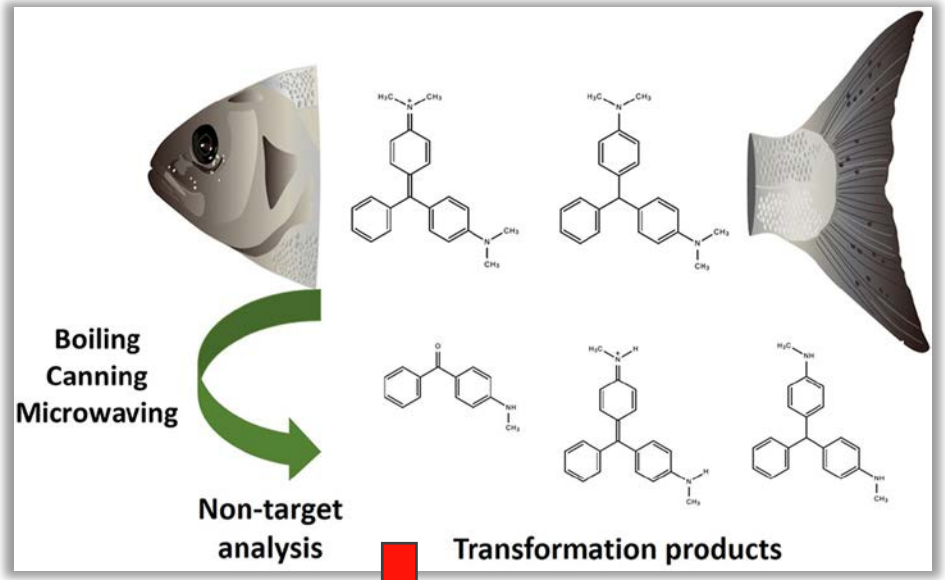
>97% are Top 2 … 100 % Top 3 …

Emerging Contaminants: Top 1-3 matches will cover a lot!

(and quickly!)

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

What about *new* Emerging Contaminants? Add them in!



zenodo Search Upload Communities

October 8, 2021 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.

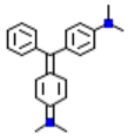
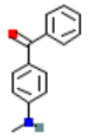
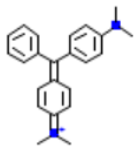
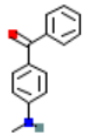
https://www.norman-network.com/nds/SLE/

PubChem 4-(Methylamino)benzophenone (Compound)

7.1 Transformations

2 items View More Details Download

SORT BY Please Choose One

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enz
	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

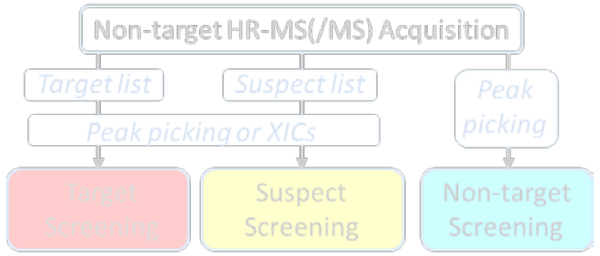
Source: NORMAN Suspect List Exchange

Record Name: 4-(Methylamino)benzophenone

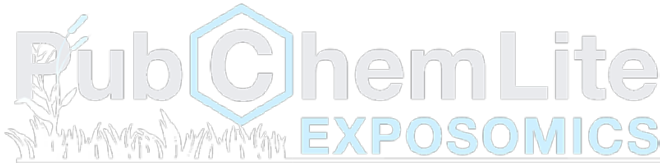
Baesu *et al*. (2021) DOI: 10.1016/j.crfs.2021.09.010. S74 | REFTPS (2021) DOI: 10.5281/zenodo.4318838

Outline of Today

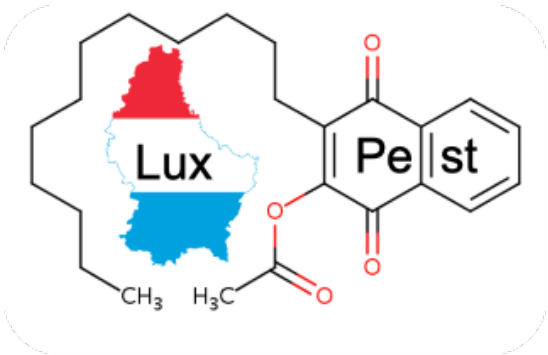
- Introduction and Background
  - Definitions, Key Resources
  - Examples



- Identification & Chemical Space
  - Identification + MetFrag
  - PubChemLite for Exposomics



- Case Study: LuxPest



- Take-home messages!

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

The Problem: Which chemicals are relevant? How to find them?

