

Discovering Contaminants & their TPs in Luxembourg Waters using Open Cheminformatics Approaches

Slides @ DOI: [10.5281/zenodo.5059829](https://doi.org/10.5281/zenodo.5059829)

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

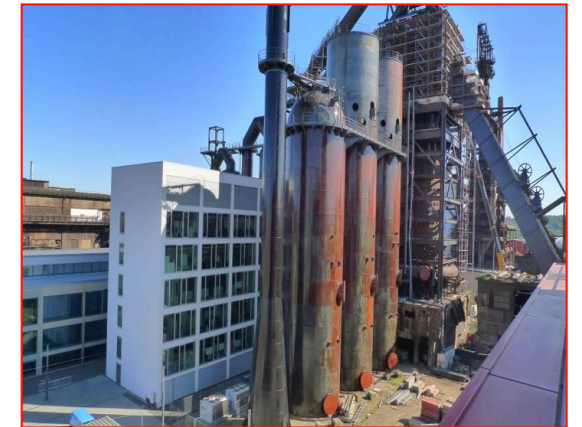
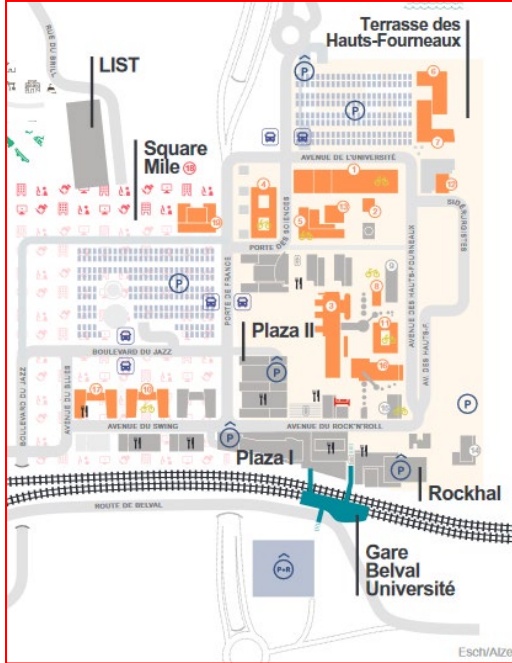
Environmental Cheminformatics Group,
Luxembourg Centre for Systems Biomedicine, University of Luxembourg

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

Web: https://www.uni.lu/lcsb/research/environmental_cheminformatics/



LCSB and the University of Luxembourg



- Uni Lu was founded in 2003
 - Teenage years!
- LCSB was founded in 2009
 - ...and is still pre-teenager
- Young and very dynamic working environment!

LCSB: Interdisciplinary Groups (PIs)



Rudi Balling
Director

Jorge Gonçalves
Systems Control

Jens Schwamborn
Developmental and
Cellular Biology

Paul Wilmes
Systems Ecology
ERC fellow

Reinhard Schneider
Bioinformatics Core

Antonio del Sol
Computational Biology

Enrico Glaab
Biomedical Data
Science

Rejko Krüger
Translational
Neuroscience
FNR PEARL chair

Jochen Klucken
Digital Medicine

Jochen Schneider
Medical Translational
Research

Frank Hertel
Interventional
Neuroscience

Michel Mittelbronn
Neuropathology
FNR PEARL chair

Evan Williams
Gene Expression
and Metabolism

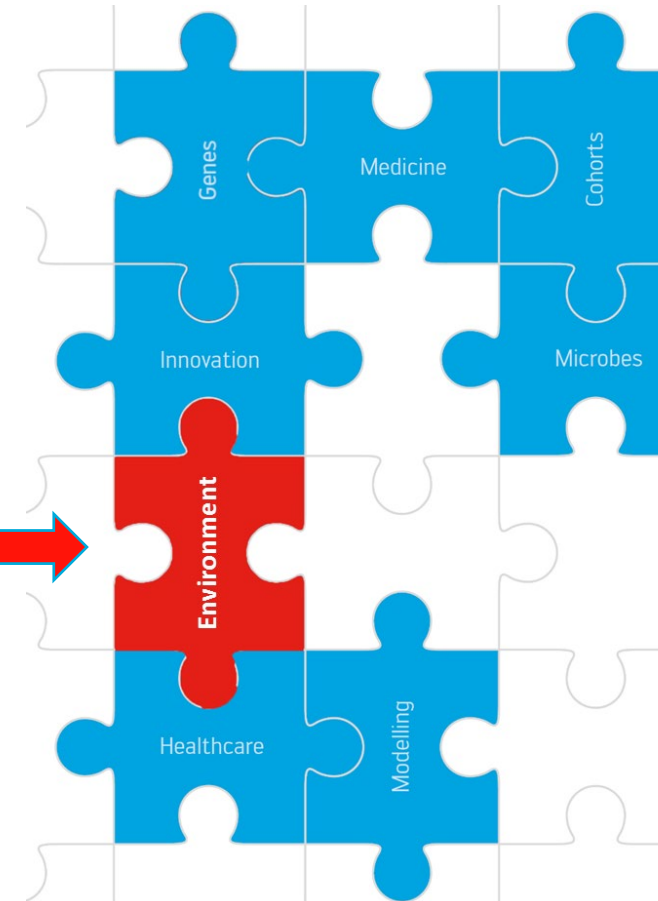
Dirk Brenner
Immunology
and Genetics

Alex Skupin
Integrative
Cell Signalling

Emma Schymanski
Environmental
Cheminformatics
FNR ATTRACT fellow

Anne Grünewald
Molecular & Functional
Neurobiology
FNR ATTRACT fellow

Carole Linster
Enzymology and
Metabolism



Environmental Cheminformatics @ LCSB



Adelene Lai
@adelene.lai
Given access 1 year ago



Emma Schymanski @emma.schymanski
Given access 1 year ago



Lorenzo Favilli
@lorenzo.favilli
Given access 1 year ago



Anjana Elapavalore
@anjana.elapavalore
Given access 1 year ago



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



Mira Narayanan
@mira.narayanan
Given access 2 months ago



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



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



Parviel Chirsir
@parviel.chirsir
Given access 1 month ago



Corey Griffith
@corey.griffith
Given access 1 year ago



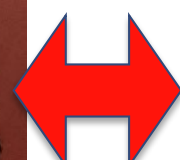
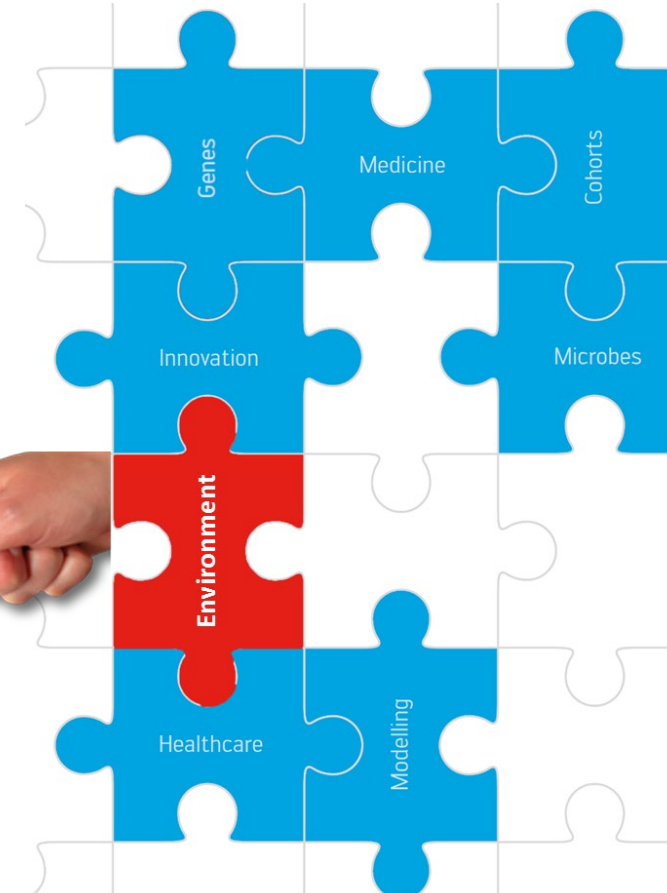
Jessy Krier @jessy.krier
Given access 1 year ago



Todor Kondic
@todor.kondic
Given access 1 year ago



Dagny Aurich
@dagny.aurich
Given access 3 months ago



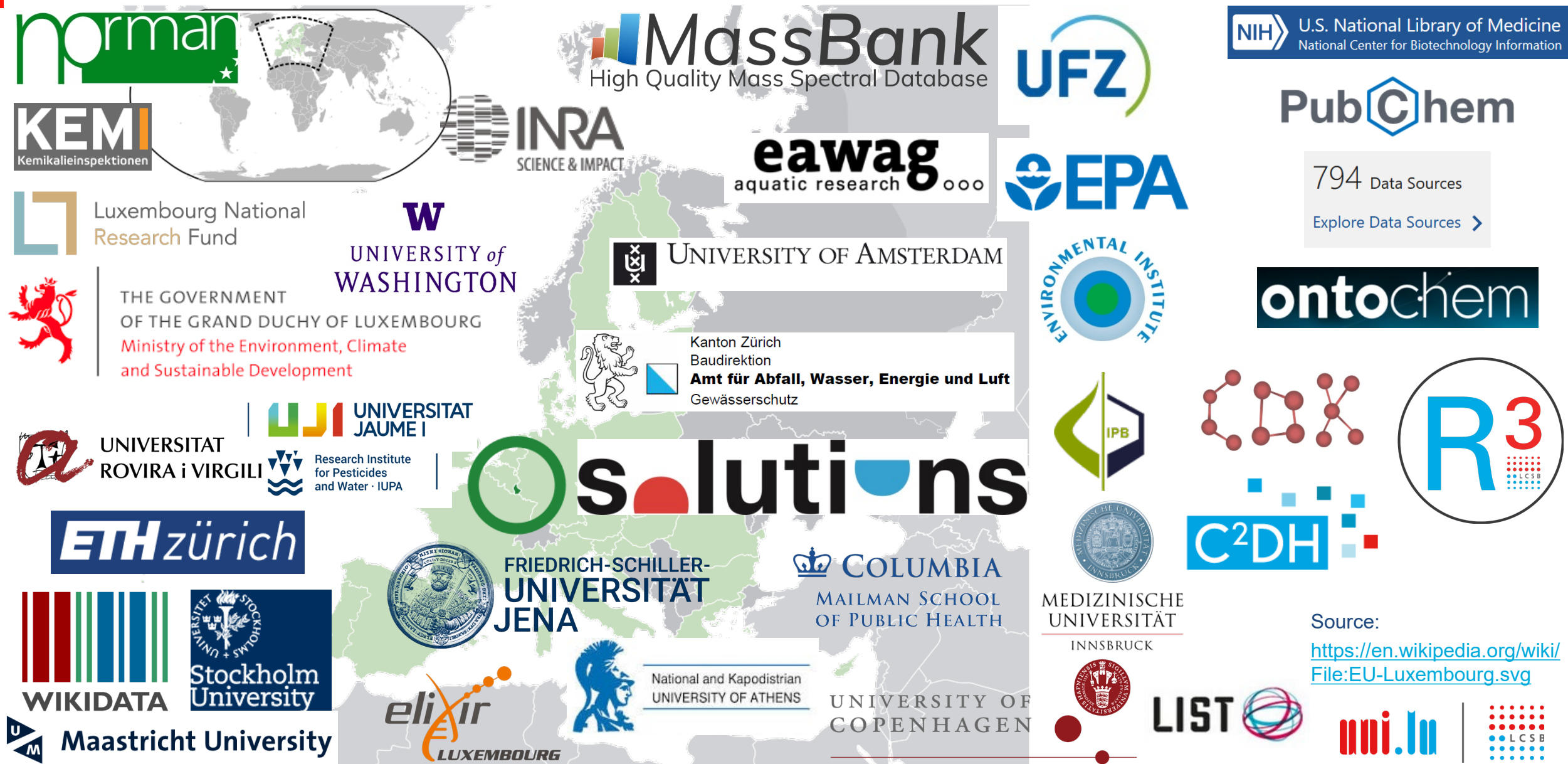
INSTITUTE FOR ADVANCED STUDIES (IAS)



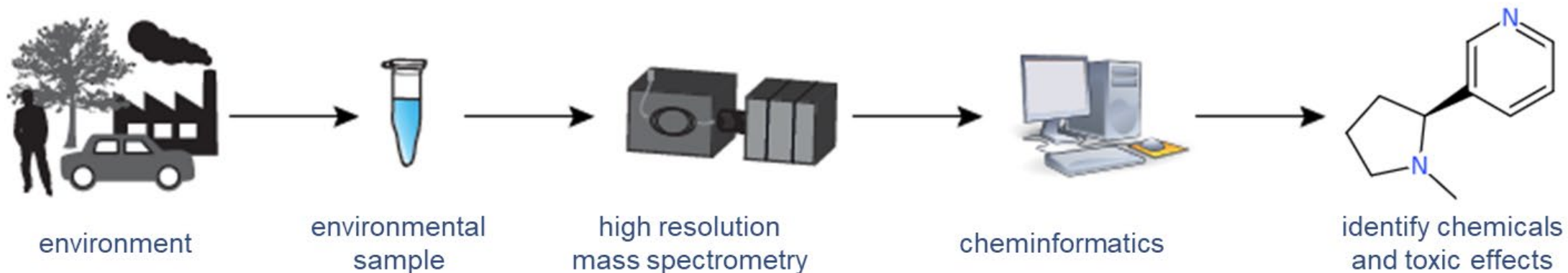
Luxembourg National
Research Fund



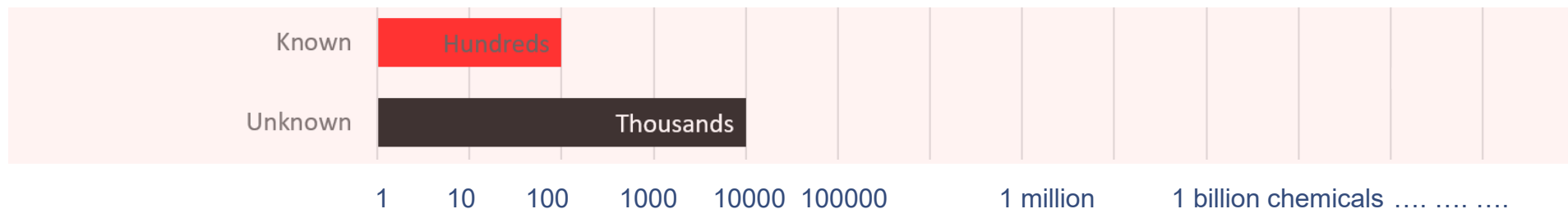
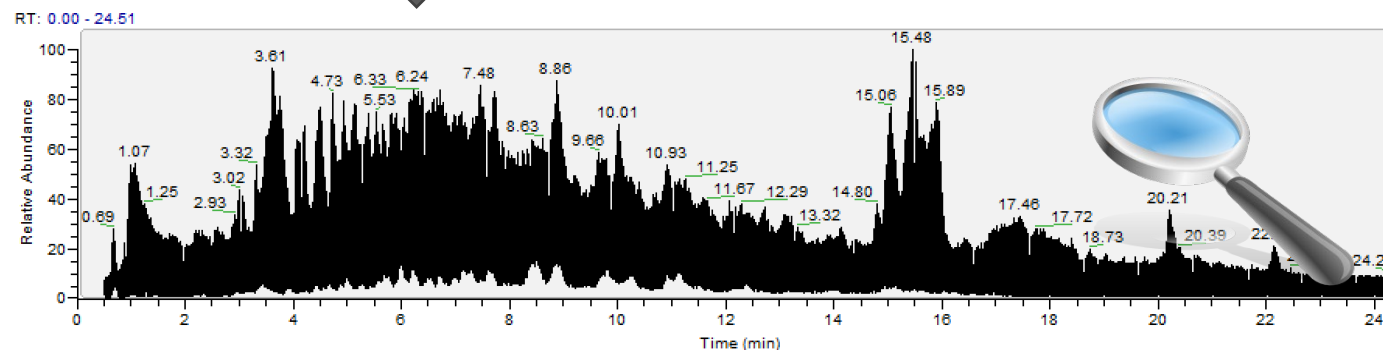
... and a lot of International Collaborations!



What is Environmental Cheminformatics?



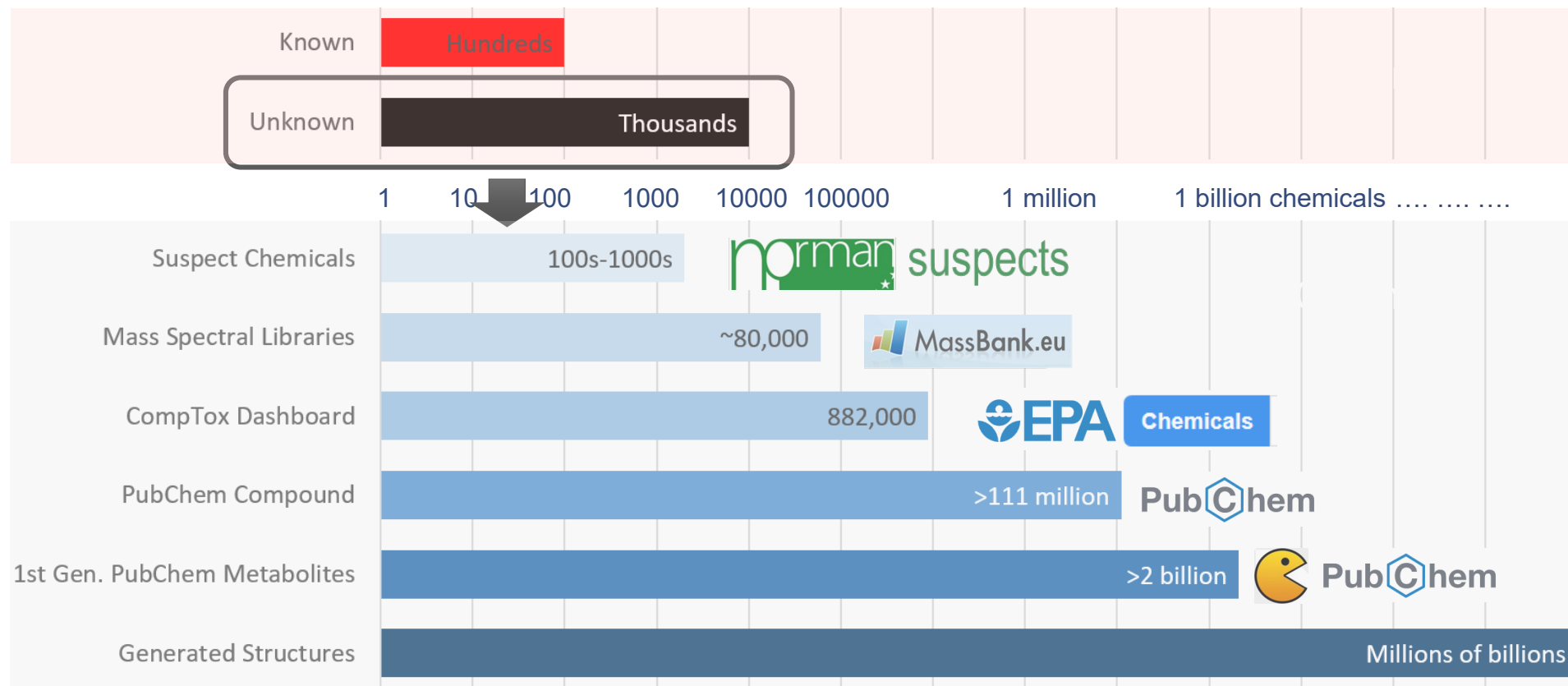
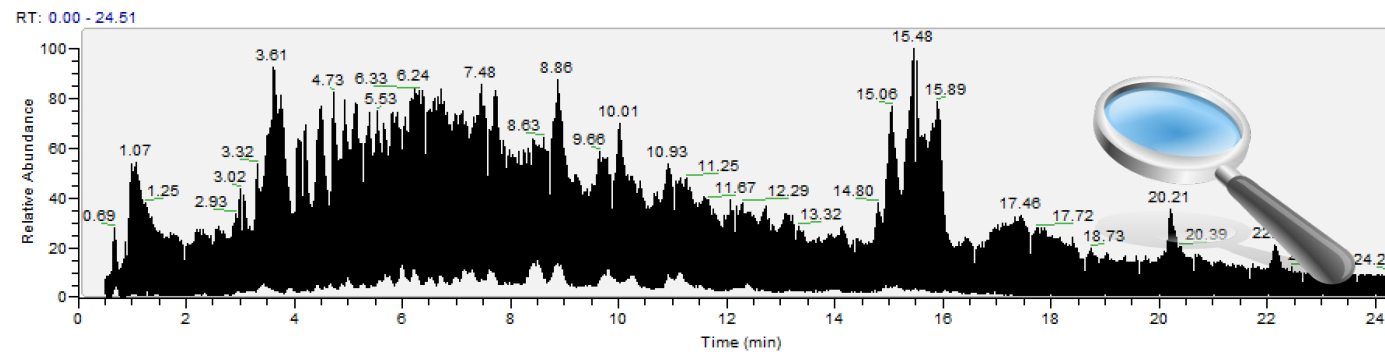
High resolution
mass spectrometry



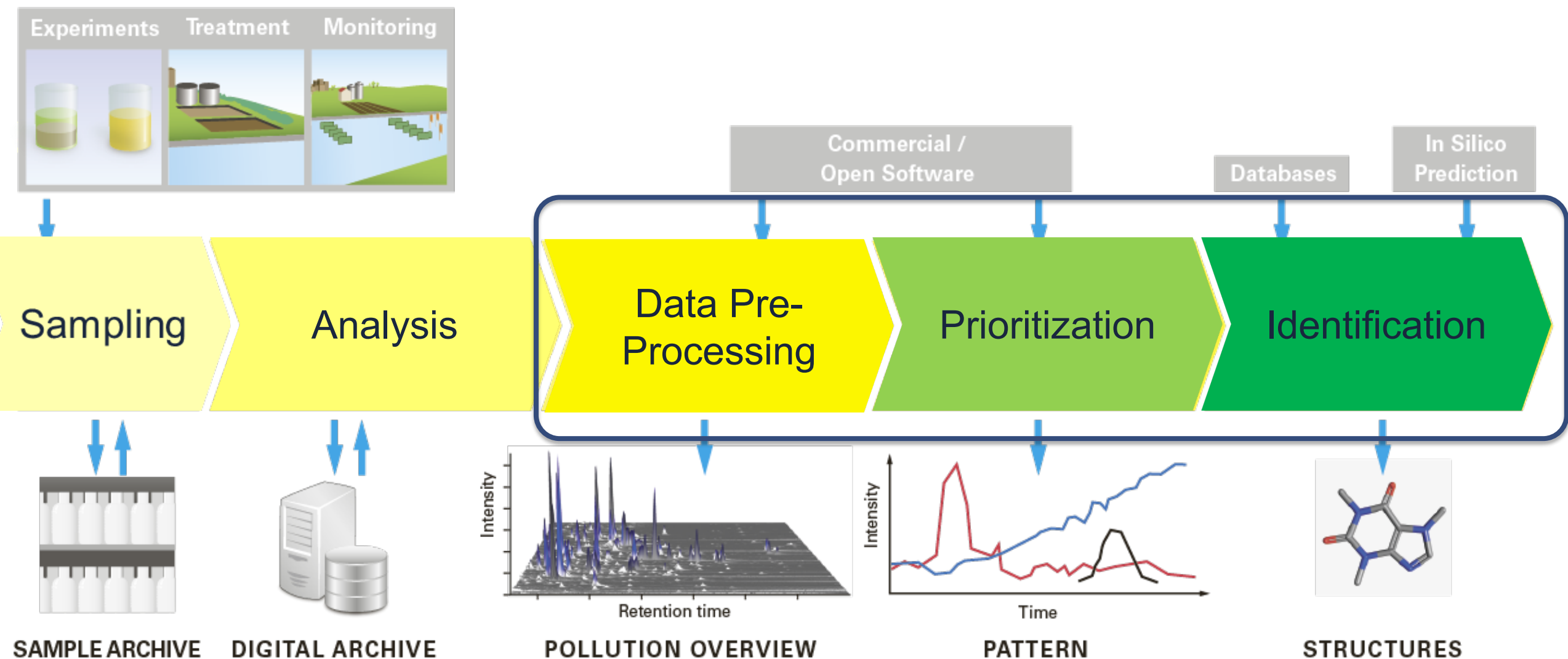
What is Environmental Cheminformatics?

High resolution
mass spectrometry

AND connecting
chemical knowledge



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



Open Source Workflows for NT-HRMS: patRoon



Rick
Helmus

Software | [Open Access](#) | Published: 06 January 2021

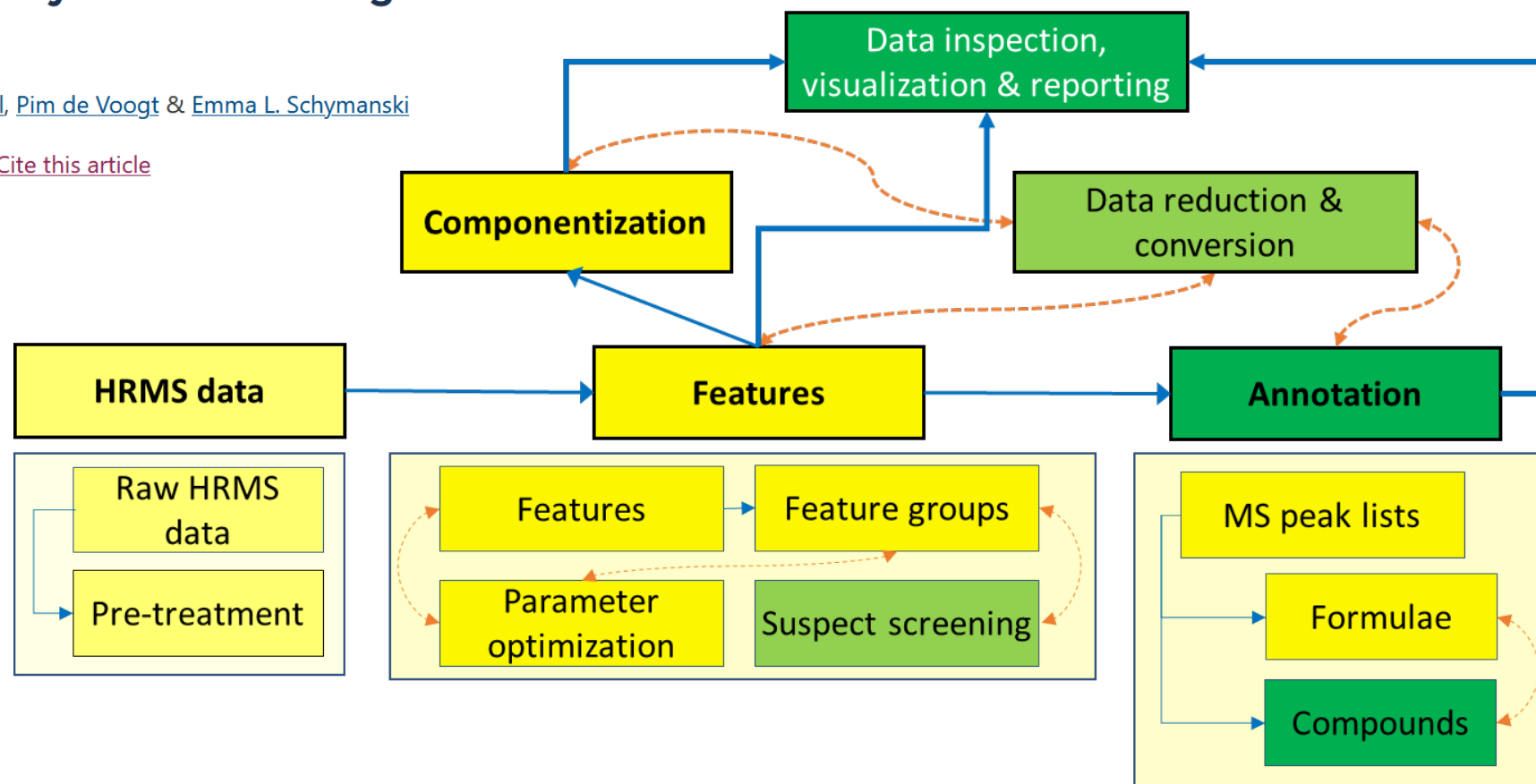
patRoon: open source software platform for environmental mass spectrometry based non-target screening

[Rick Helmus](#) , [Thomas L. ter Laak](#), [Annemarie P. van Wezel](#), [Pim de Voogt](#) & [Emma L. Schymanski](#)

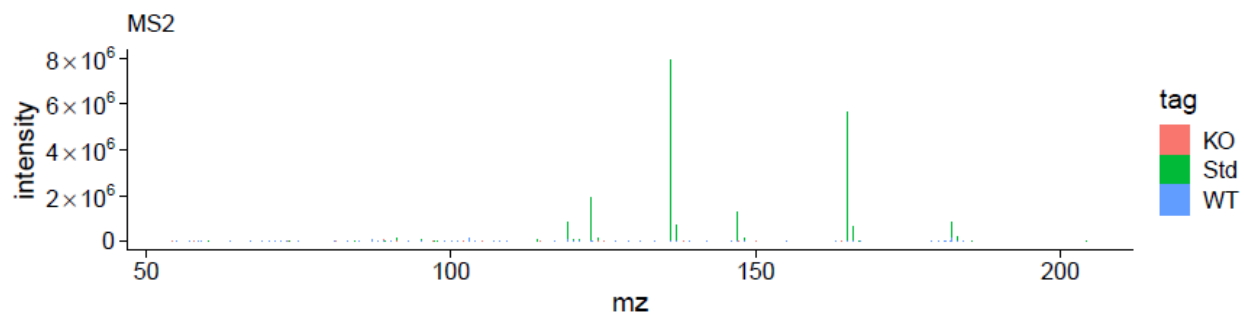
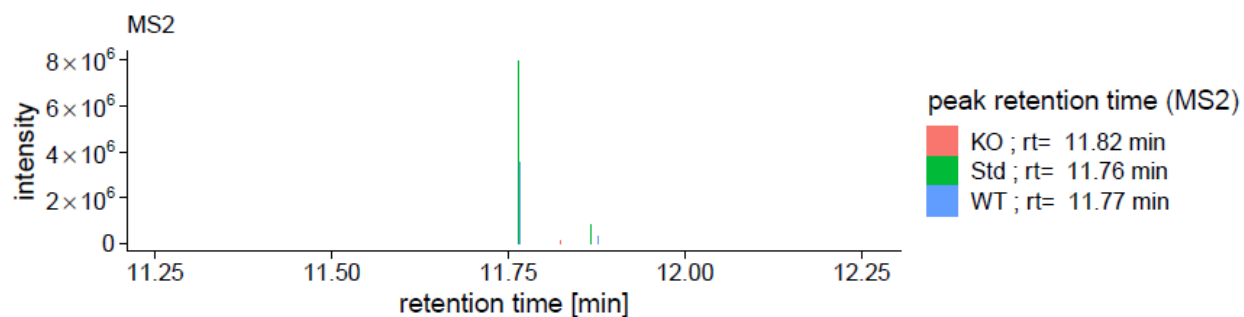
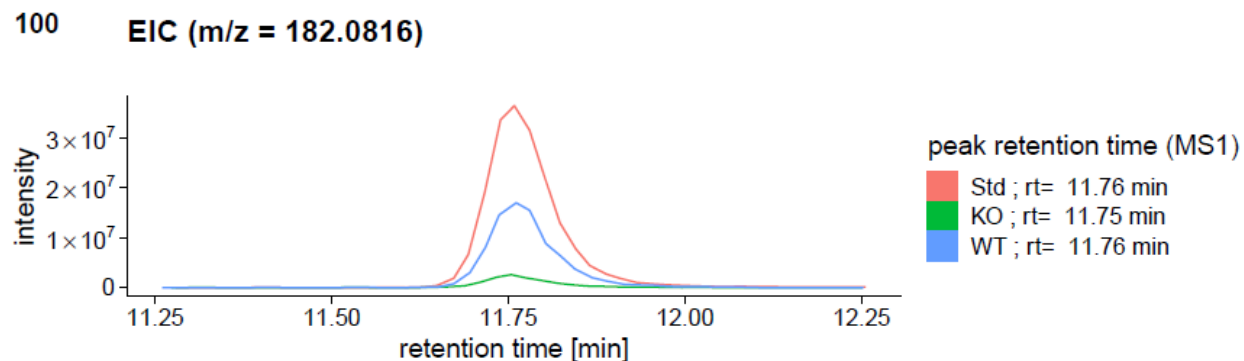
Journal of Cheminformatics **13**, Article number: 1 (2021) | [Cite this article](#)

1394 Accesses | 2 Citations | 20 Altmetric | [Metrics](#)

Journal of Cheminformatics



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.