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

180 million

PubChem

111 million

ChemSpider
Search and share chemistry

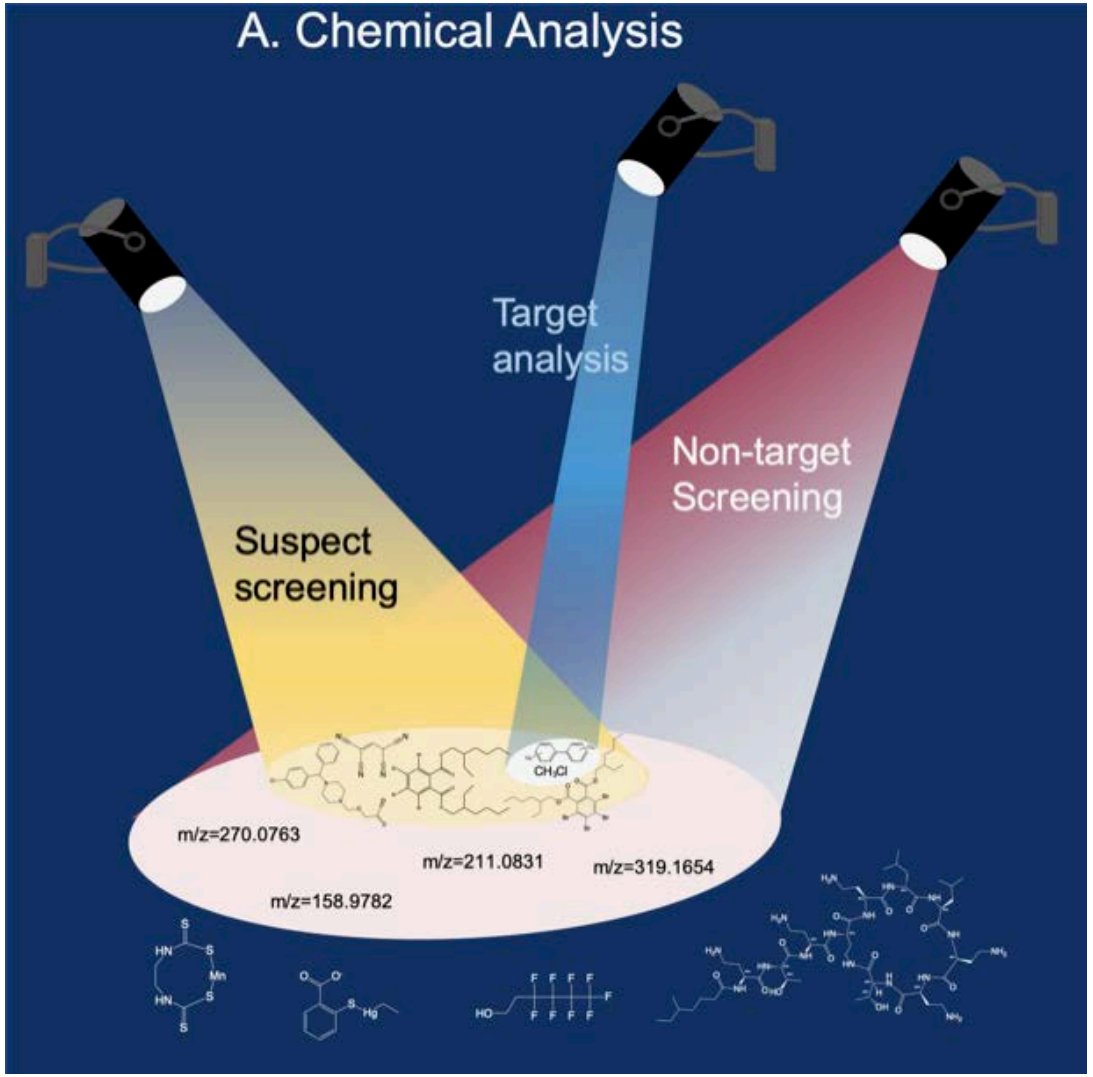
114 million

EPA Chemicals

883,000

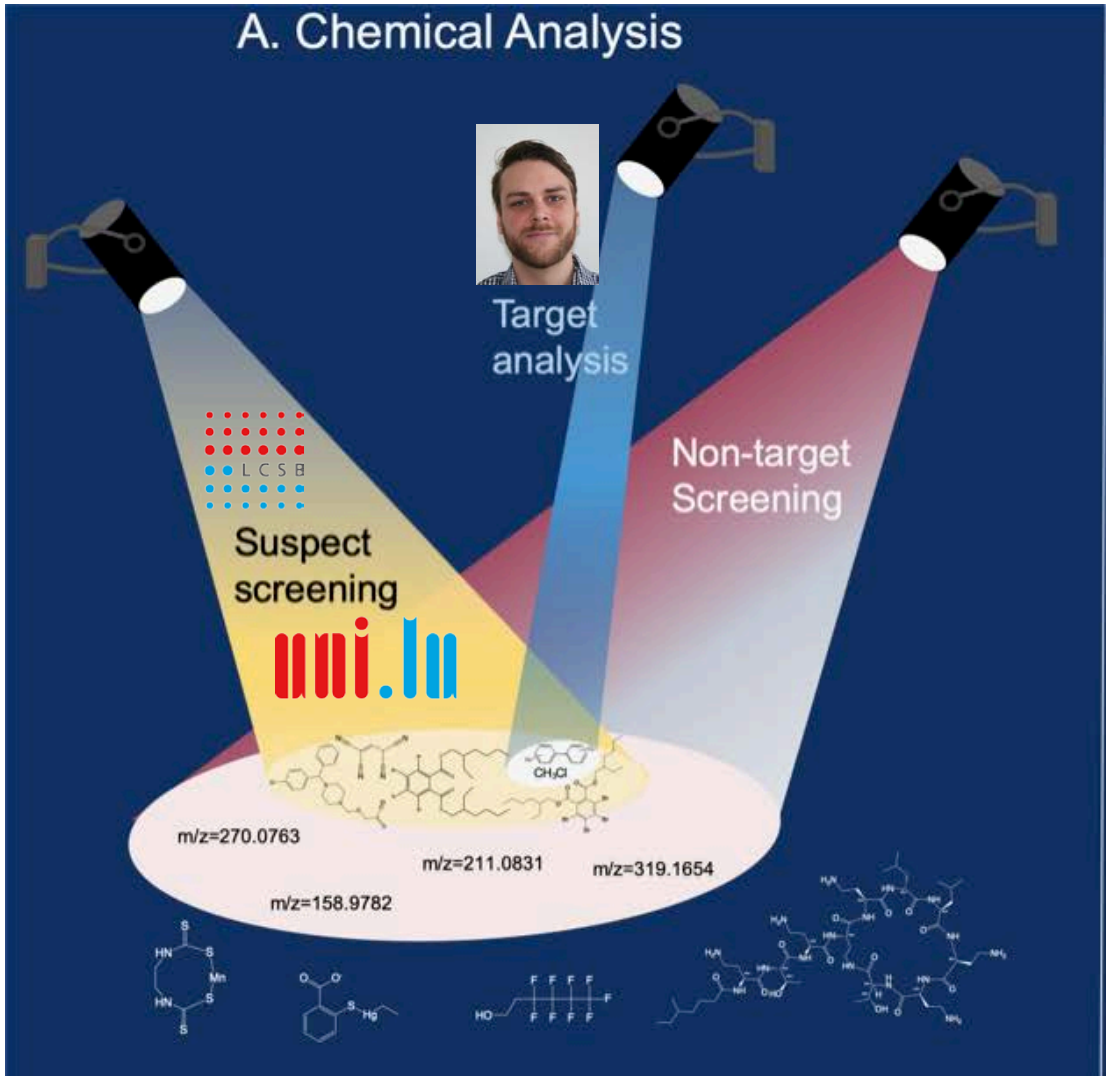
norman suspects

113,000



Source: ESO/IDA/Danish 1.5 m http://www.eso.org/public/images/potw1015a/ Mod. from Escher *et al* (2020). Science. DOI: 10.1126/science.aay6636

The Problem: Which chemicals are relevant? How to find them?

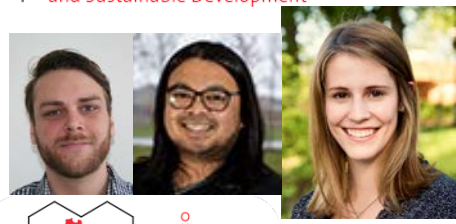


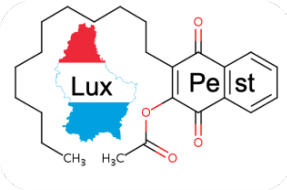
Monthly sampling
8-9 locations / year
4 fixed, 4-5 rotate
3 year cycle

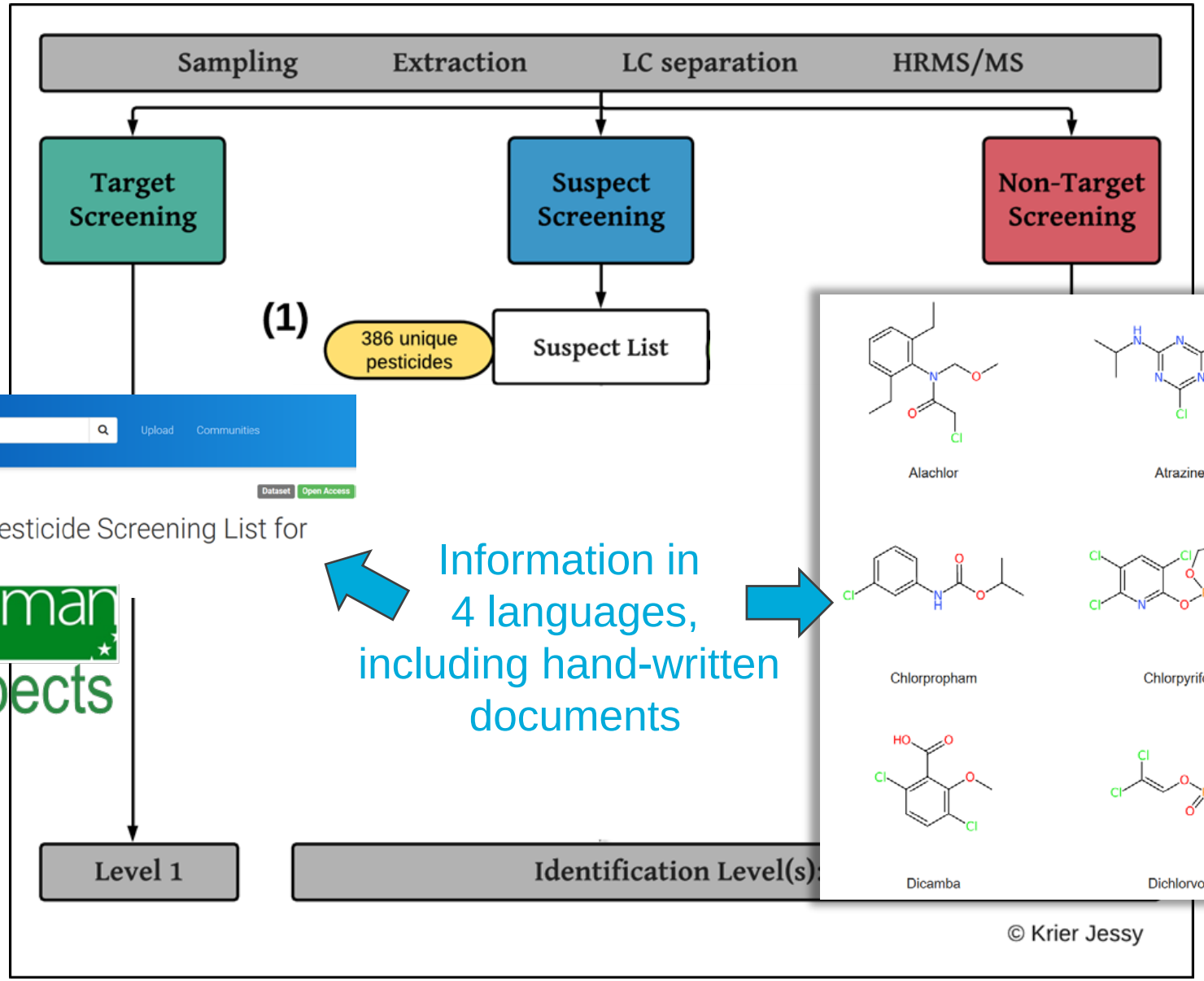
Modified from Escher, Stapleton, Schymanski (2020). Science. DOI: 10.1126/science.aay6636. Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885.

LuxPest – Suspect List Generation

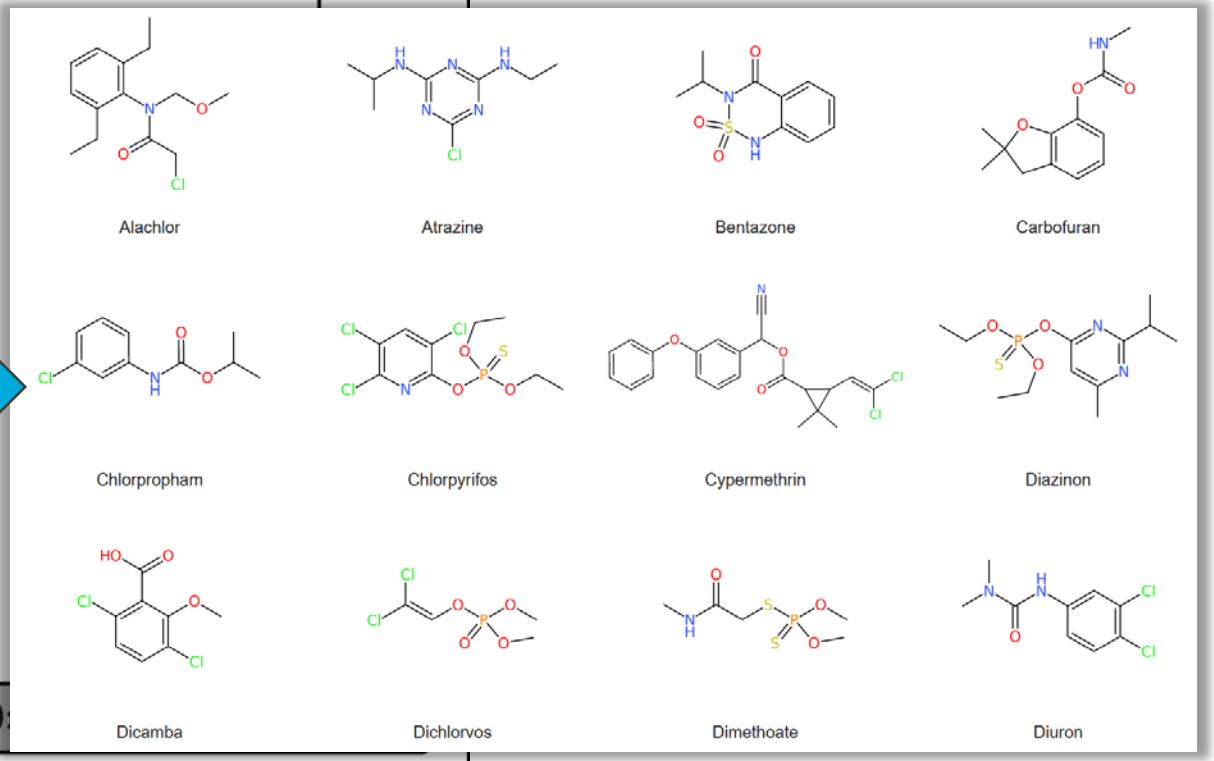
THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development







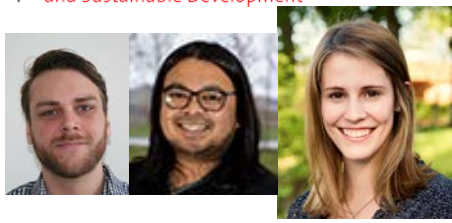
Information in 4 languages, including hand-written documents