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

>> Learn More

Search for:

→ [Mass calculator](#)

Basic Search

Peak List

Peaks

Peak Differences

InChIKey

SPLASH

Compound Information

Compound name

Exact Mass

Tolerance

AND

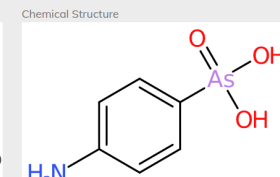
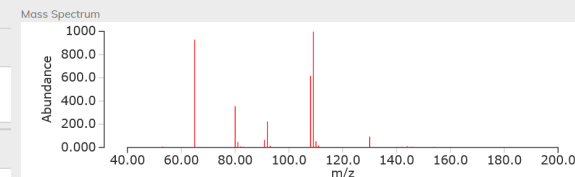
Formula (e.g. C₆H₇N₅, C₅H⁺N₅, C₅⁺)

AND

Search

MassBank Record: LU040605

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



Expert Knowledge: NORMAN Database System

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



NORMAN Database System

NORMAN

 [NORMAN WEBSITE](#) | [NORMAN DATABASE SYSTEM](#) | [HOME](#) | [LOGIN](#)

NORMAN SUBSTANCE DATABASE

SEARCH

Search

Note: C

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](https://doi.org/10.5281/zenodo.2664077)).



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



Chemical Occurrence Data

A database of geo-referenced monitoring data on emerging substances



Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmental matrices



Ecotoxicology

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



MassBank Europe

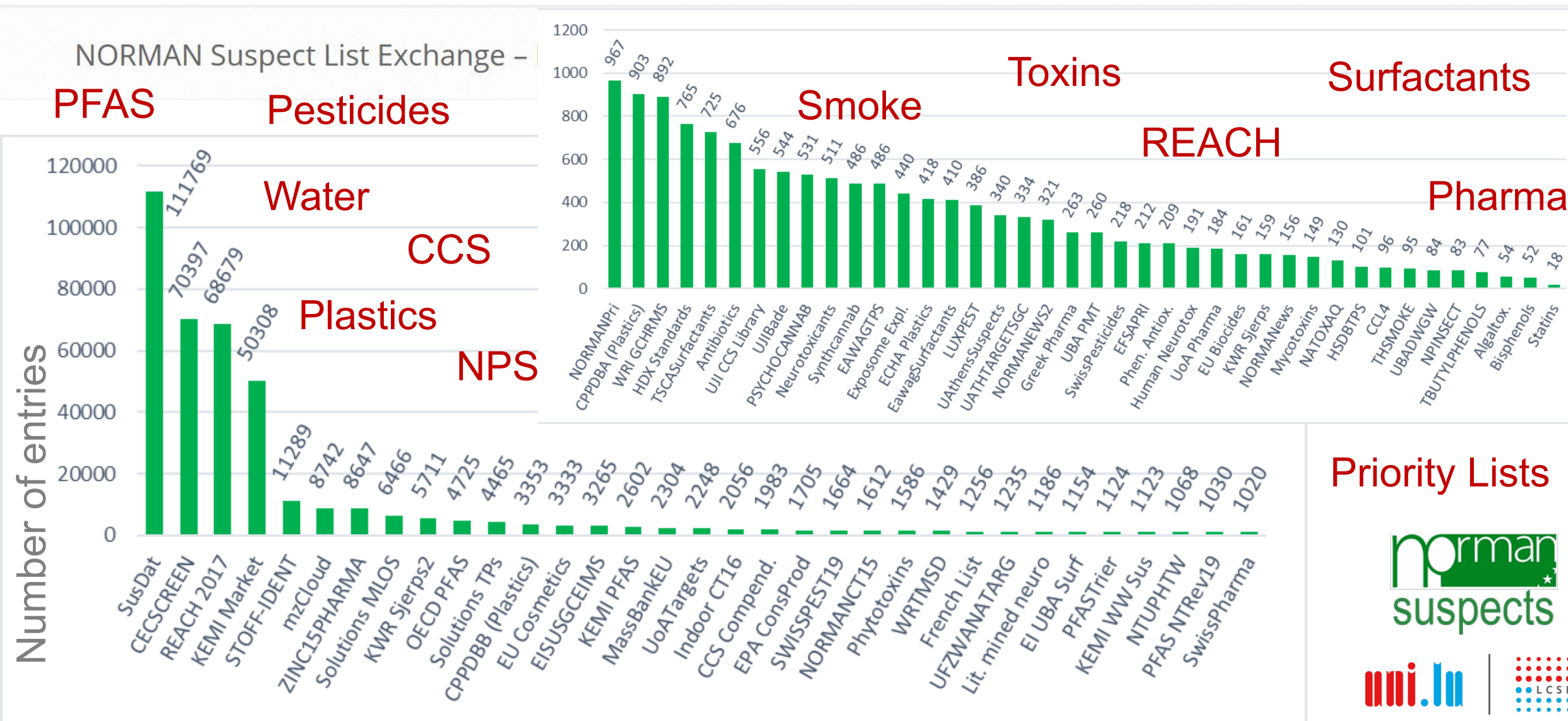
A database of mass spectra of emerging substances to support identification of unknown substances



Expert Knowledge: NORMAN Suspect List Exchange (>80 lists!)

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

<https://zenodo.org/communities/norman-sle>

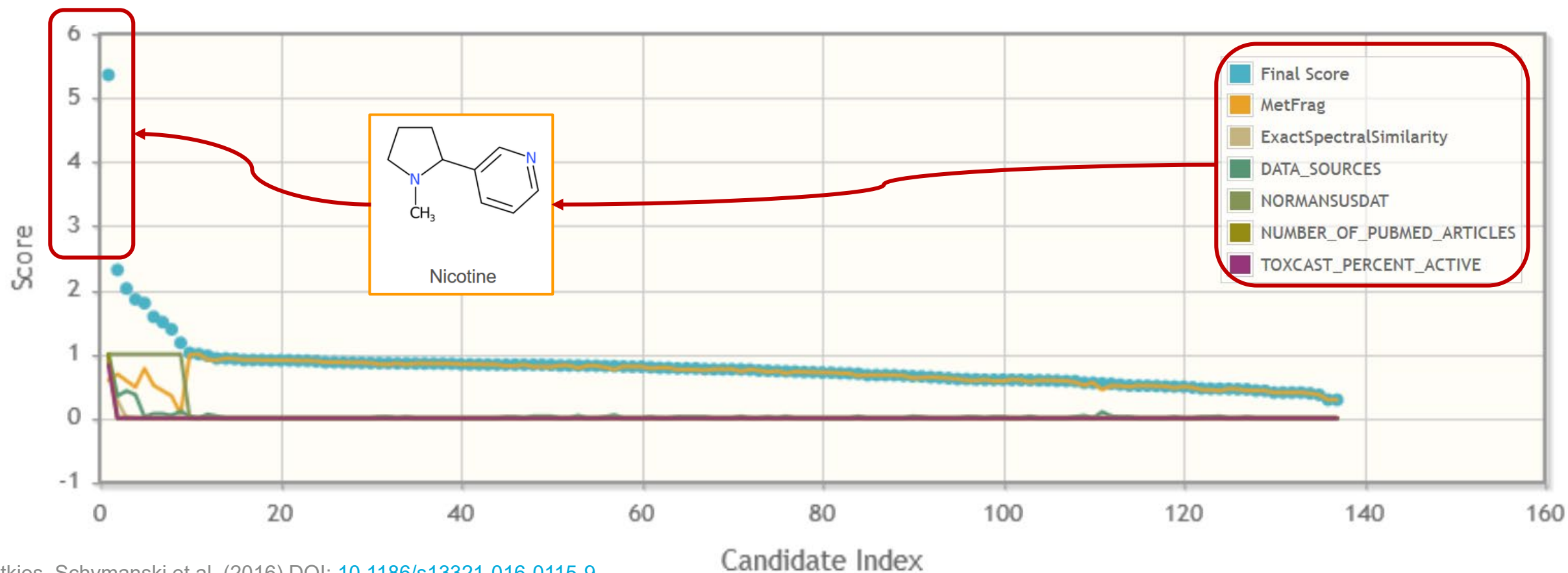


Connecting Knowledge for Chemical Identification: MetFrag



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

<http://ipb-halle.github.io/MetFrag/>

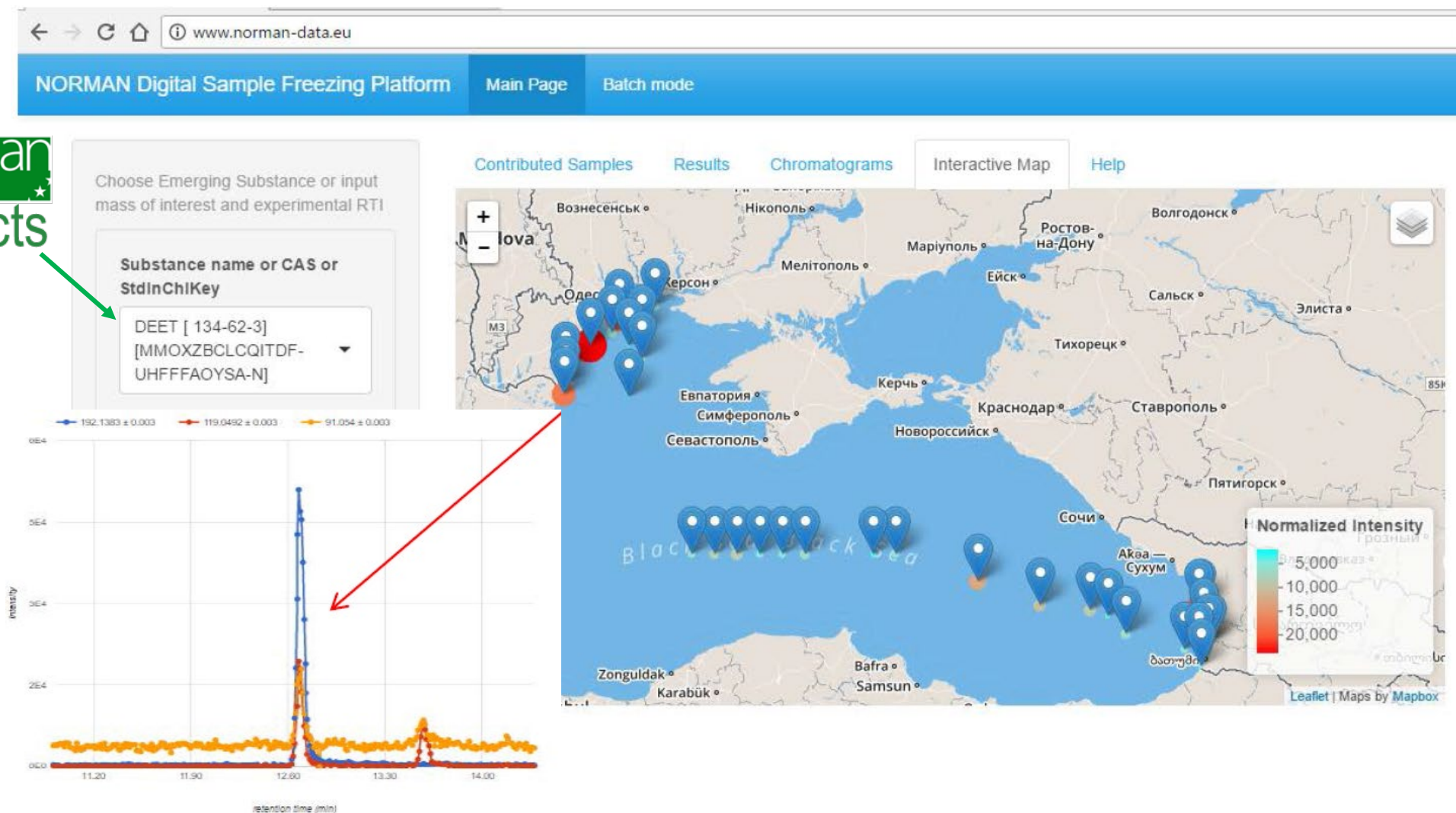


Examples: NORMAN Digital Sample Freezing Platform

“Live” retrospective screening of known and unknown chemicals in European samples

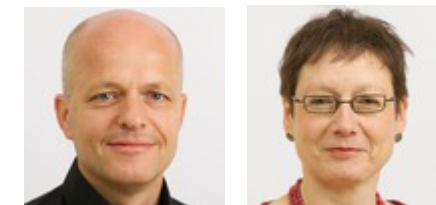


norman
suspects

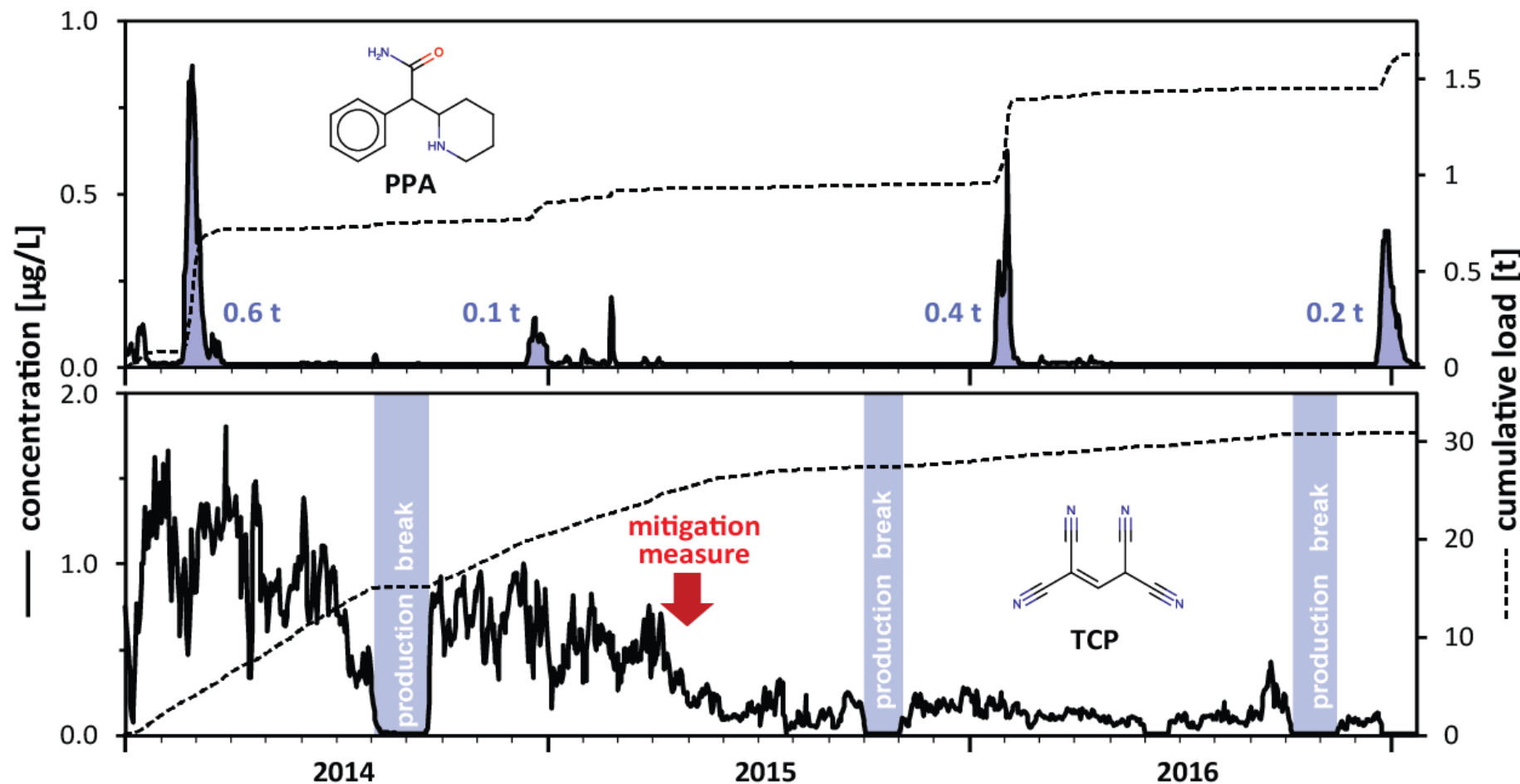


Examples: Real Time Monitoring of the Rhine River

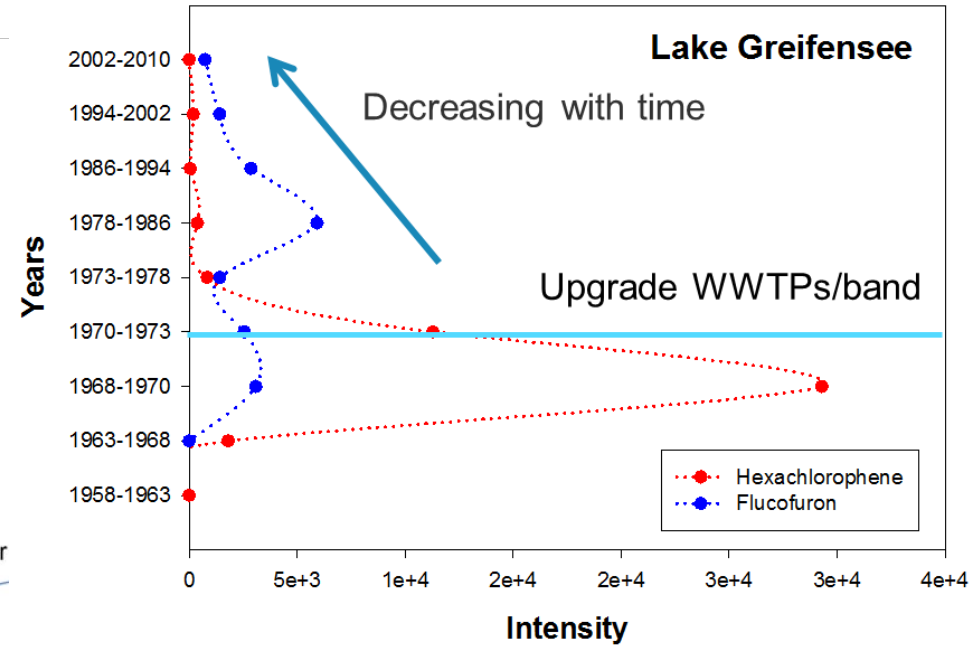
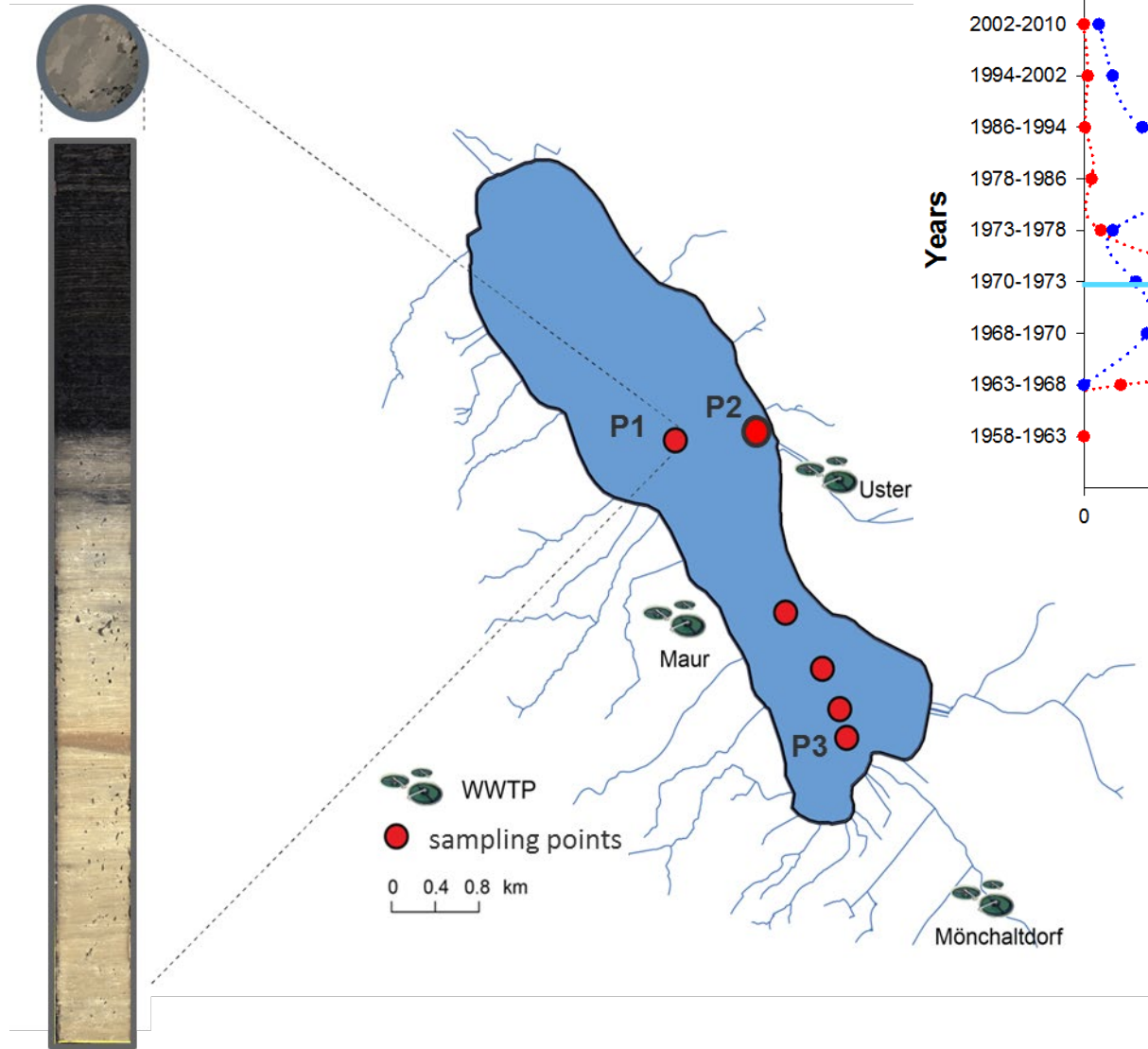
Previously unknown chemicals detected due to “stand-out” patterns



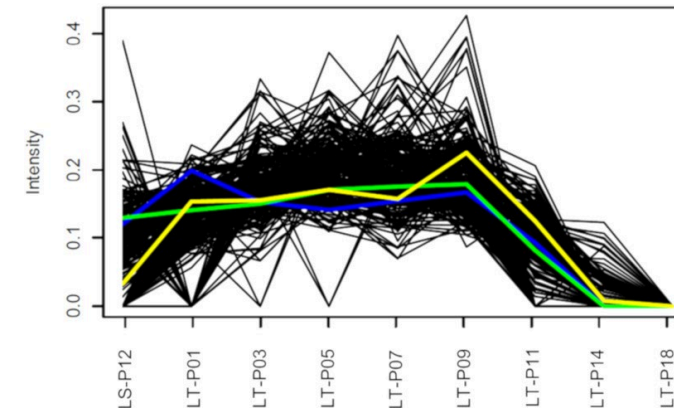
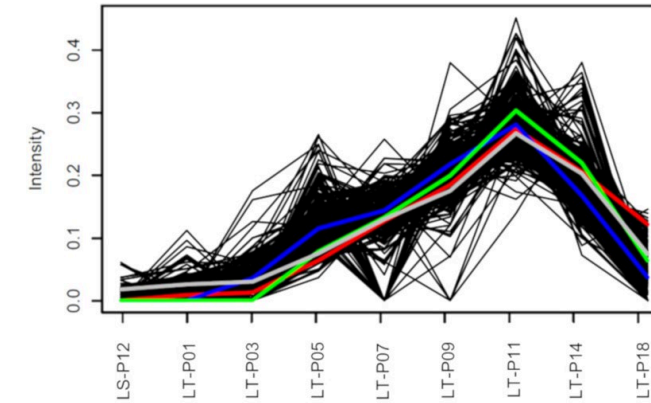
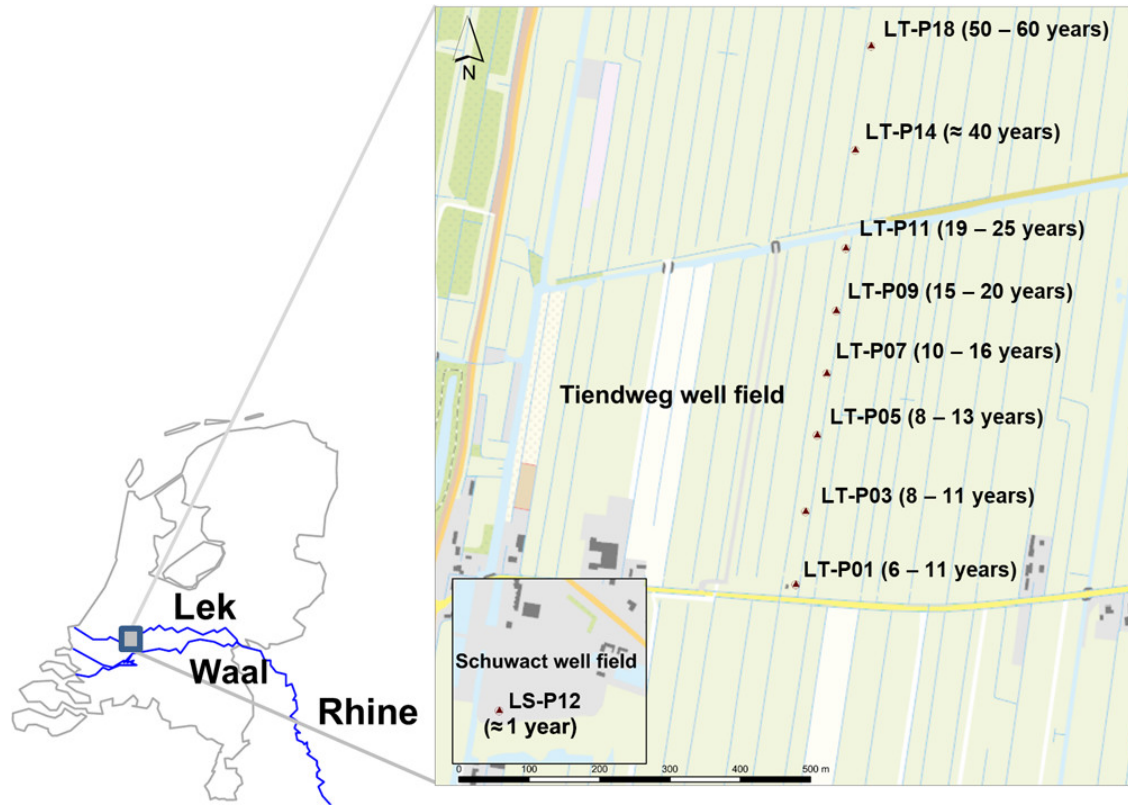
eawag
aquatic research



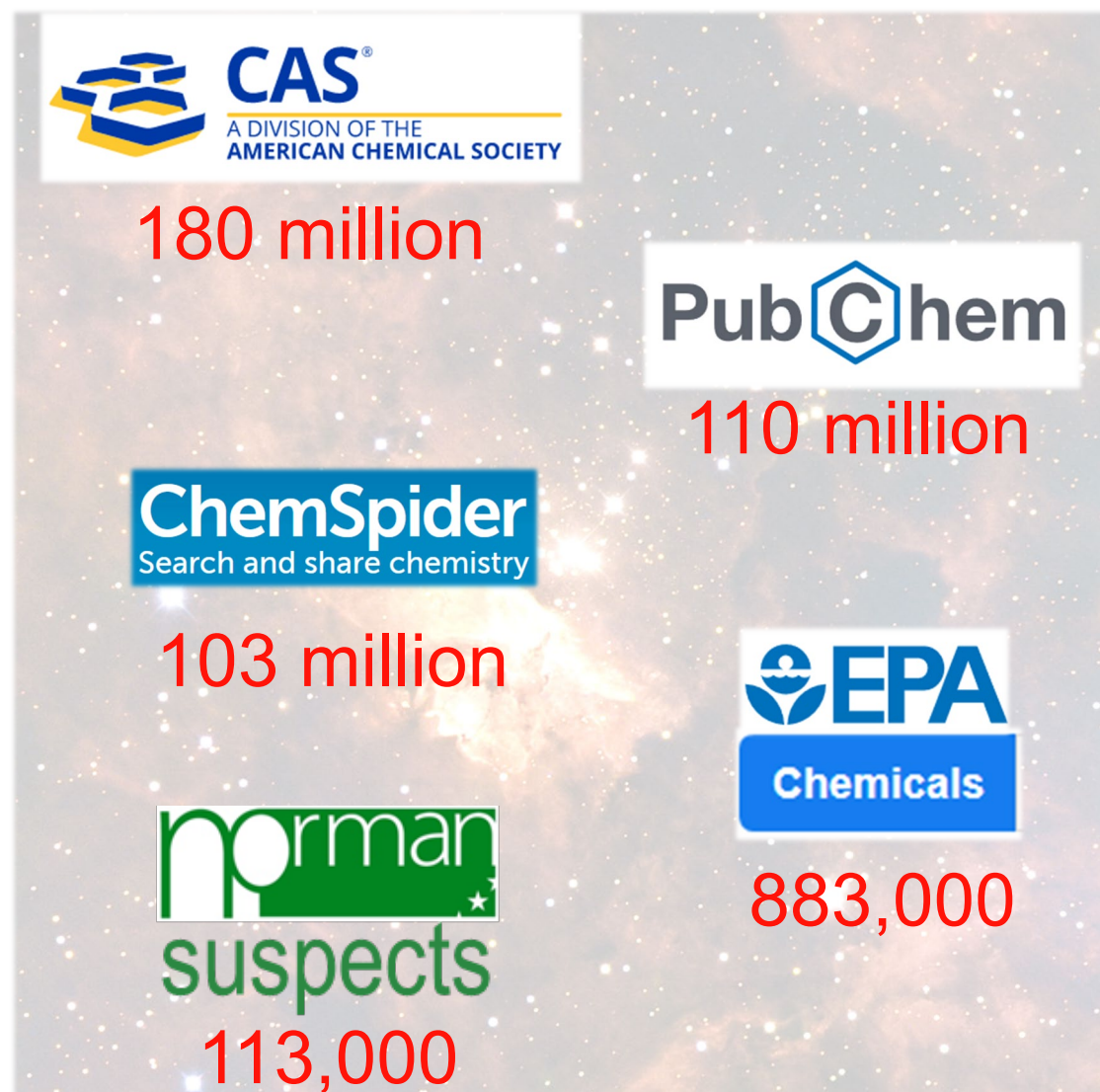
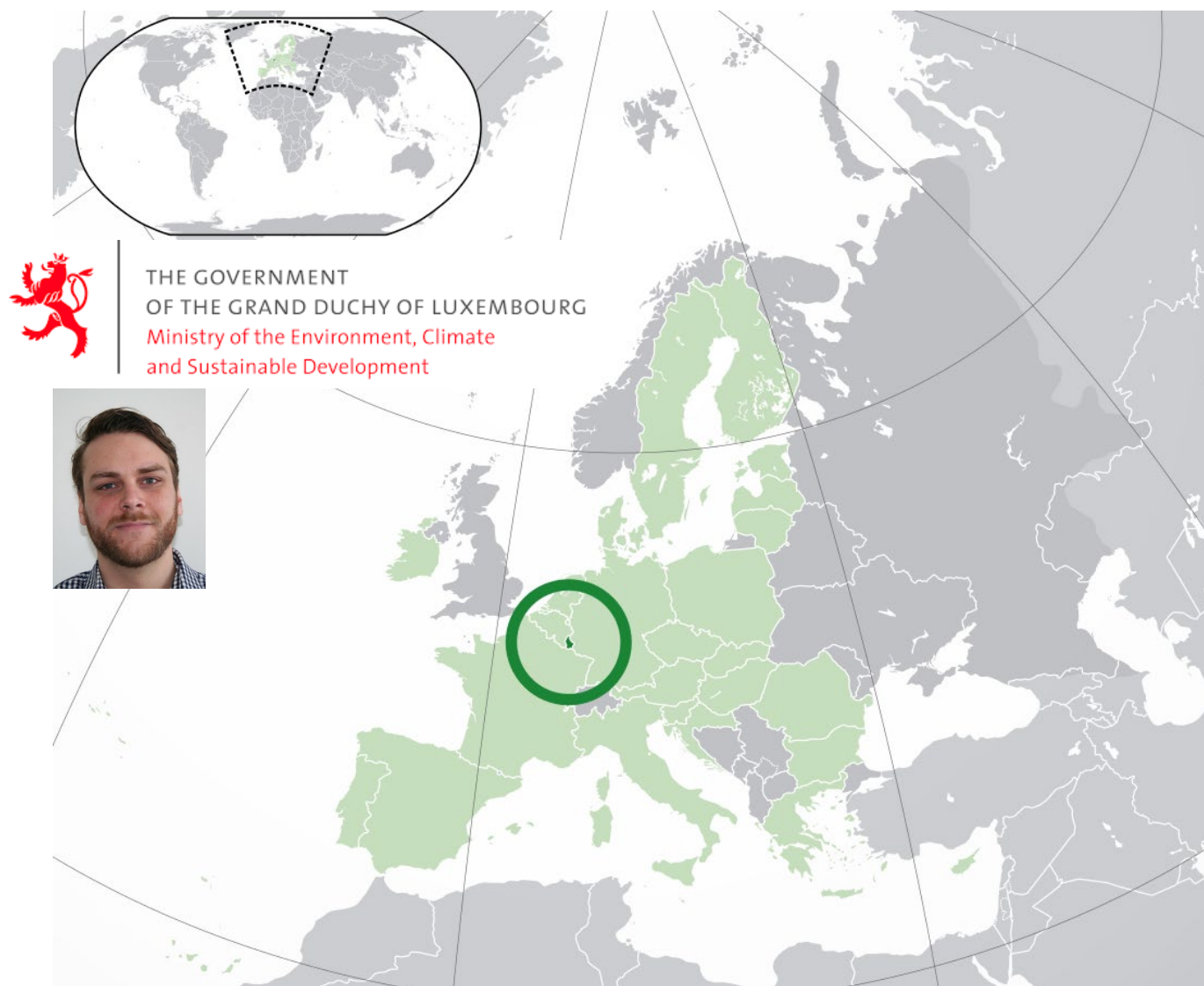
Examples: Historical Contamination in Lake Sediments