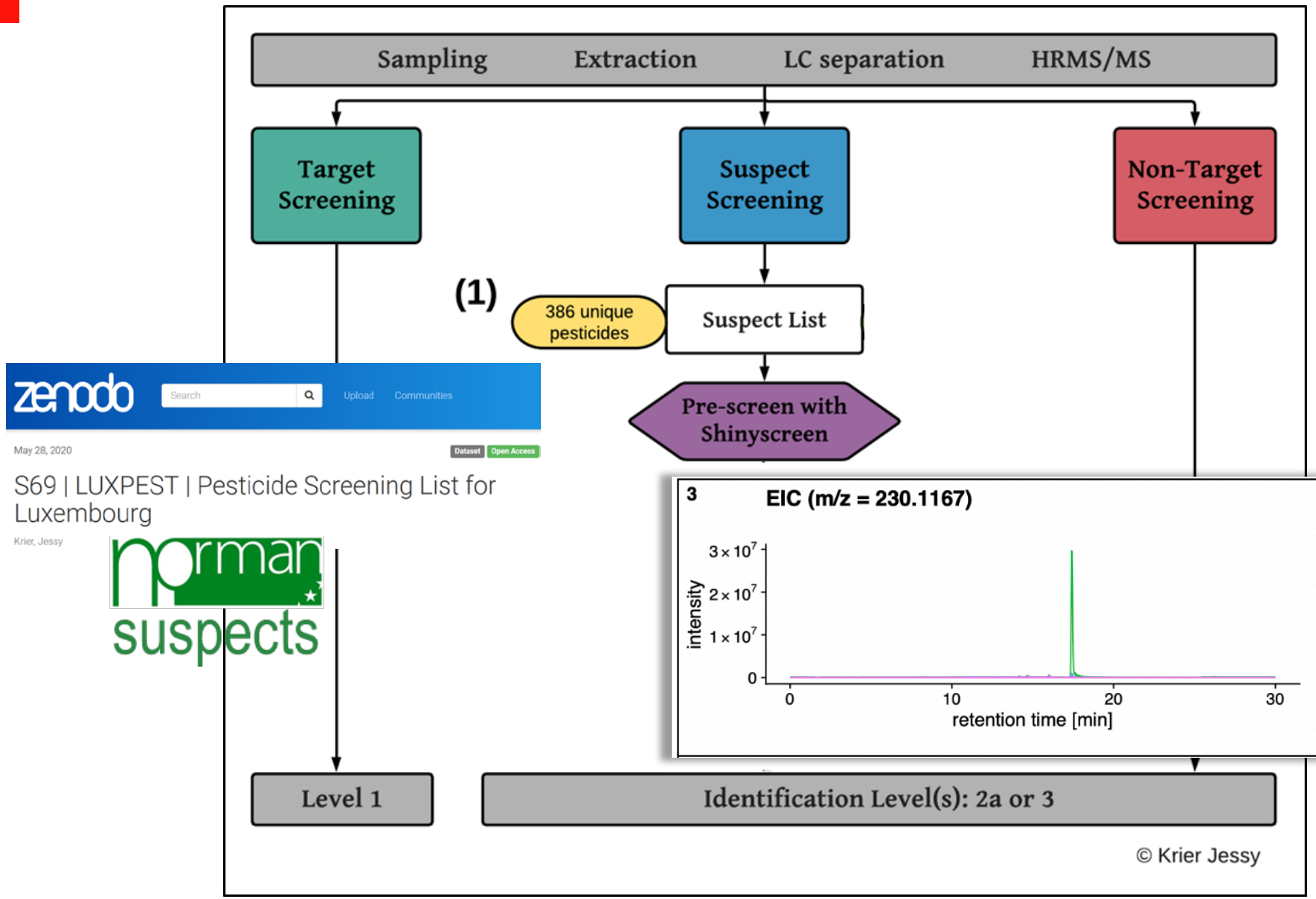


Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885. Jessy Krier (2020) S69 | LUXPEST. DOI: 10.5281/zenodo.3862689

LuxPest – Pre-screening (+QC) with Shinyscreen

THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development

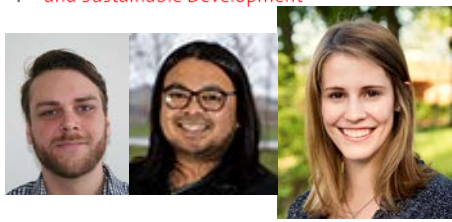


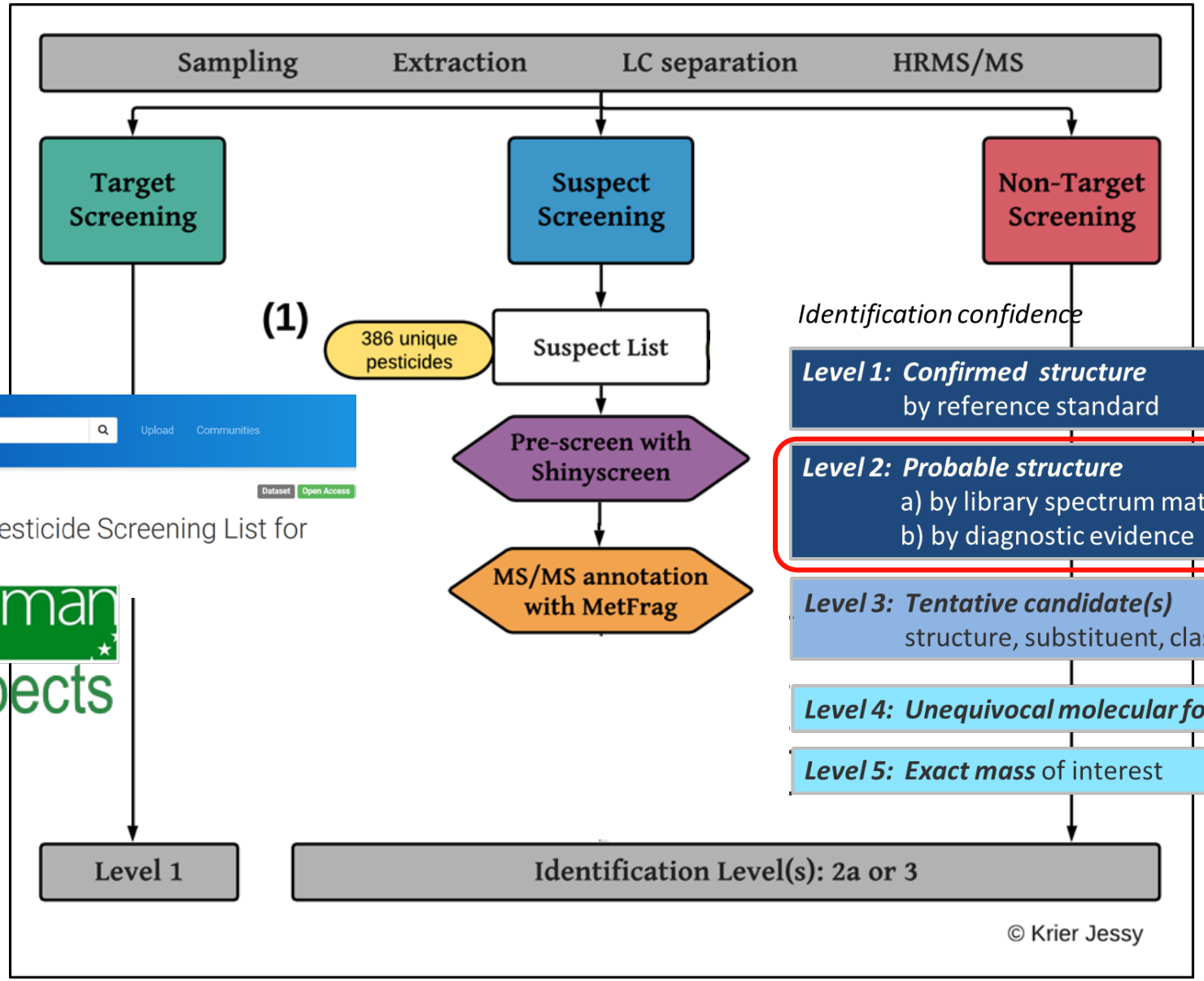


Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885. Jessy Krier (2020) S69 | LUXPEST. DOI: 10.5281/zenodo.3862689

LuxPest – MS/MS Annotation with MetFrag

THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development



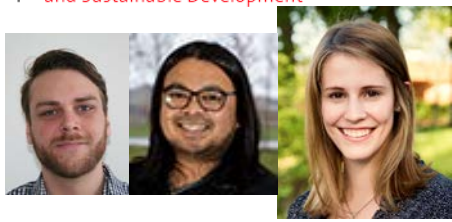


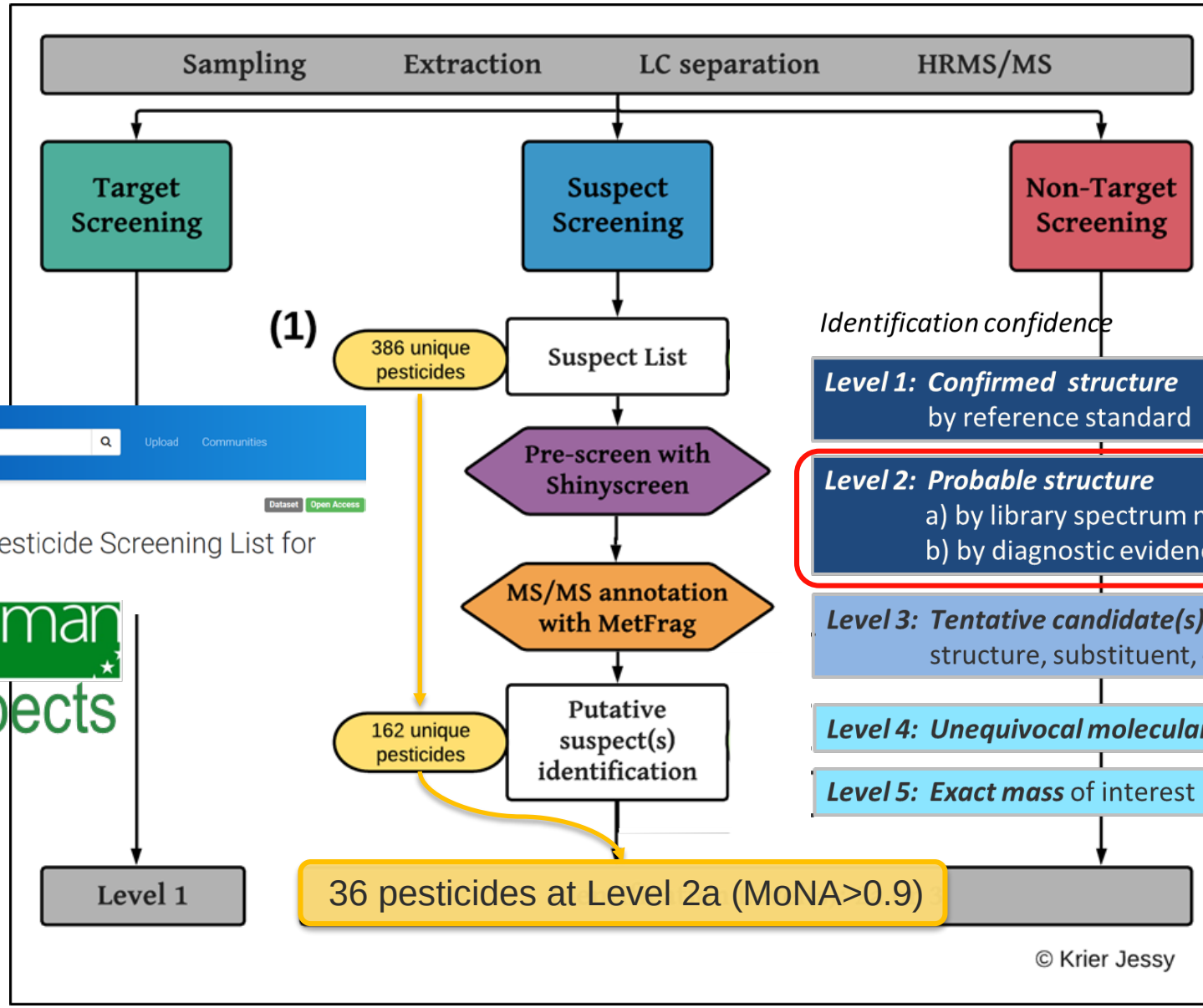
Identification confidence	*Minimum data requirements*
Level 1: Confirmed structure by reference standard	MS, MS^2, RT, Reference Std.
Level 2: Probable structure a) by library spectrum match b) by diagnostic evidence	MS, MassBank MS, High Quality Mass Spectral Database
Level 3: Tentative candidate(s) structure, substituent, class	MS, MS^2, Exp. data
Level 4: Unequivocal molecular formula	MS isotope/adduct
Level 5: Exact mass of interest	MS

Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885. Jessy Krier (2020) S69 | LUXPEST. DOI: 10.5281/zenodo.3862689

LuxPest – MS/MS Annotation with MetFrag

THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development





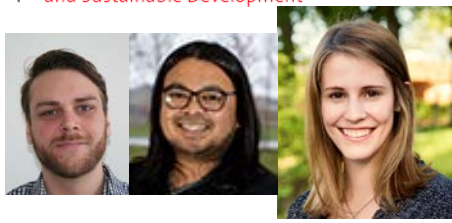
Identification confidence — *Minimum data requirements*

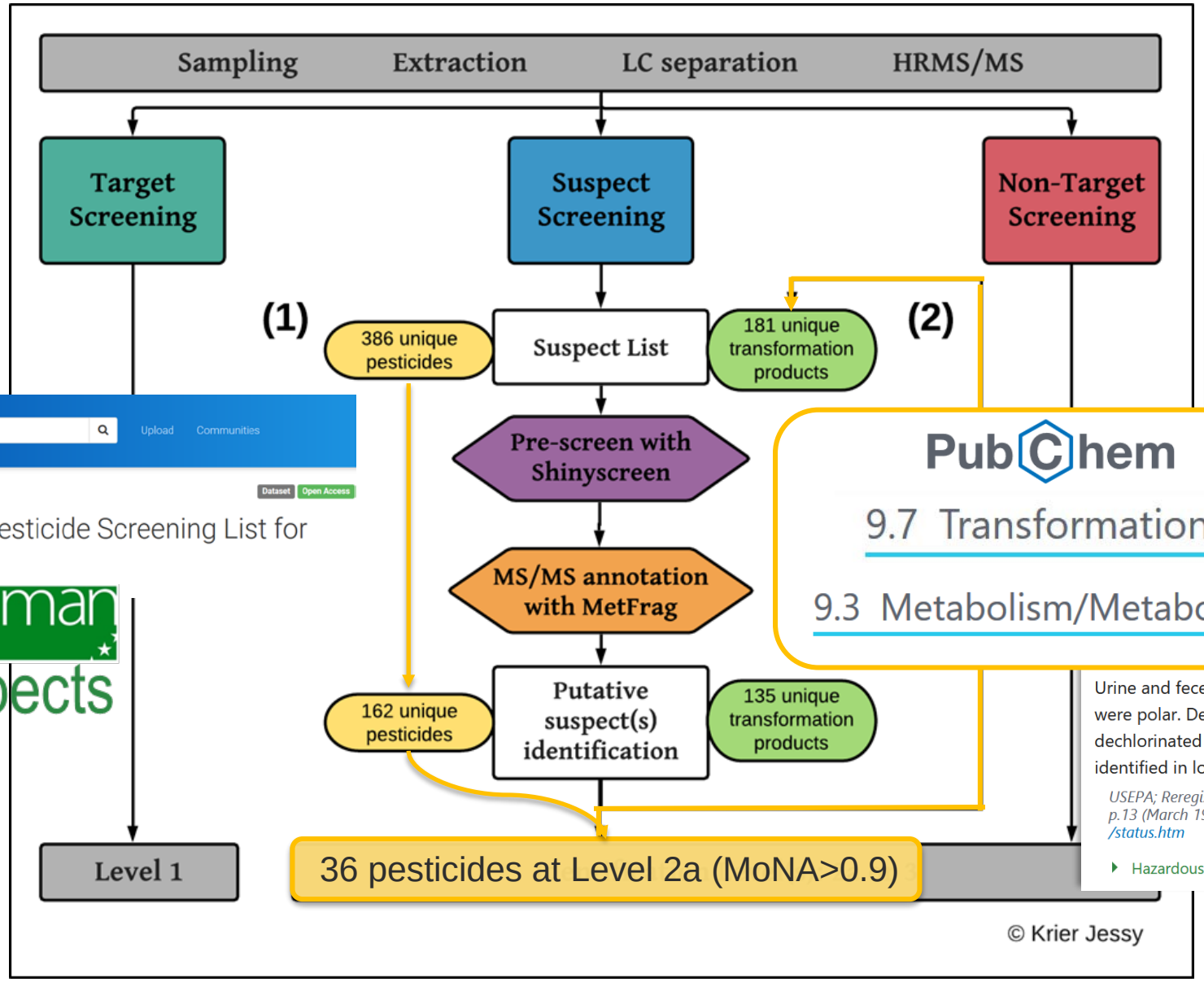
Level 1: Confirmed structure by reference standard	MS, MS^2, RT, Reference Std.
Level 2: Probable structure a) by library spectrum match b) by diagnostic evidence	MS, MassBank MS, High Quality Mass Spectral Database
Level 3: Tentative candidate(s) structure, substituent, class	MS, MS^2, Exp. data
Level 4: Unequivocal molecular formula	MS isotope/adduct
Level 5: Exact mass of interest	MS

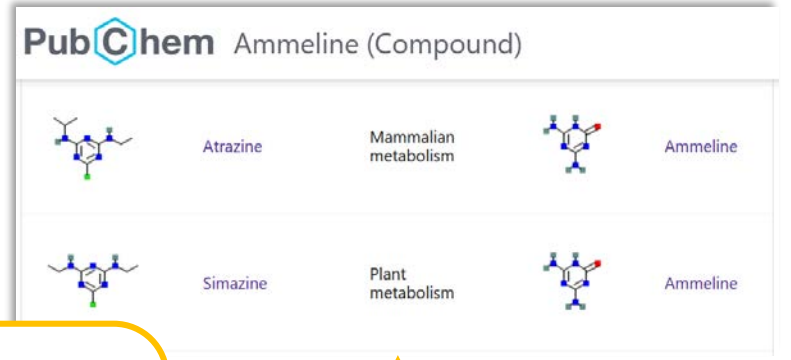
Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885. Jessy Krier (2020) S69 | LUXPEST. DOI: 10.5281/zenodo.3862689

LuxPest – Transformation Product Workflow

THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development





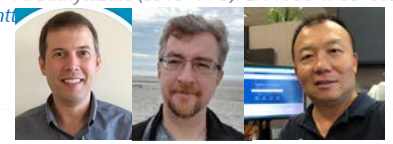


Schymanski *et al*. (2021)
DOI: 10.1186/s13321-016-0115-9

Urine and feces contained up to 25 and 15 identified metabolites, respectively, most of which were polar. Degradation of the triazine ring did not occur. Ammeline and ammelide, 2 dechlorinated and dealkylated/hydroxylated metabolites common to all triazines, were identified in low amounts in the feces.

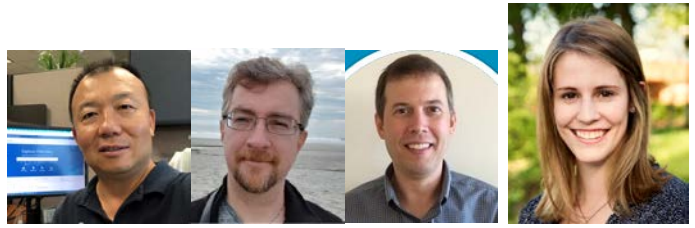
USEPA; Reregistration Eligibility Decision (RED) Database for Terbuthylazine (5915-41-3). EPA 738-R-95-005 p.13 (March 1995). Available from, as of October 11, 2012: http /status.htm

Hazardous Substances Data Bank (HSDB)



Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885. Jessy Krier (2020) S69 | LUXPEST. DOI: 10.5281/zenodo.3862689

“Circle of Data”: Literature Mining for Metabolites / TPs



PubChem Terbuthylazine (Compound)

8.3 Metabolism/Metabolites

Metabolism of terbuthylazine in rats is similar to other chloro-s-triazine herbicides. The major routes of metabolism are hydrolysis of the chlorine moiety and mono- or didealkylation. Hydroxylation of one or both of the dealkylated amine groups may also occur.

USEPA; Reregistration Eligibility Decision (RED) Database for Terbuthylazine (5915-41-3). EPA 738-R-95-005 p.12 (March 1995). Available from, as of October 11, 2012: http://www.epa.gov/pesticides/reregistration/status.htm

- Hazardous Substances Data Bank (HSDB)

Urine and feces contained up to 25 and 15 identified metabolites, respectively, most of which were polar. Degradation of the triazine ring did not occur. Ammeline and ammelide, 2 dechlorinated and dealkylated/hydroxylated metabolites common to all triazines, were identified in low amounts in the feces.

USEPA; Reregistration Eligibility Decision (RED) Database for Terbuthylazine (5915-41-3). EPA 738-R-95-005 p.13 (March 1995). Available from, as of October 11, 2012: http://www.epa.gov/pesticides/reregistration/status.htm

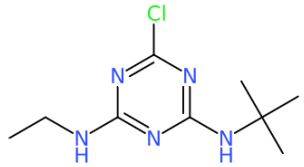
- Hazardous Substances Data Bank (HSDB)

In mammals, following oral administration, ...a de-ethyl metabolite forms rapidly, followed by conjugates of products formed by oxidation of one methyl group of the tert-butyl moiety. All are rapidly excreted.

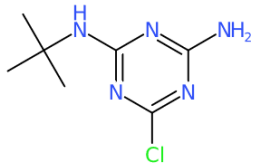
Tomlin CDS, ed. Terbuthylazine (5915-41-3). In: The e-Pesticide Manual, Version 2.2 (2002). Surrey UK, British Crop Protection Council.