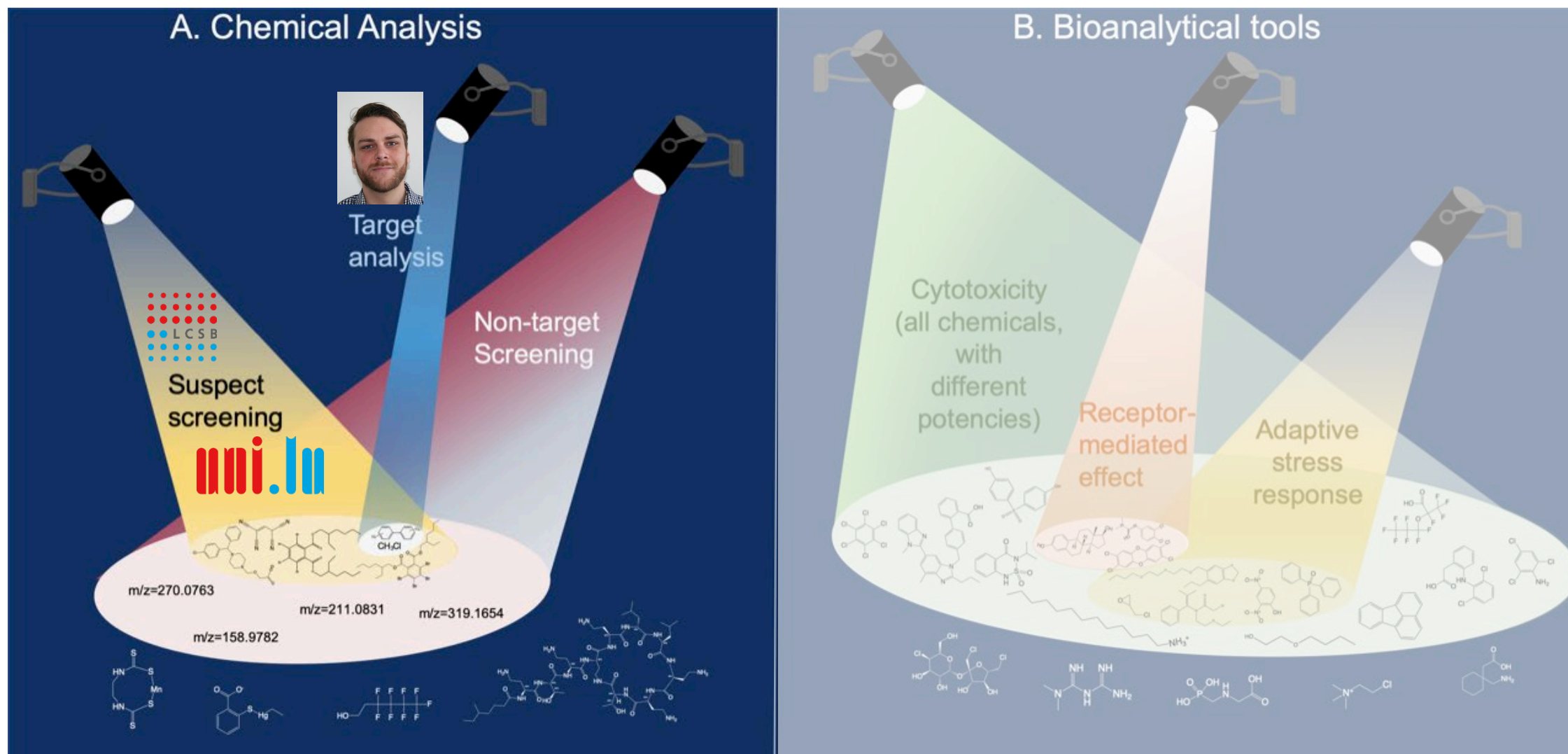
Examples: Time Trends in Riverbank Filtration Systems



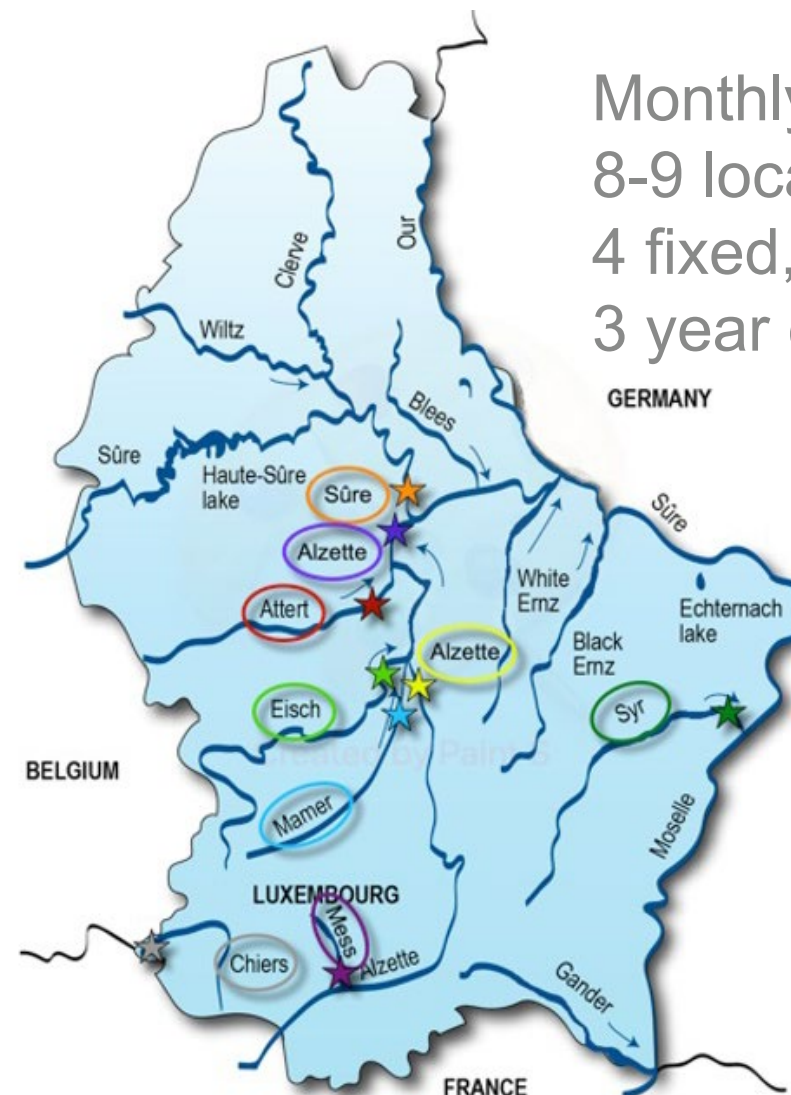
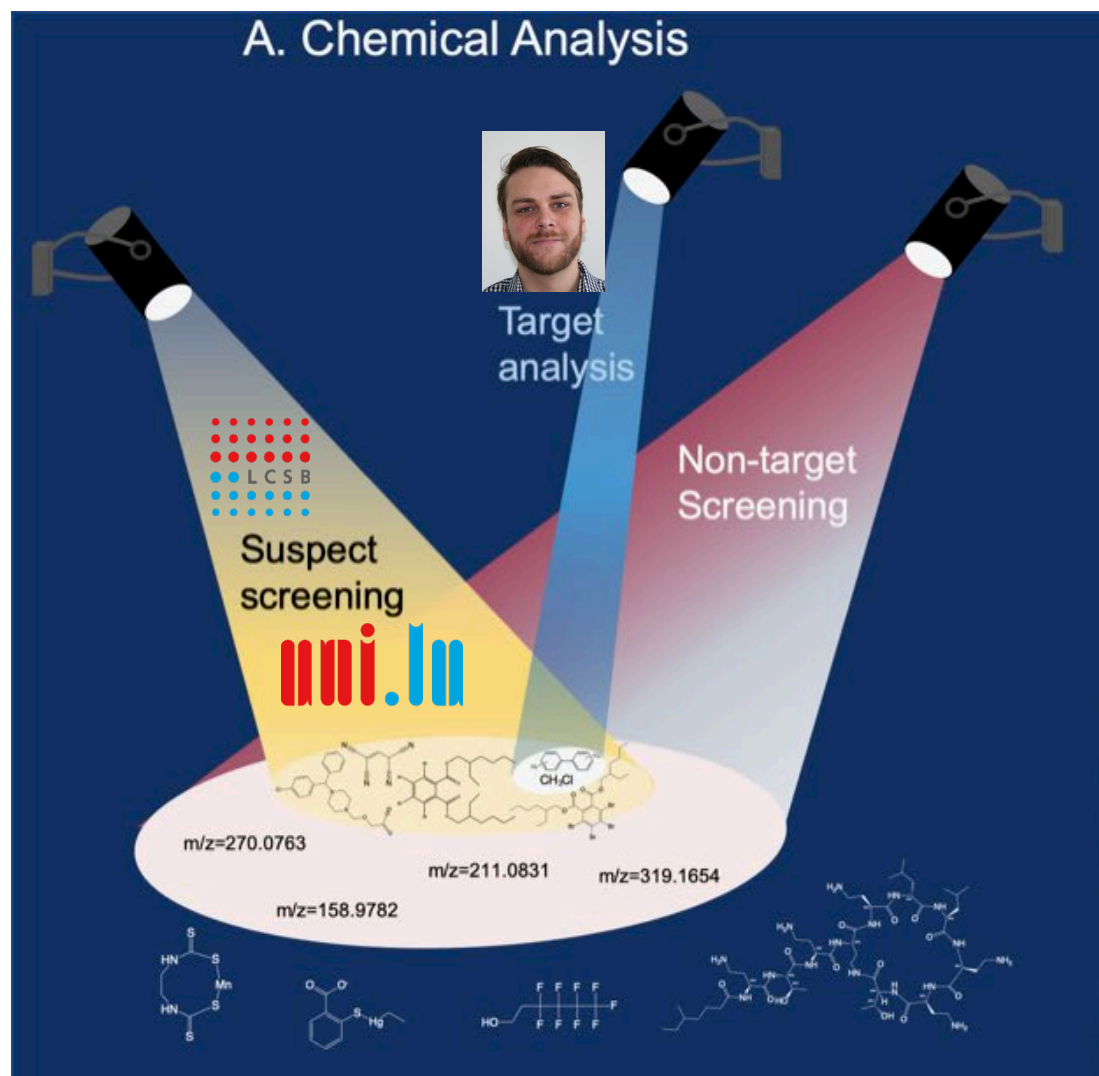
The Problem: Which chemicals are relevant? How to find them?