- Hazardous Substances Data Bank (HSDB)

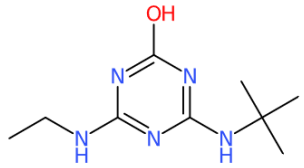
CDK DEPICT



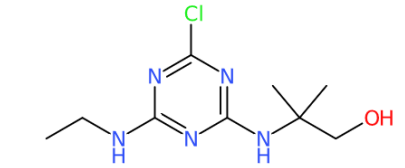
Terbutylazine CID:22206



desethyl-terbutylazine CID:108201



2-hydroxy-terbutylazine CID:135495928



(hydroxy-t-butyl)-Terbutylazine CID:779516

Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885.
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

“Living data connections”

zenodo Search Upload Communities

June 11, 2020 Dataset Open

S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem

LCSB-ECI; Krier, Jessy; Schymanski, Emma; PubChem Team; Bolton, Evan; Thiessen, Paul; Zhang, Jeff

This is the collection associated with list S68 HSDBTPS Transformation Products Extracted from HSDB Content in PubChem on the NORMAN Suspect List Exchange.

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

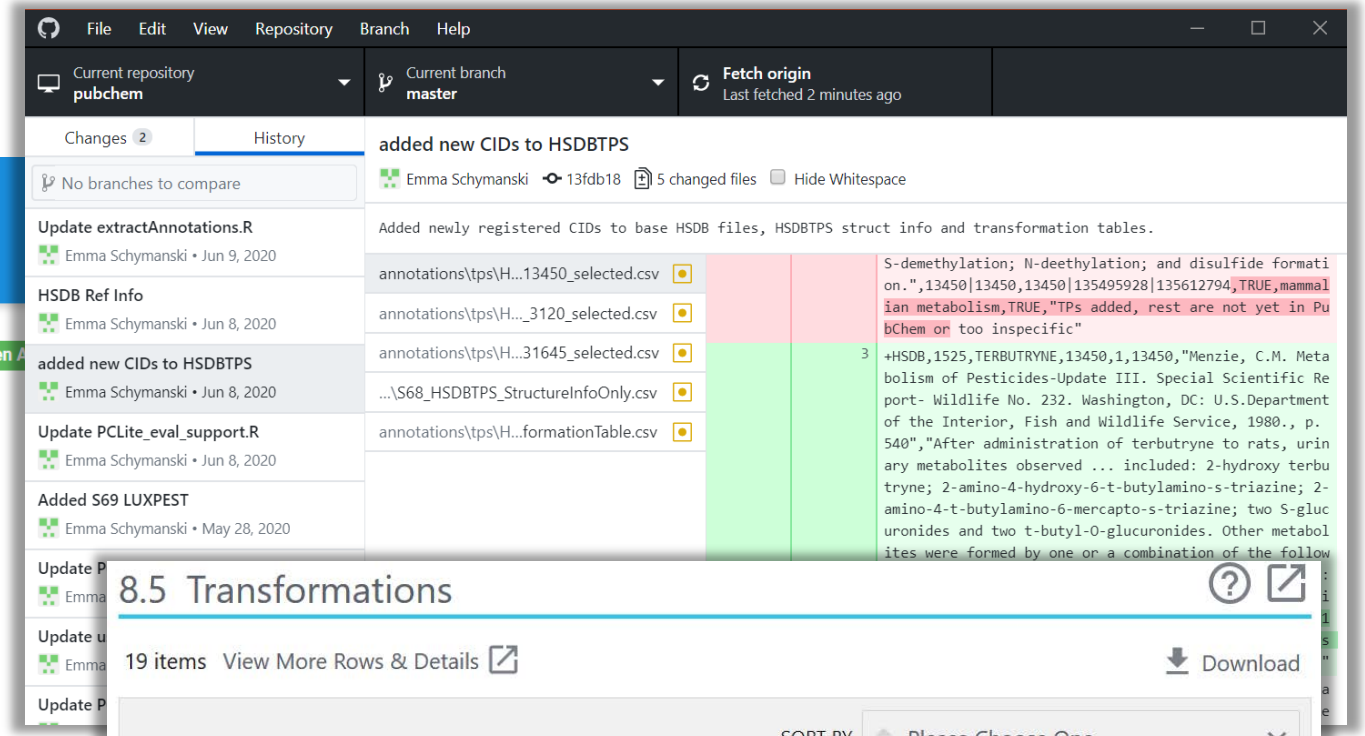
HSDBTPS is a list of metabolites / transformation products extracted from the "Metabolites/Metabolism" section from HSDB (Hazardous Substance Data Bank) in PubChem (https://pubchem.ncbi.nlm.nih.gov/source/11933). Dataset DOI: 10.5281/zenodo.3827487.

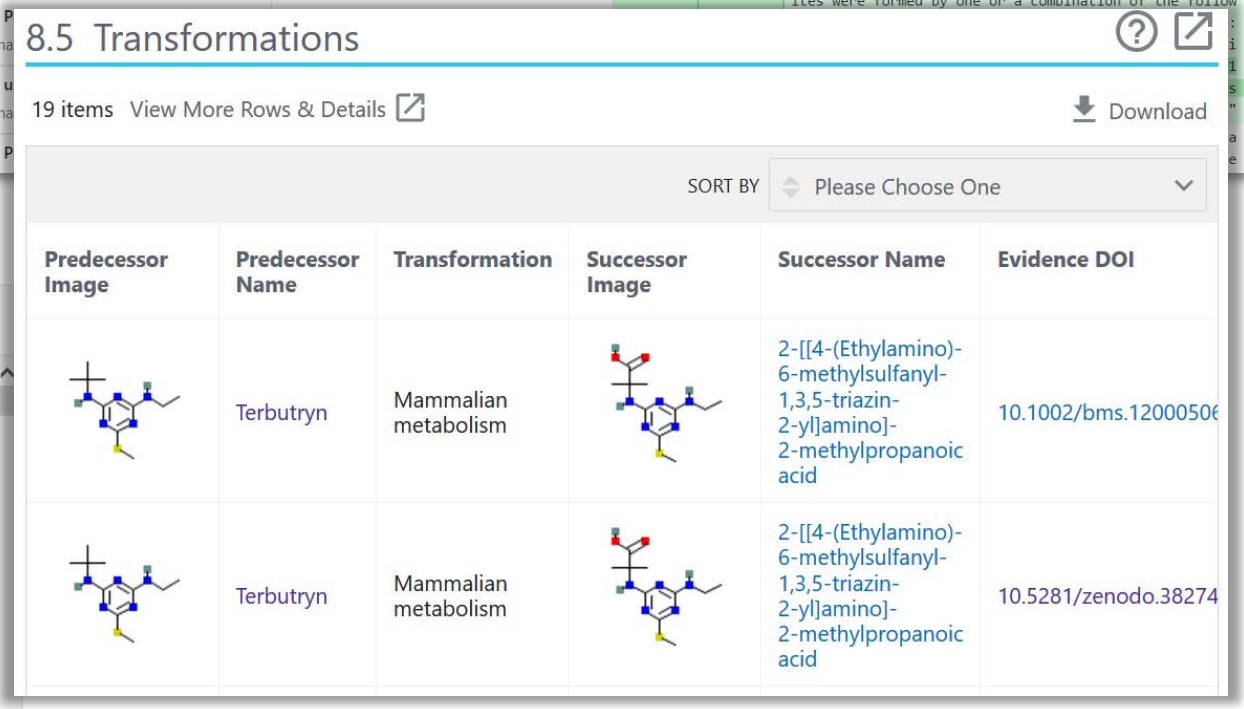
Preview

Predecessor_CID	Predecessor_Name	Successor_CID	Successor_Name	Transf
13450	Terbutryn	110189337	2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	mamma metabc
13450	Terbutryn	110189337	2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	mamma metabc

https://git-r3lab.uni.lu/eci/pubchem/

LCSB-ECI & PubChem Team. DOI 10.5281/zenodo.3890392





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

uni.lu LCSB

LuxPest – Terbutylazine and (tentative) TPs – in PubChem

Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885

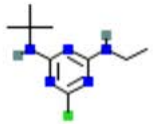
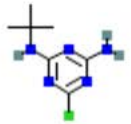
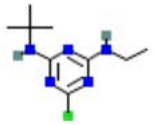
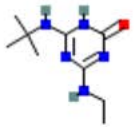
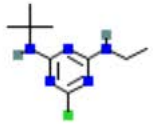
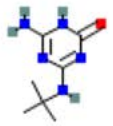
PubChem Terbuthylazine (Compound)

8.5 Transformations

30 items View More Rows & Details

Download

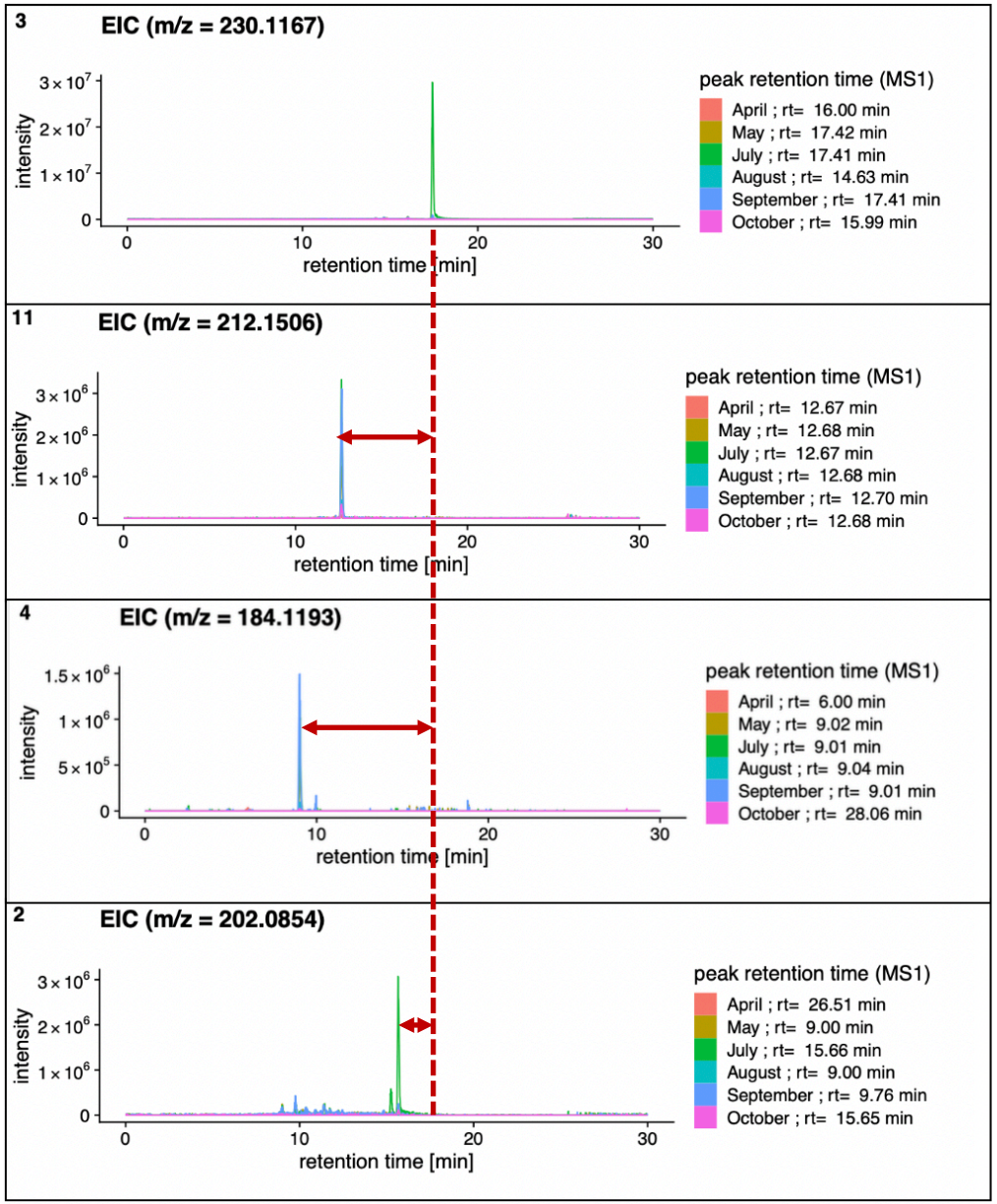
SORT BY Please Choose One

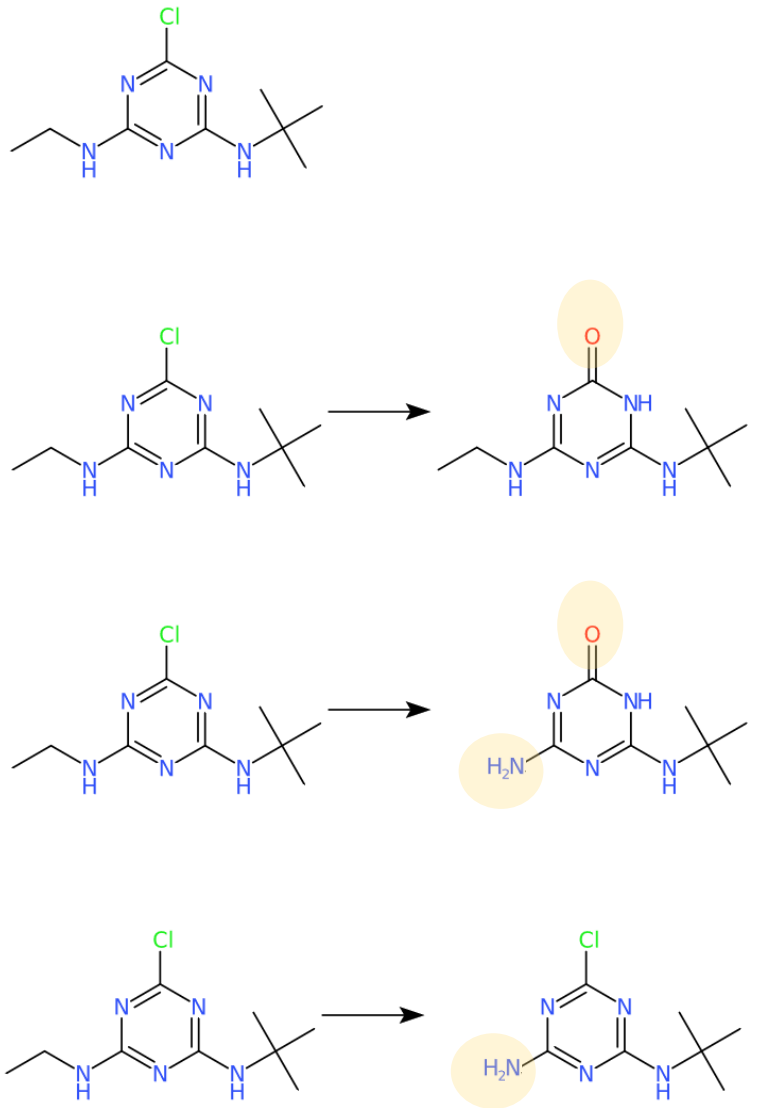
Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	E
	terbuthylazine		desethyl-terbuthylazine	Environmental		1(
	terbuthylazine		hydroxy-terbuthylazine	Environmental		1(
	terbuthylazine		2-hydroxy-desethyl-terbuthylazine	Environmental		1(

https://pubchem.ncbi.nlm.nih.gov/compound/22206#section=Transformations

LuxPest – Terbutylazine and (tentative) TPs – in water!

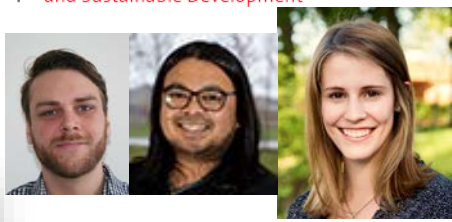
Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885

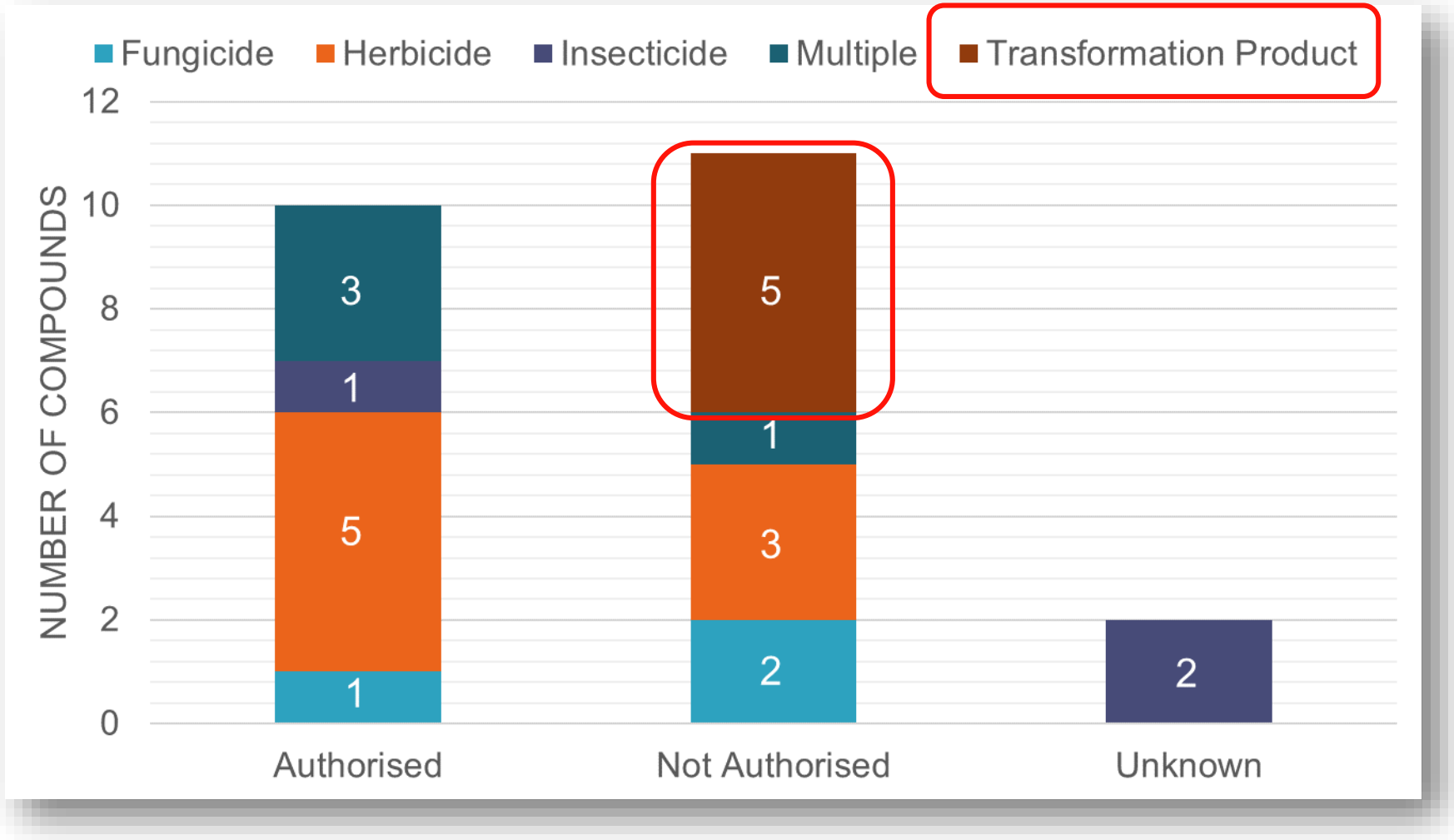




LuxPest – Verification and Quantification

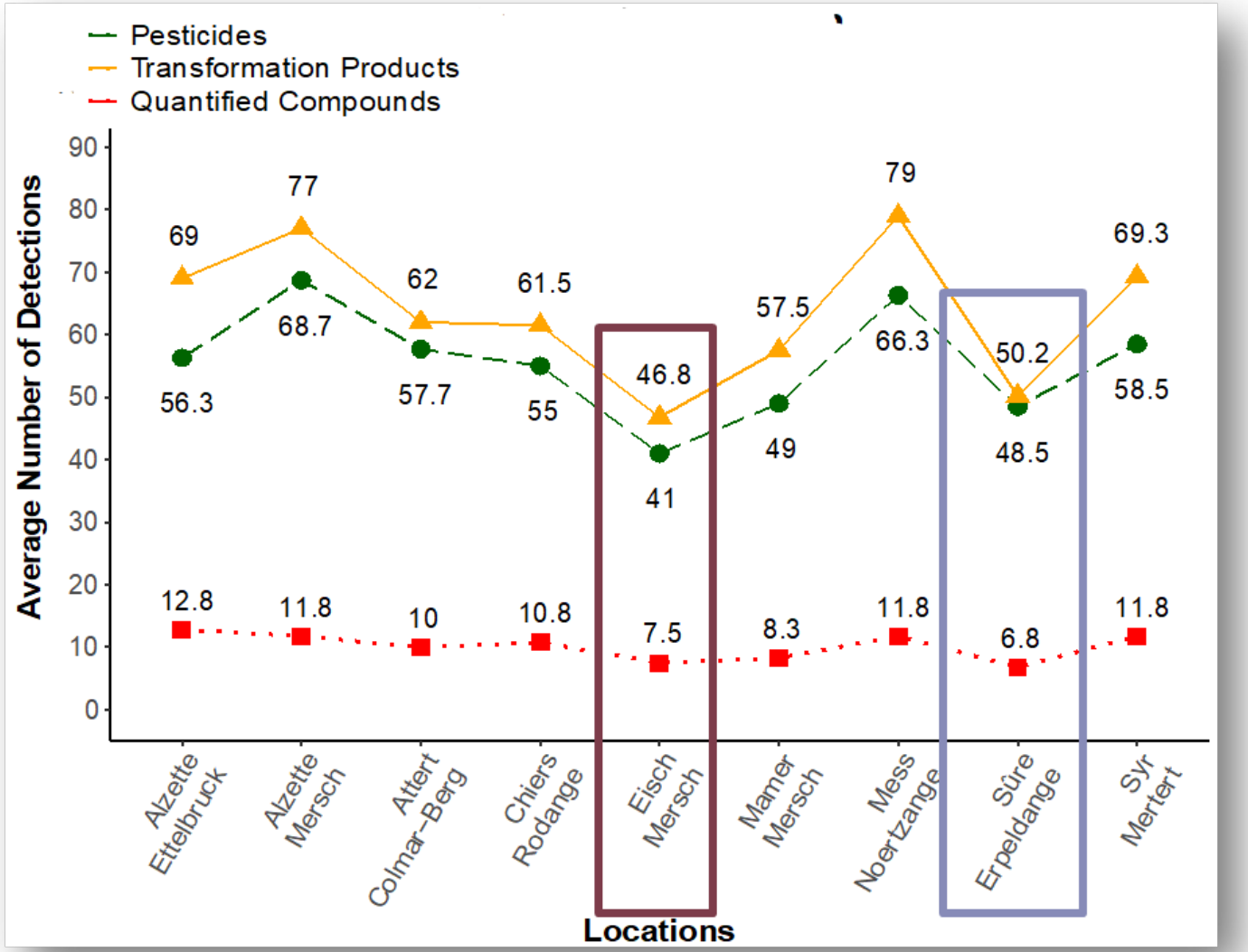
THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development