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



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

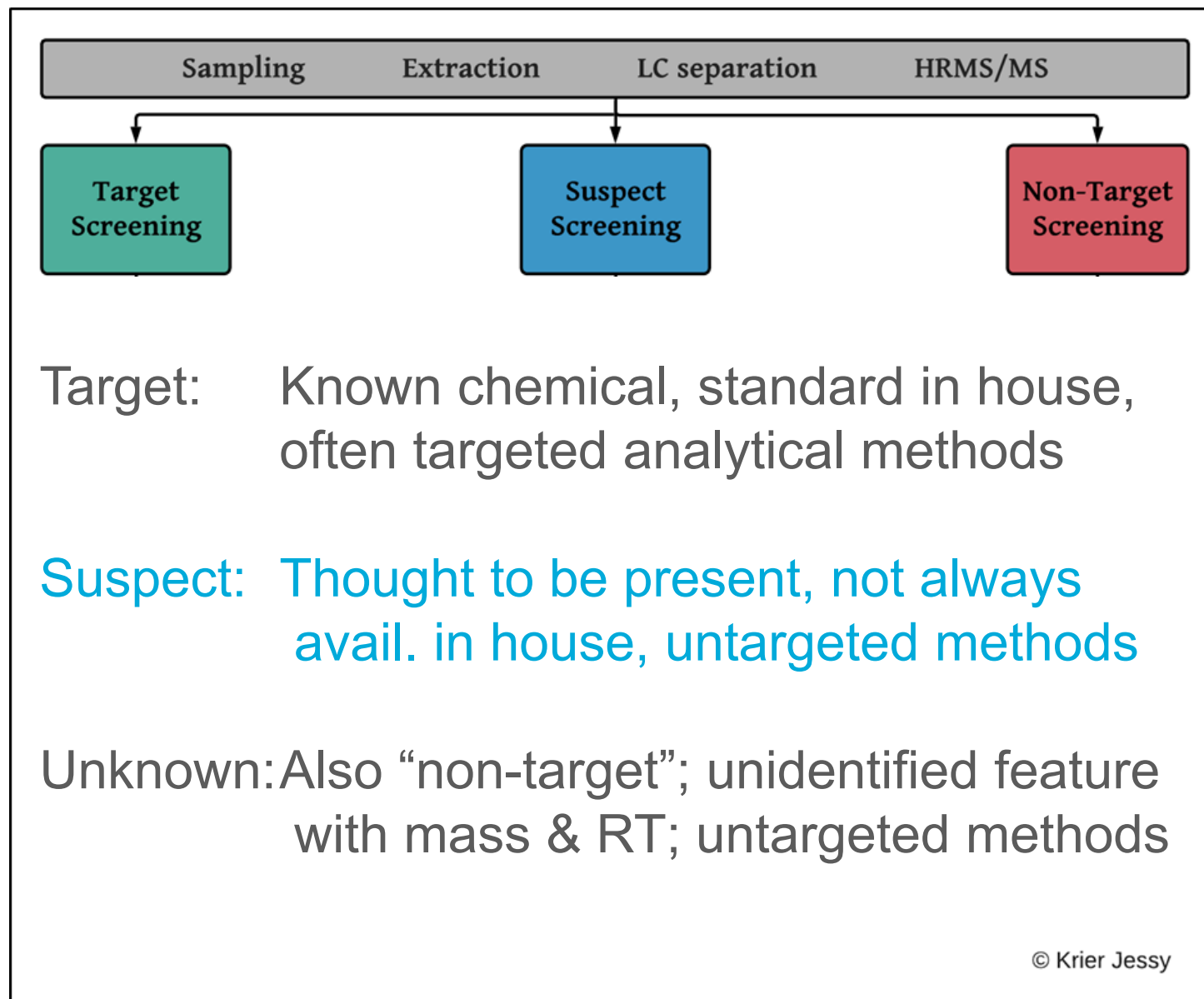


Monthly sampling
8-9 locations / year
4 fixed, 4-5 rotate
3 year cycle

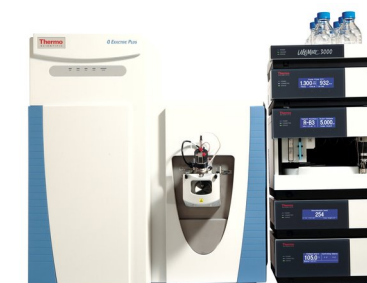
Targets, Suspects and Non-targets



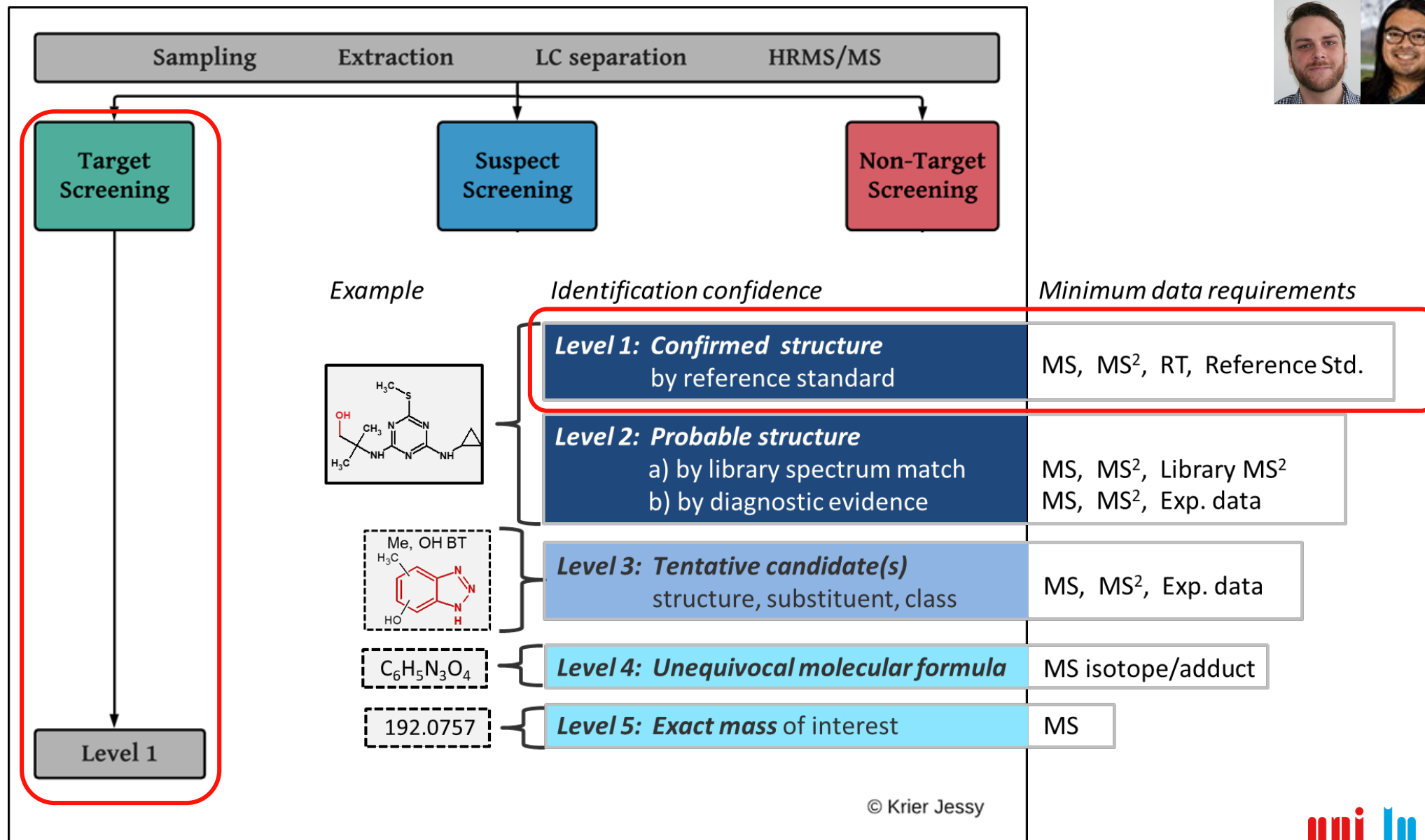
THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development



Extraction at AGE
RP (C₁₈) pos/neg
Generic NT method



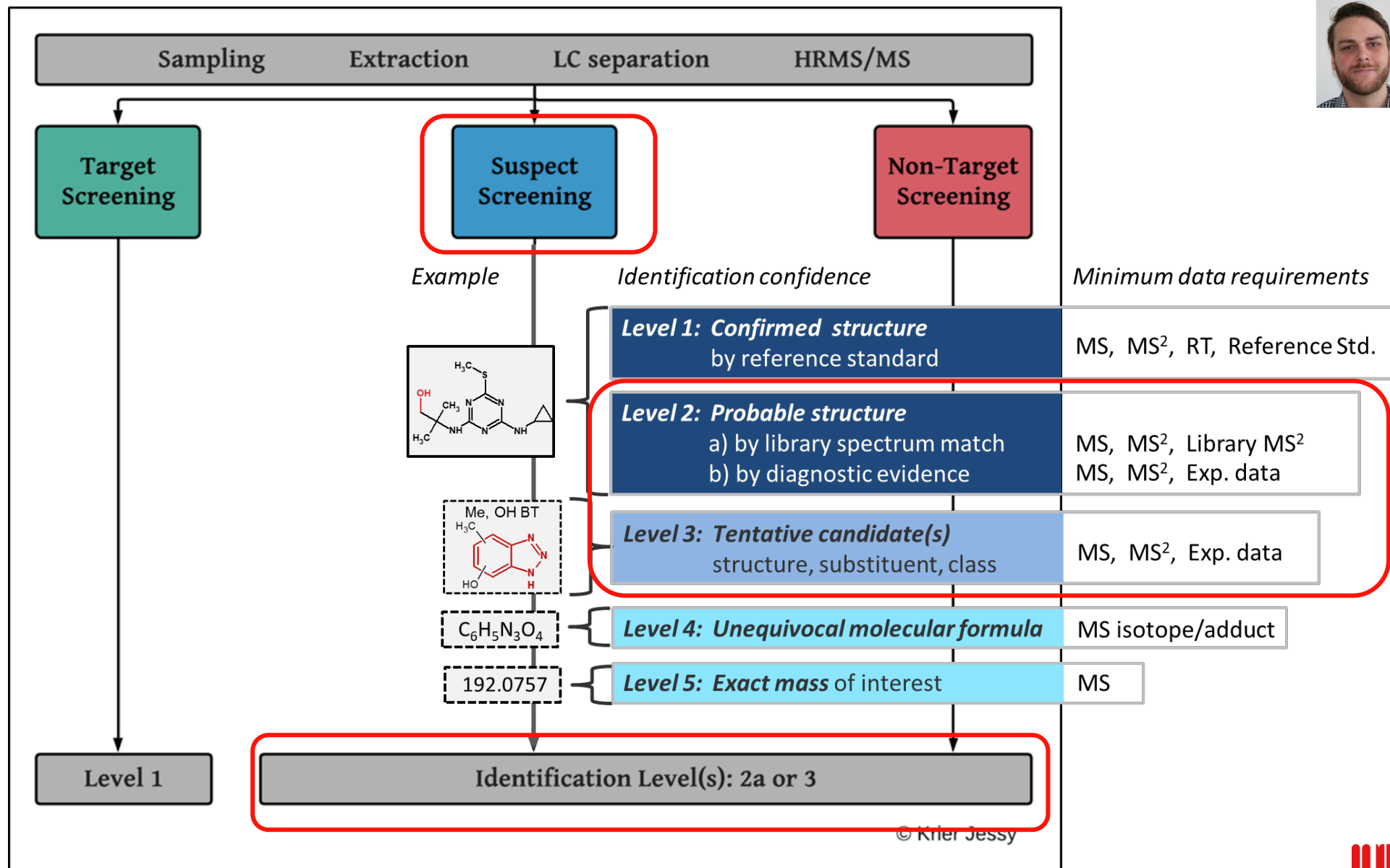
Targets, Suspects and Non-targets – and ID Levels



Targets, Suspects and Non-targets – and ID Levels



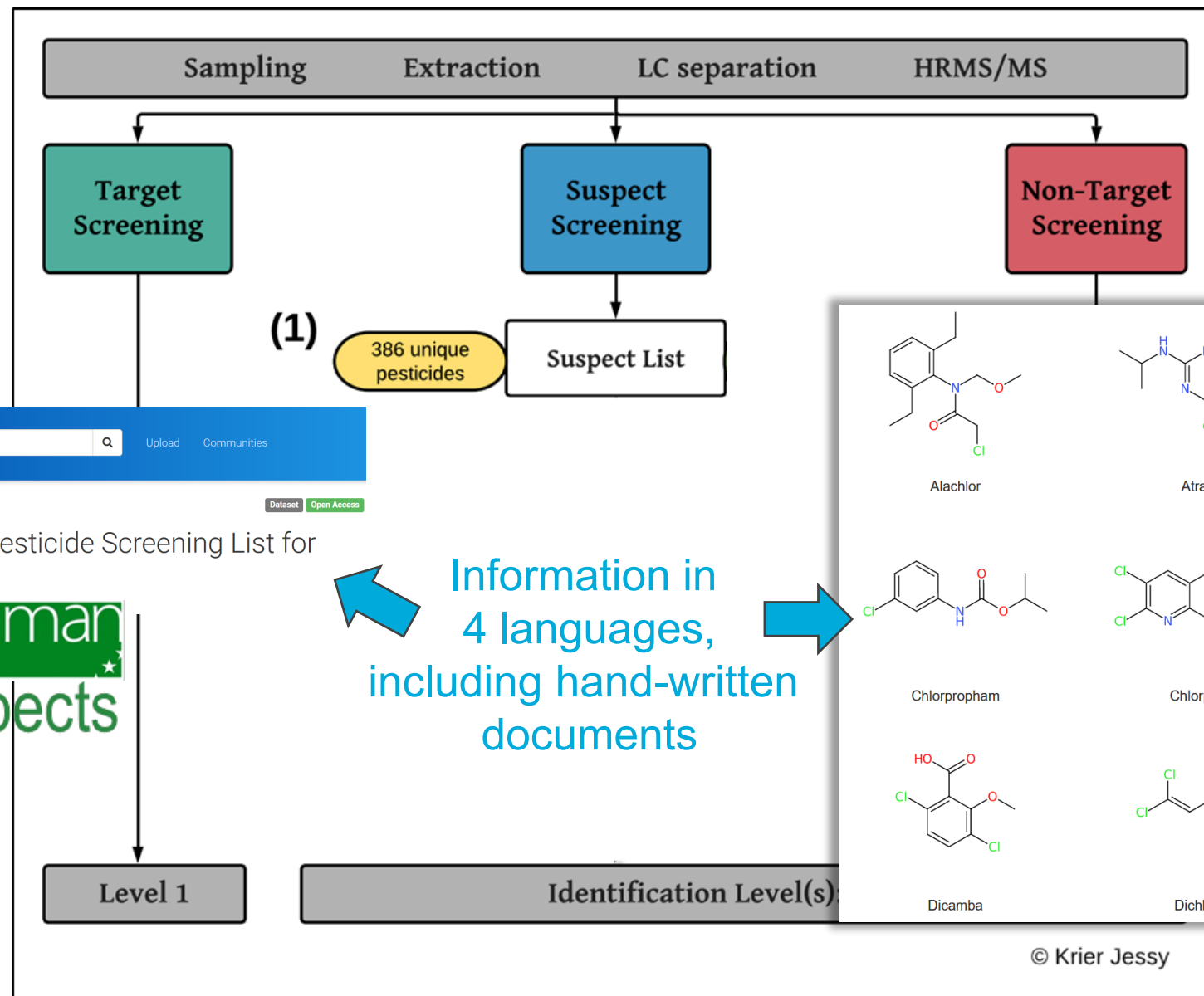
THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development



LuxPest – Suspect List Generation



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



(1)

386 unique
pesticides

Suspect List

zenodo

Search

q

Upload

Communities

Dataset Open Access

May 28, 2020

S69 | LUXPEST | Pesticide Screening List for
Luxembourg

Krier, Jessy

norman
suspects

Level 1

Identification Level(s)

Information in
4 languages,
including hand-written
documents

Alachlor

Atrazine

Bentazone

Carbofuran

Chlorpropham

Chlorpyrifos

Cypermethrin

Diazinon

Dicamba

Dichlorvos

Dimethoate

Diuron

© Krier Jessy



LuxPest - Classification

zenodo

Search



Upload

Communities

Log in

Sign up

May 28, 2020

Dataset Open Access

828

views

733

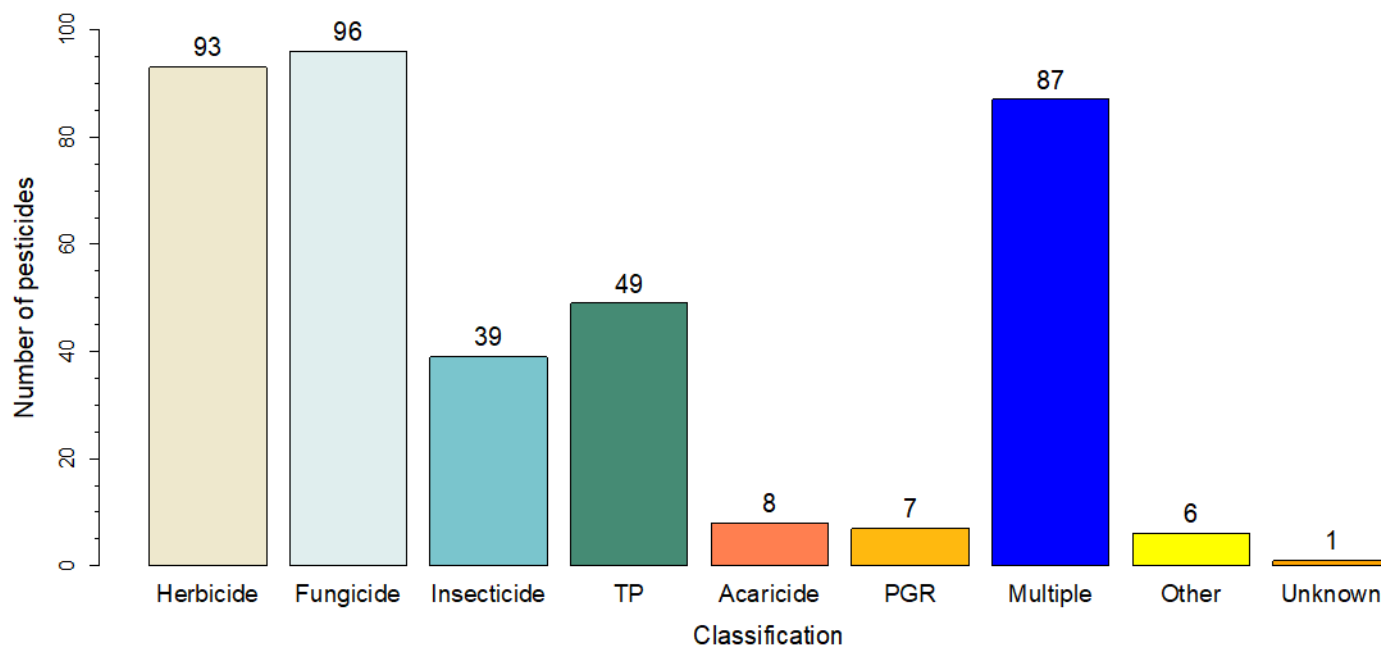
downloads

See more details...

S69 | LUXPEST | Pesticide Screening List for Luxembourg

Krier, Jessy

Other(s)



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

Browse NORMAN Suspect List Exchange Tree

NORMAN Suspect List Exchange Classification ? 113,235

S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) a Inventory (2006) ? 3,983

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 ?

S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ? 1,344

S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ?

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

Authorization status in Luxembourg ? 386

No information 21

Not permitted 169

Permitted 196

Acaricides 43

Adjuvants 4

Algicides 2

Algists 1

Antimicrobials 2

Bactericides 7

Bird repellents 1

Fungicides 119

Herbicides 116

Insect attractants 4

Insecticides 91

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

PubChem

mini

LC SB

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

Krier *et al* (in review). Preprint DOI: [10.21203/rs.3.rs-478324/v1](https://doi.org/10.21203/rs.3.rs-478324/v1). Jessy Krier (2020) S69 | LUXPEST. DOI: [10.5281/zenodo.3862689](https://doi.org/10.5281/zenodo.3862689)

LuxPest - Classification



About Blog Submit Contact

SEARCH FOR

NORMAN Suspect List Exchange: S69 | LUXPEST | Pesticide Screening List for

Treating this as a previously computed list of identifiers.