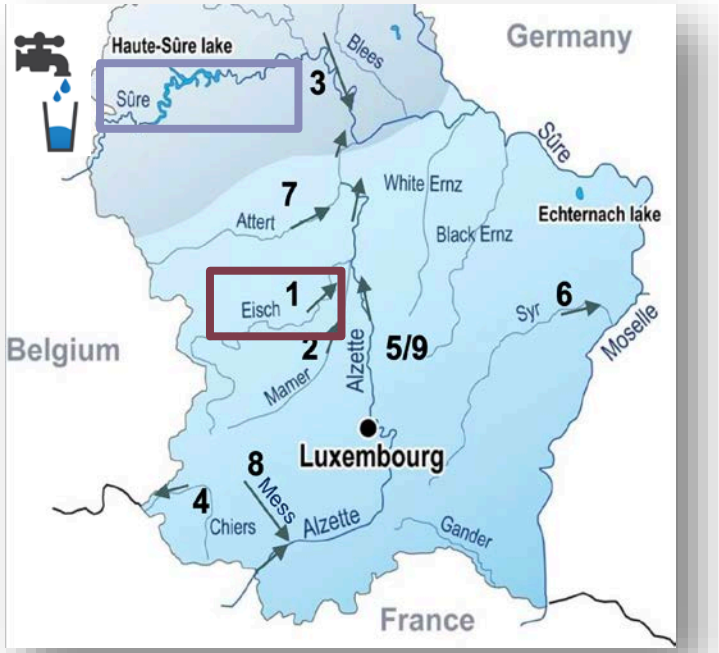




Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885.

LuxPest – Spatial Distribution

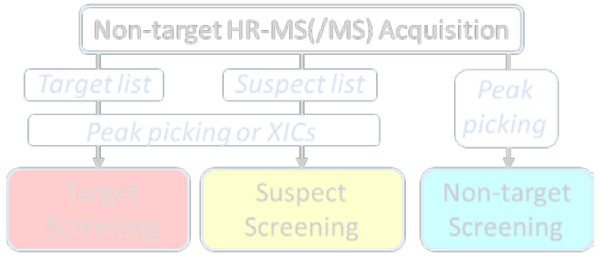




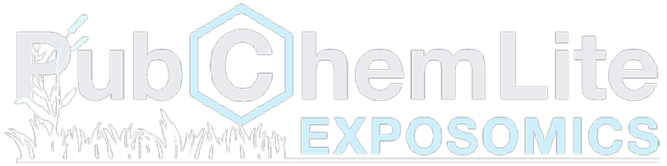
Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885.

Outline of Today

- Introduction and Background
  - Definitions, Key Resources
  - Examples



- Identification & Chemical Space
  - Identification + MetFrag
  - PubChemLite for Exposomics

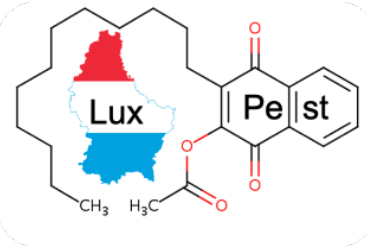


- Case Study: LuxPest

- Take-home messages!

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

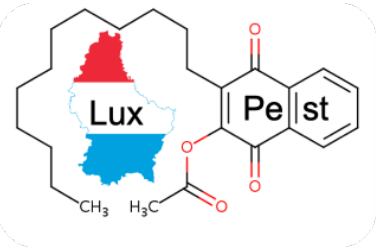
Open Source Workflows for Emerging Contaminants

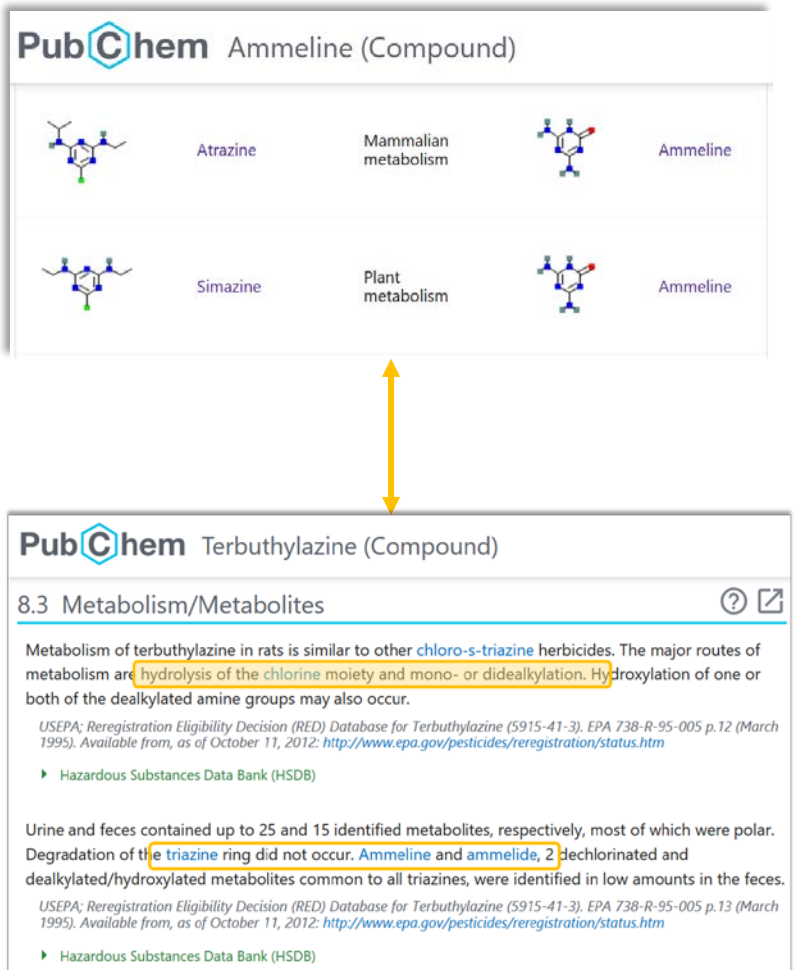


- Open and FAIR Expert Knowledge Exchange
- Open Source pre-screening with auto-QC & manual review
- Comprehensive & open annotation combining MetFrag & MassBank

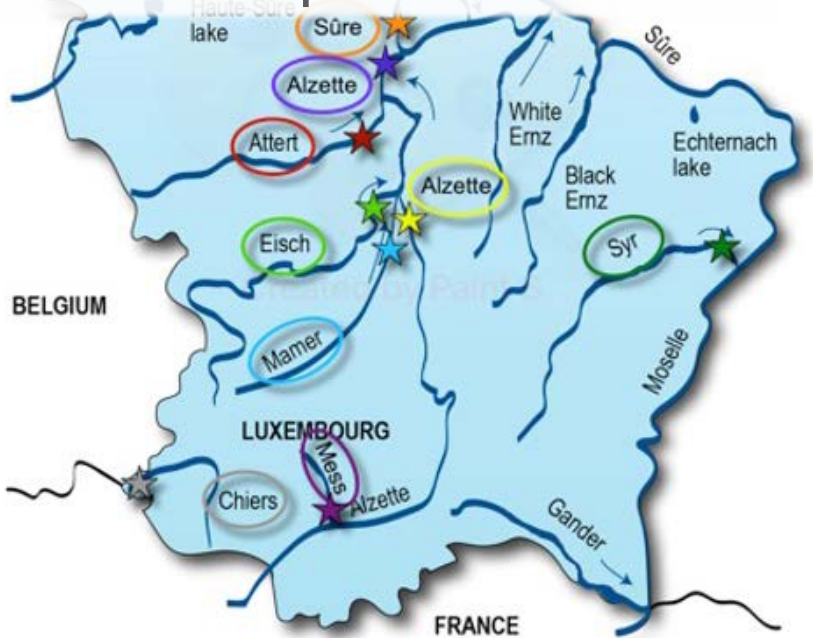
Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885.

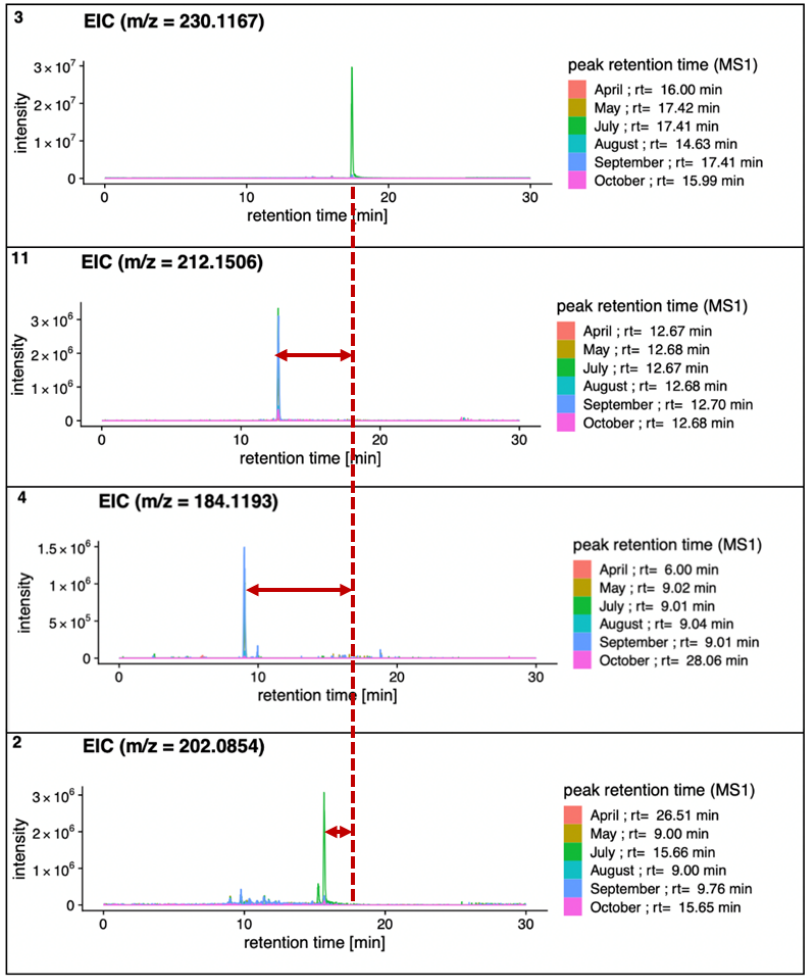
Open Source Workflows for Emerging Contaminants





- Extensive transformation product efforts
- Enhanced visualisation and interpretation





Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885.

Outcomes?

THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development

- PubChem Compound TOC ? 49,493,641
  - Agrochemical Information ? 3,045
  - Associated Disorders and Diseases ? 20,847
  - Biologic Description ? 2,056,521
  - Biological Test Results ? 3,622,920
  - Biomolecular Interactions and Pathways ? 125,253
  - Chemical and Physical Properties ? 263,015
  - Classification ? 1,454,824
  - Drug and Medication Information ? 17,922
  - Food Additives and Ingredients ? 8,414
  - Identification ? 4,968
  - Information Sources ? 20,271,277
  - Literature ? 1,833,941
  - Names and Identifiers ? 1,275,170
  - Patents ? 36,351,418
  - Pharmacology and Biochemistry ? 110,628
  - Related Records ? 9,224,590
  - Safety and Hazards ? 149,319
  - Spectral Information ? 480,730
  - Structures ? 9,117,635
  - Toxicity ? 114,012
  - Use and Manufacturing ? 115,321

Chemical Safety ? 147,023

Continued efforts for improved monitoring of chemicals (and actions!) in Luxembourg ... ... and the world!

PubChemLite
EXPOSOMICS

~370,000 entries “small”

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

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

uni.lu

LCSB

Expert Knowledge is YOUR Knowledge!

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

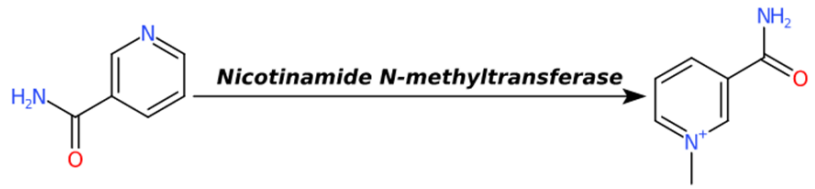
- Help us help you! Add data with FAIR templates

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



Nicotinamide to MNAM DOI:10.1124/dmd.112.049734

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

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 (submitted) FAIR-ifying the Exposome: Preprint DOI: 10.5281/zenodo.5495109

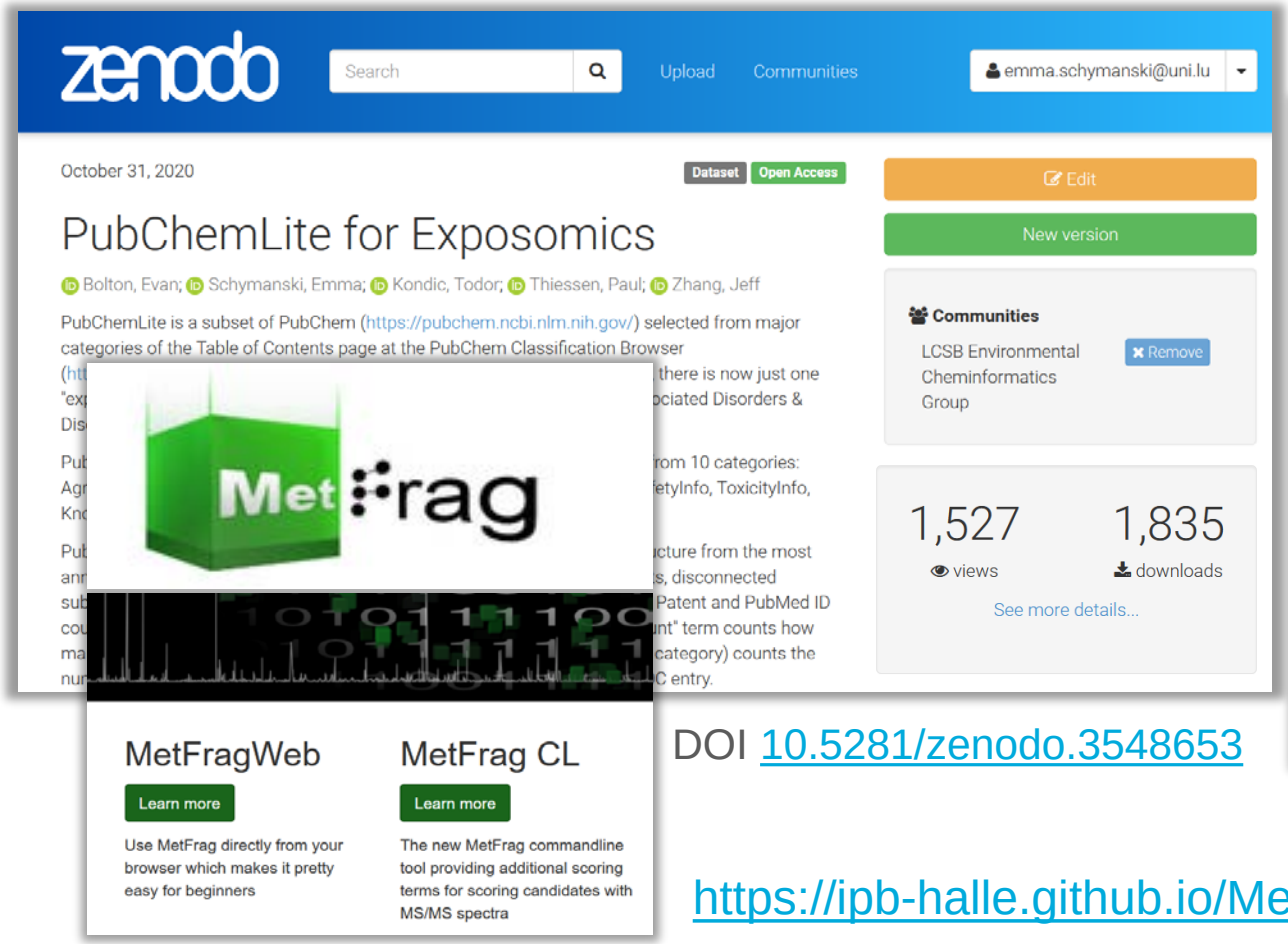
Schymanski & Bolton (2021) FAIR Chemical Structures. *J. Cheminform*. DOI: 10.1186/s13321-021-00520-4

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

Putting PubChemLite to Practice?

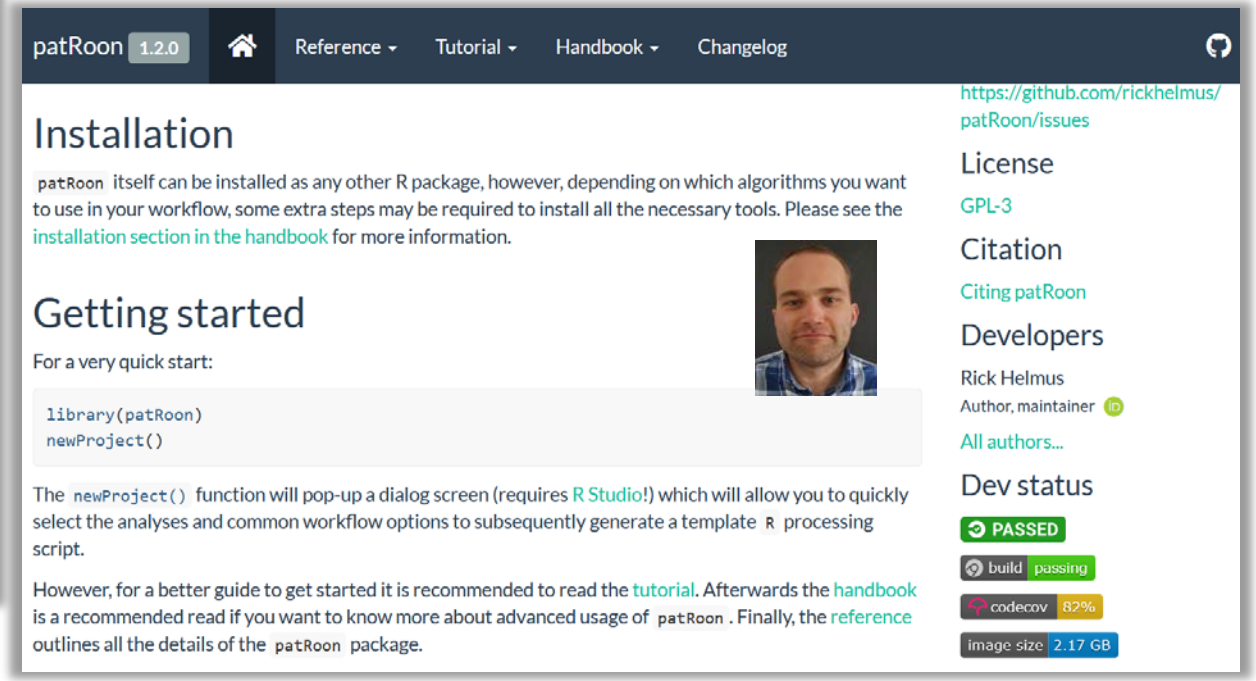
PubChemLite EXPOSOMICS

- Downloadable & fully integrated in patRoon, MetFragWeb & MetFrag CL!



DOI 10.5281/zenodo.3548653

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



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

Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0
Helmus *et al.* (2021) patRoon *J Cheminform* **13,** 1. DOI: 10.1186/s13321-020-00477-w

IPB Halle
University of Amsterdam
NIH U.S. National Library of Medicine National Center for Biotechnology Information
uni.lu
LCSB

Thank you!

Email: emma.schymanski@uni.lu
Twitter: @ESchymanski

Slides @ DOI:10.5281/zenodo.5595630

https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics/

PubChemLite EXPOSOMICS

… and team

Community Efforts!

MassBank consortium