Compounds

386 results

Filters

SORT BY



3-hydroxybenzaldehyde; 100-83-4; M-Benzaldehyde, 3-hydroxy-; M-Formylp

Compound CID: 101

MF: $C_7H_6O_2$ MW: 122.12g/mol

InChIKey: IAVREABSGIHHMO-UHFFFAOYSA-N

IUPAC Name: 3-hydroxybenzaldehyde

Create Date: 2004-09-16



Acetic Acid; Ethanoic Acid; 64-19-7; Et

Compound CID: 176

MF: $C_2H_4O_2$ MW: 60.05g/mol

InChIKey: QTBSBXVTEAMEQO-UHFFFAOYSA-N

PubChem Diazinon (Compound)

8 Agrochemical Information

8.1 Agrochemical Category

Acaricides, Insecticides

► EU Pesticides Database

Repellents, Veterinary substances, Acaricides, Insecticides

S69 | LUXPEST | Pesticide Screening List for Luxembourg | DOI:10.5281/zenodo.3862688

► NORMAN Suspect List Exchange

Pesticides -> Insecticides -> Organophosphorus insecticides -> Organothiophosphate insecticides -> Heterocyclic organothiophosphate insecticides

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

► NORMAN Suspect List Exchange

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

Browse NORMAN Suspect List Exchange Tree

NORMAN Suspect List Exchange Classification ? 113,235

S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386

Authorization status in Luxembourg ? 386

No information 21

Not permitted 169

Permitted 196

Acaricides 43

Adjuvants 4

Algicides 2

Algistsats 1

Antimicrobials 2

Bactericides 7

Bird repellents 1

Fungicides 119

Herbicides 116

Insect attractants 4

Insecticides 91

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

PubChem



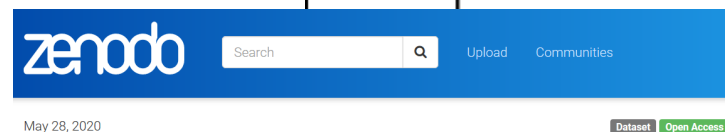
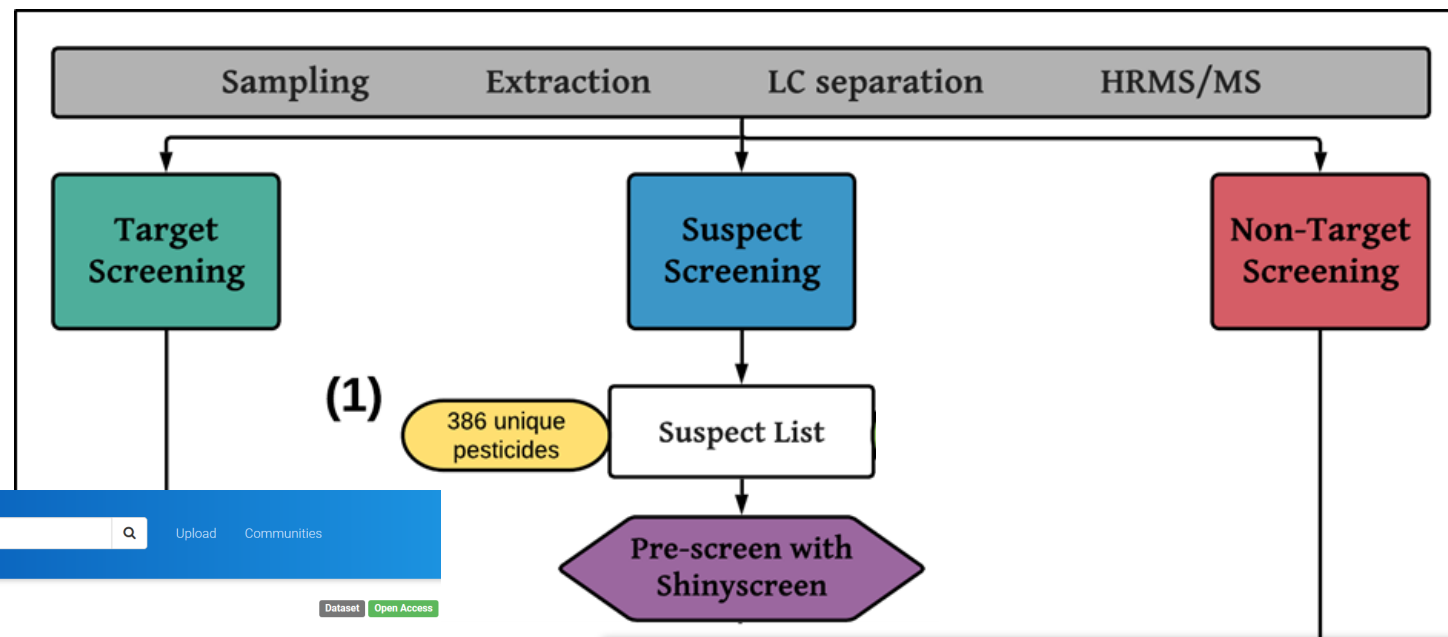
<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101>

Krier *et al* (in review). Preprint DOI: [10.21203/rs.3.rs-478324/v1](https://doi.org/10.21203/rs.3.rs-478324/v1). Jessy Krier (2020) S69 | LUXPEST. DOI: [10.5281/zenodo.3862689](https://doi.org/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



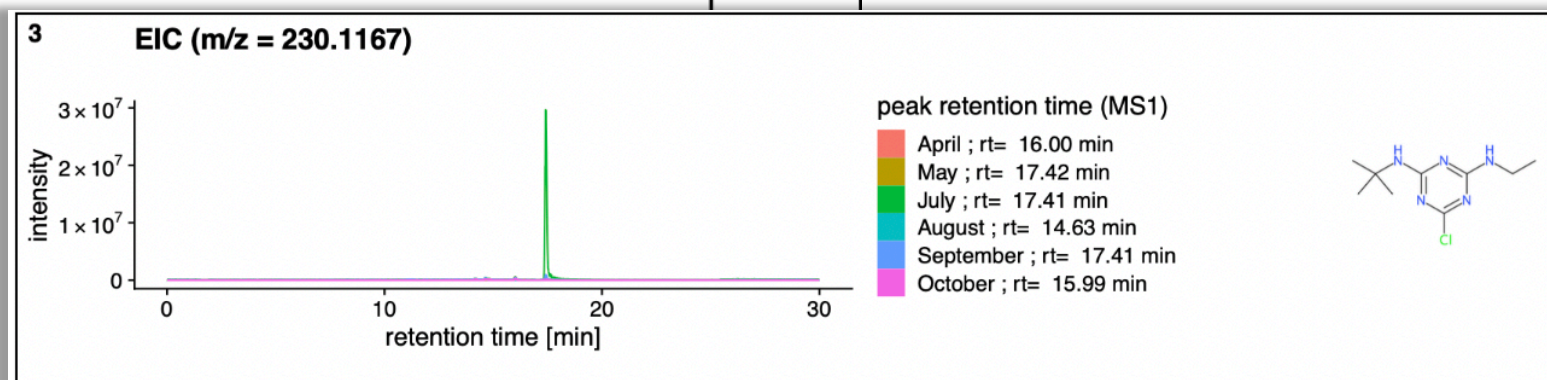
S69 | LUXPEST | Pesticide Screening List for
Luxembourg

Krier, Jessy



Level 1

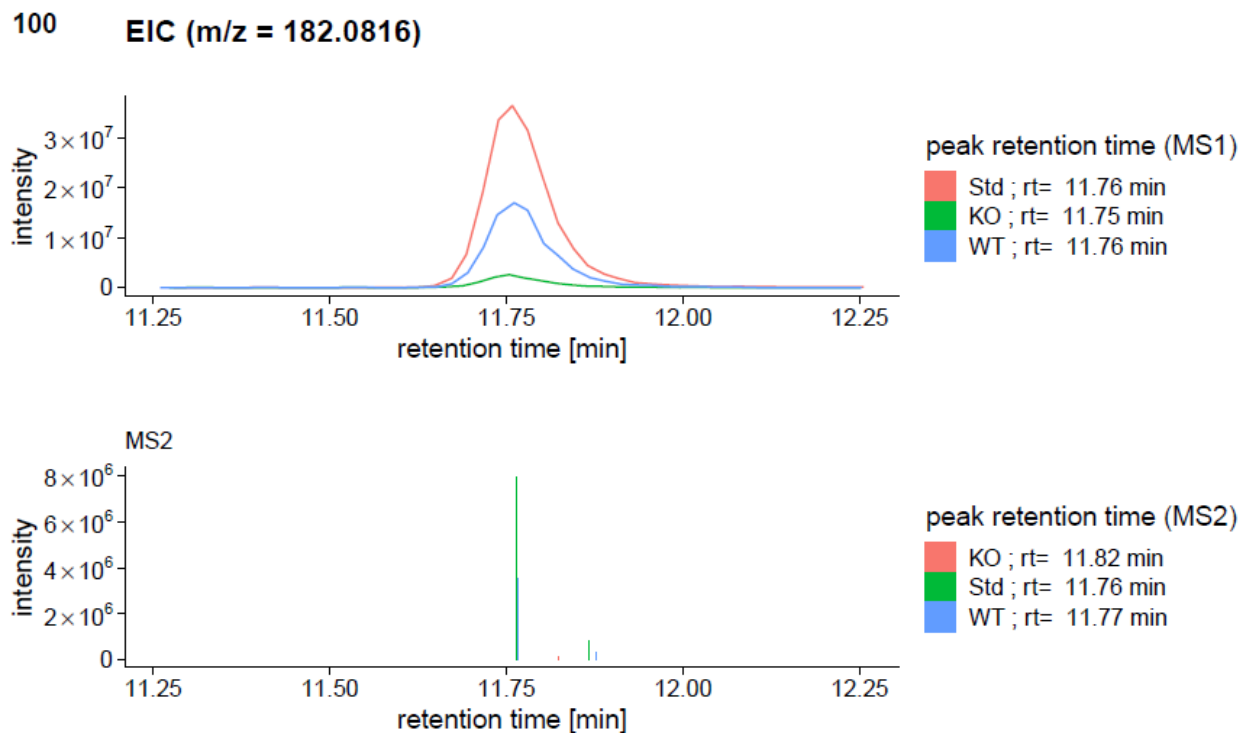
Identification Level(s): 2a or 3



© Krier Jessy



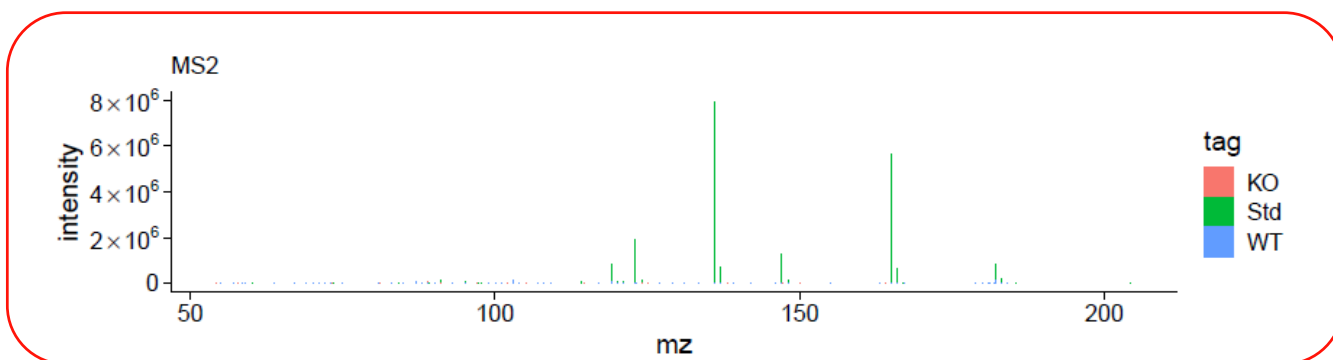
Open Source Pre-screening Workflow (+QC): ShinyScreen



<https://git-r3lab.uni.lu/eci/shinyScreen>



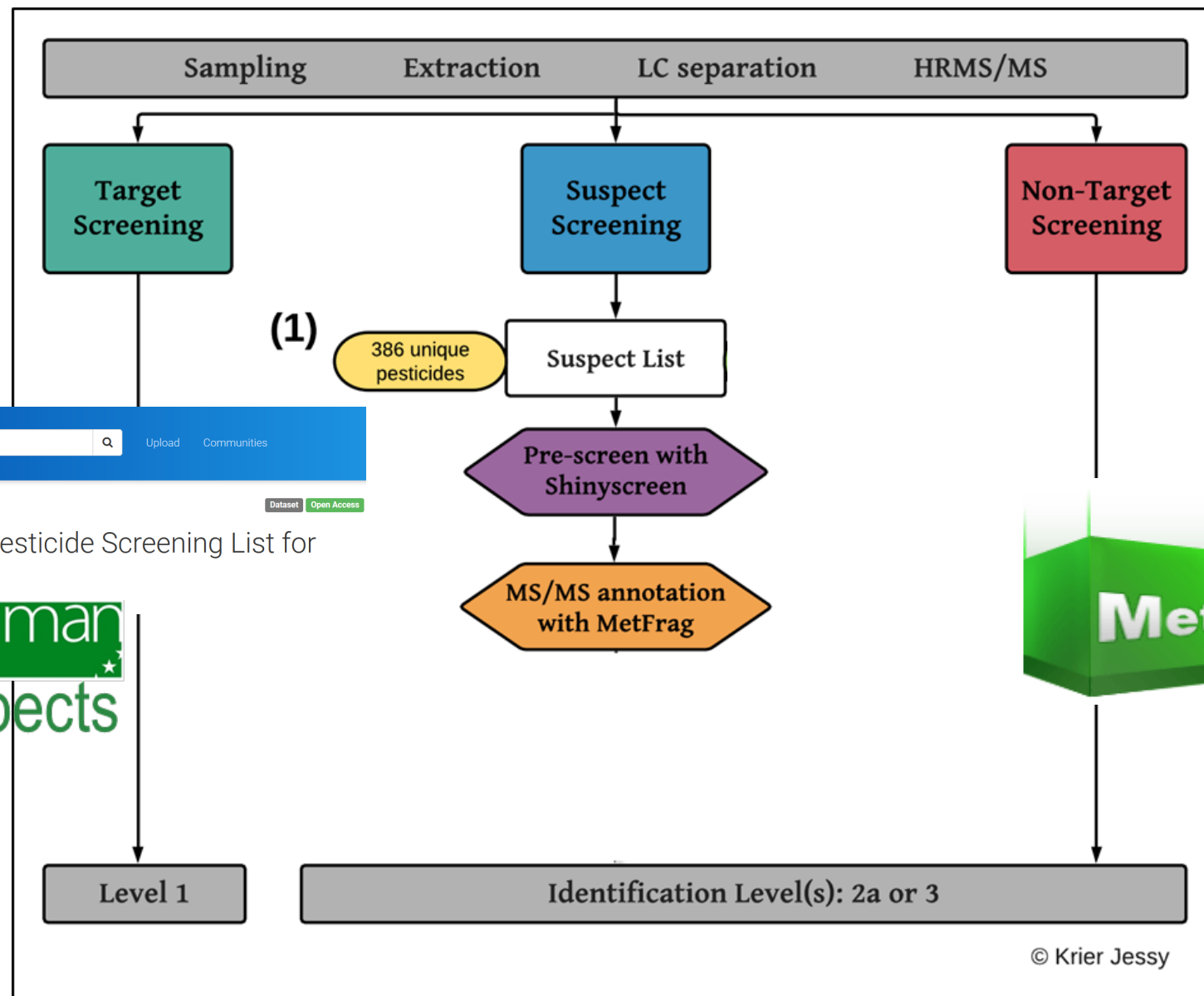
Anjana Elapavalore, Mira Narayanan,
Todor Kondic, Jessy Krier,
Hiba Mohammed Taha.



LuxPest – MS/MS Annotation with MetFrag



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



May 28, 2020

Dataset Open Access

S69 | LUXPEST | Pesticide Screening List for
Luxembourg

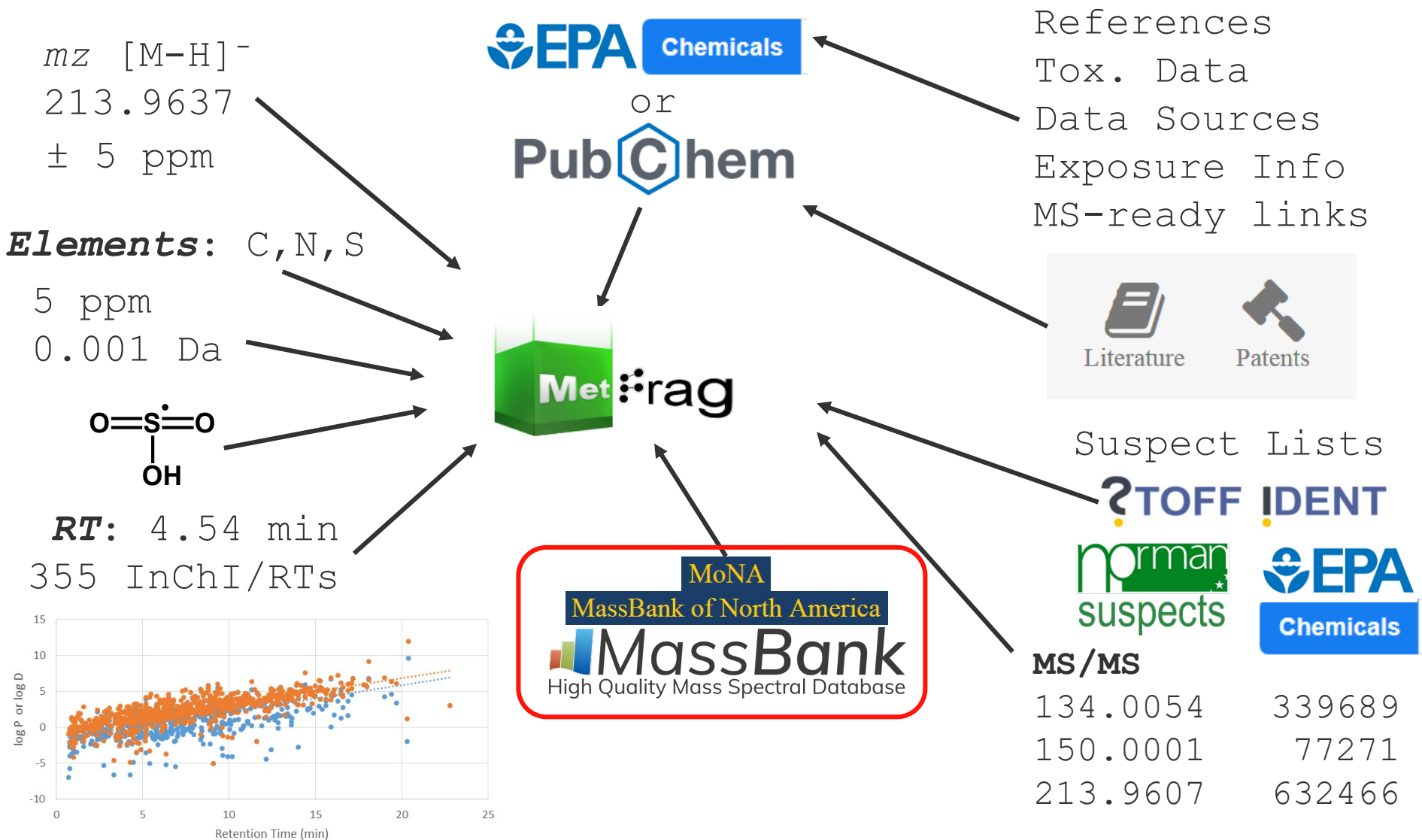
Krier, Jessy



© Krier Jessy



Annotation with MetFrag (Relaunched) ...



eawag
aquatic research

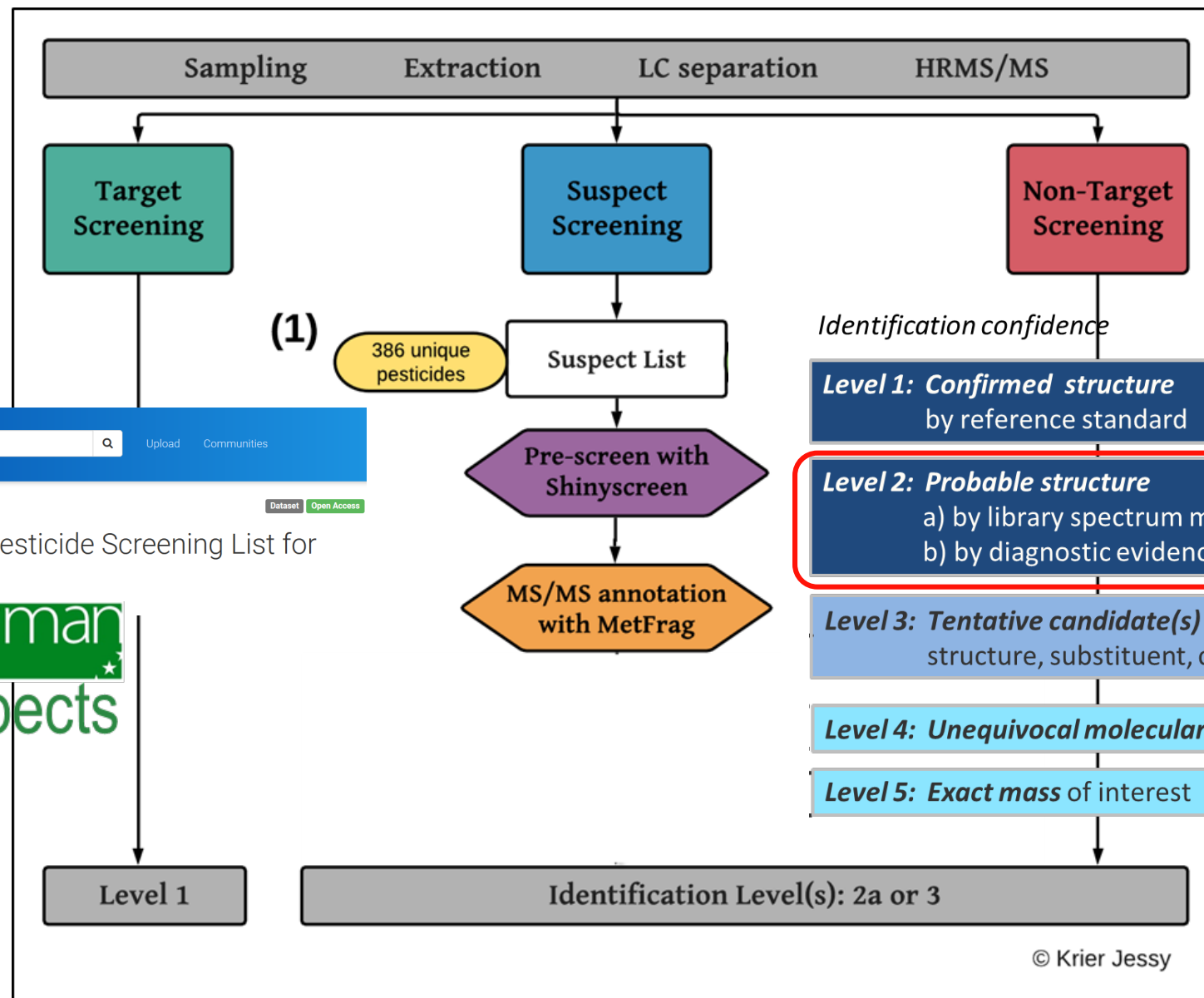
solutions



LuxPest – MS/MS Annotation with MetFrag



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



Dataset Open Access

May 28, 2020

S69 | LUXPEST | Pesticide Screening List for Luxembourg

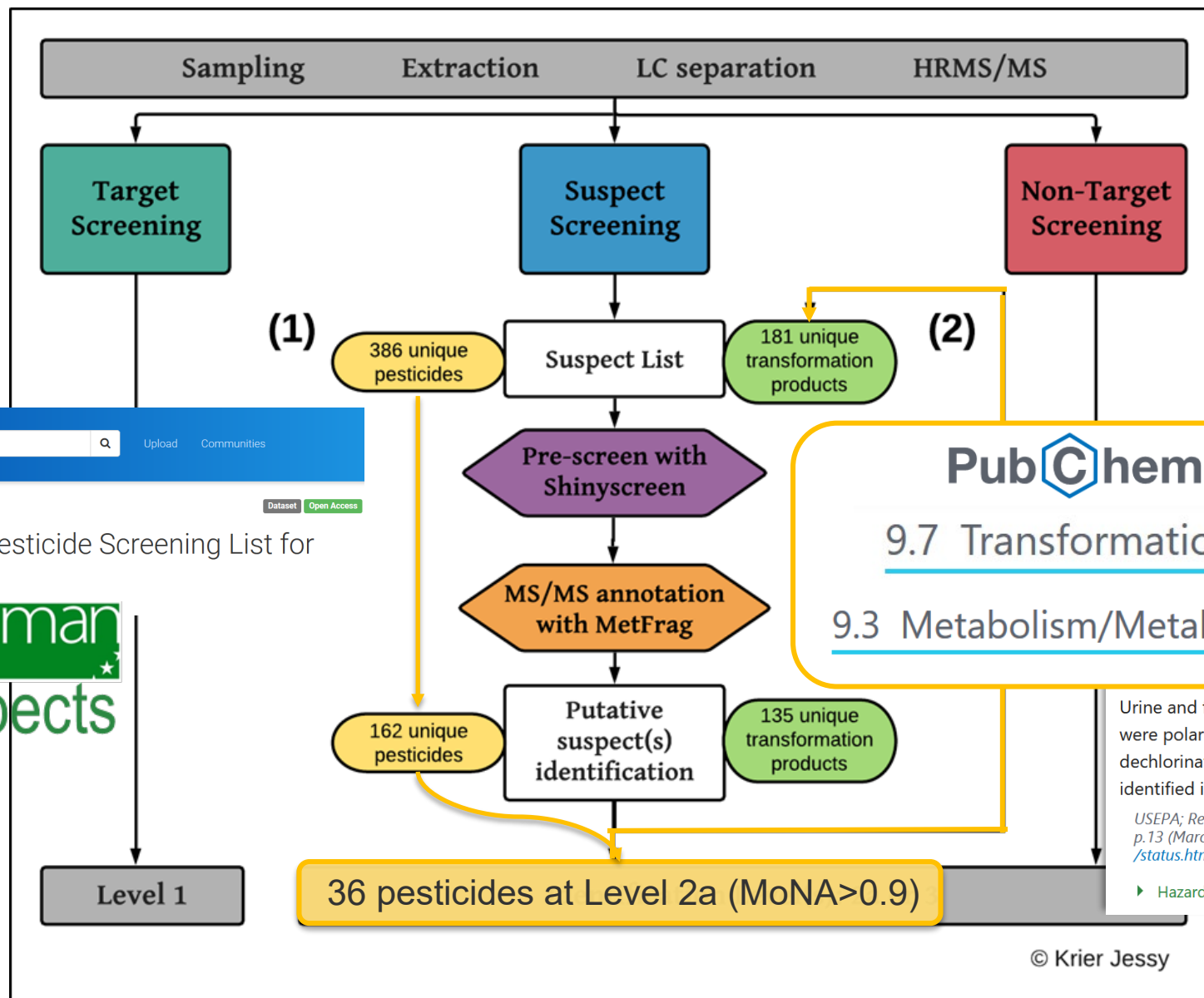
Krier, Jessy



LuxPest - Overview



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



PubChem Ammeline (Compound)

	Atrazine	Mammalian metabolism		Ammeline
	Simazine	Plant metabolism		Ammeline

PubChem

9.7 Transformations

9.3 Metabolism/Metabolites

Schymanski *et al.* (2021)
DOI: [10.1186/s13321-016-0115-9](https://doi.org/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 Terbutylazine (5915-41-3). EPA 738-R-95-005 p.13 (March 1995). Available from, as of October 11, 2012: <http://www.epa.gov/pesticides/red/status.htm>

► Hazardous Substances Data Bank (HSDB)



© Krier Jessy



Dataset Open Access

May 28, 2020

S69 | LUXPEST | Pesticide Screening List for Luxembourg

Krier, Jessy



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

PubChem Terbutylazine (Compound)

8.3 Metabolism/Metabolites



Metabolism of terbutylazine 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 Terbutylazine (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.

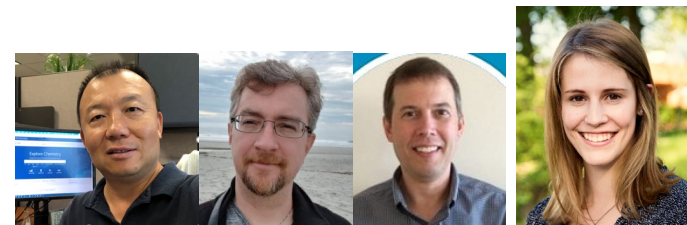
USEPA; Reregistration Eligibility Decision (RED) Database for Terbutylazine (5915-41-3). EPA 738-R-95-005 p.13 (March 1995). Available from, as of October 11, 2012: <http://www.epa.gov/pesticides/reregistration/status.htm>

► Hazardous Substances Data Bank (HSDB)

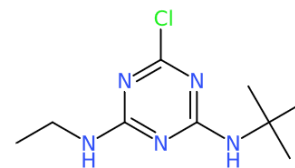
In mammals, following oral administration, ...a de-ethyl metabolite forms rapidly, followed by conjugates of products formed by oxidation of one methyl group of the tert-butyl moiety. All are rapidly excreted.

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

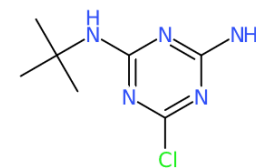
► Hazardous Substances Data Bank (HSDB)



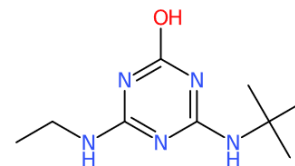
CDK
DEPICT



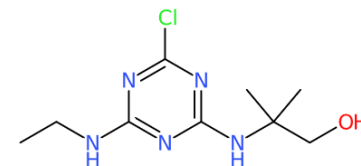
Terbutylazine CID:22206



desethyl-terbutylazine CID:108201



2-hydroxy-terbutylazine CID:135495928



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

Krier et al (in review). Preprint DOI: [10.21203/rs.3.rs-478324/v1](https://doi.org/10.21203/rs.3.rs-478324/v1) &
Schymanski et al. (2021) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)



“Living data connections”

[Upload](#)[Communities](#)

June 11, 2020

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](https://doi.org/10.5281/zenodo.3827487).

Preview

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

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

LCSB-ECI & PubChem Team. DOI [10.5281/zenodo.3890392](https://doi.org/10.5281/zenodo.3890392)

File Edit View Repository Branch Help

Current repository pubchem Current branch master Fetch origin Last fetched 2 minutes ago

Changes 2 History

No branches to compare

Update extractAnnotations.R Emma Schymanski • Jun 9, 2020

HSDB Ref Info Emma Schymanski • Jun 8, 2020

added new CIDs to HSDBTPS Emma Schymanski • Jun 8, 2020

Update PCLite_eval_support.R Emma Schymanski • Jun 8, 2020

Added S69 LUXPEST Emma Schymanski • May 28, 2020

added new CIDs to HSDBTPS Emma Schymanski • 13fdb18 5 changed files Hide Whitespace

Added newly registered CIDs to base HSDB files, HSDBTPS struct info and transformation tables.

annotations\tps\H...13450_selected.csv

annotations\tps\H...3120_selected.csv

annotations\tps\H...31645_selected.csv

...\S68_HSDBTPS_StructureInfoOnly.csv

annotations\tps\H...formationTable.csv

S-demethylation; N-deethylation; and disulfide formation." 13450|13450,13450|135495928|135612794,TRUE,mammalian metabolism,TRUE,"TPs added, rest are not yet in PubChem or too inspecific"

+HSDB,1525,TERBUTRYNE,13450,1,13450,"Menzie, C.M. Metabolism of Pesticides-Update III. Special Scientific Report- Wildlife No. 232. Washington, DC: U.S.Department of the Interior, Fish and Wildlife Service, 1980., p. 540", "After administration of terbutryne to rats, urinary metabolites observed ... included: 2-hydroxy terbutryne; 2-amino-4-hydroxy-6-t-butylamino-s-triazine; 2-amino-4-t-butylamino-6-mercapto-s-triazine; two S-glucuronides and two t-butyl-O-glucuronides. Other metabolites were formed by one or a combination of the follow

8.5 Transformations

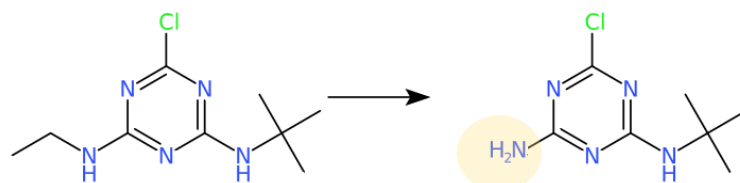
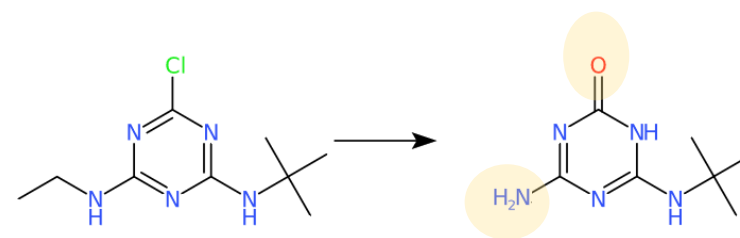
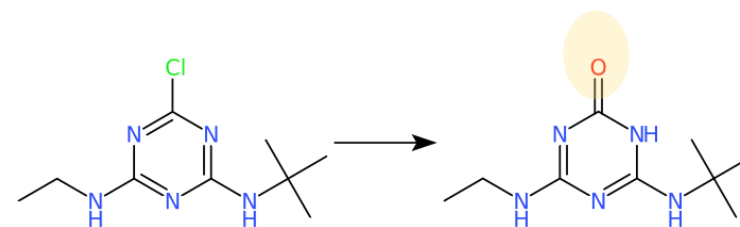
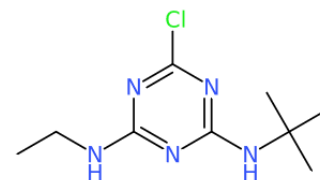
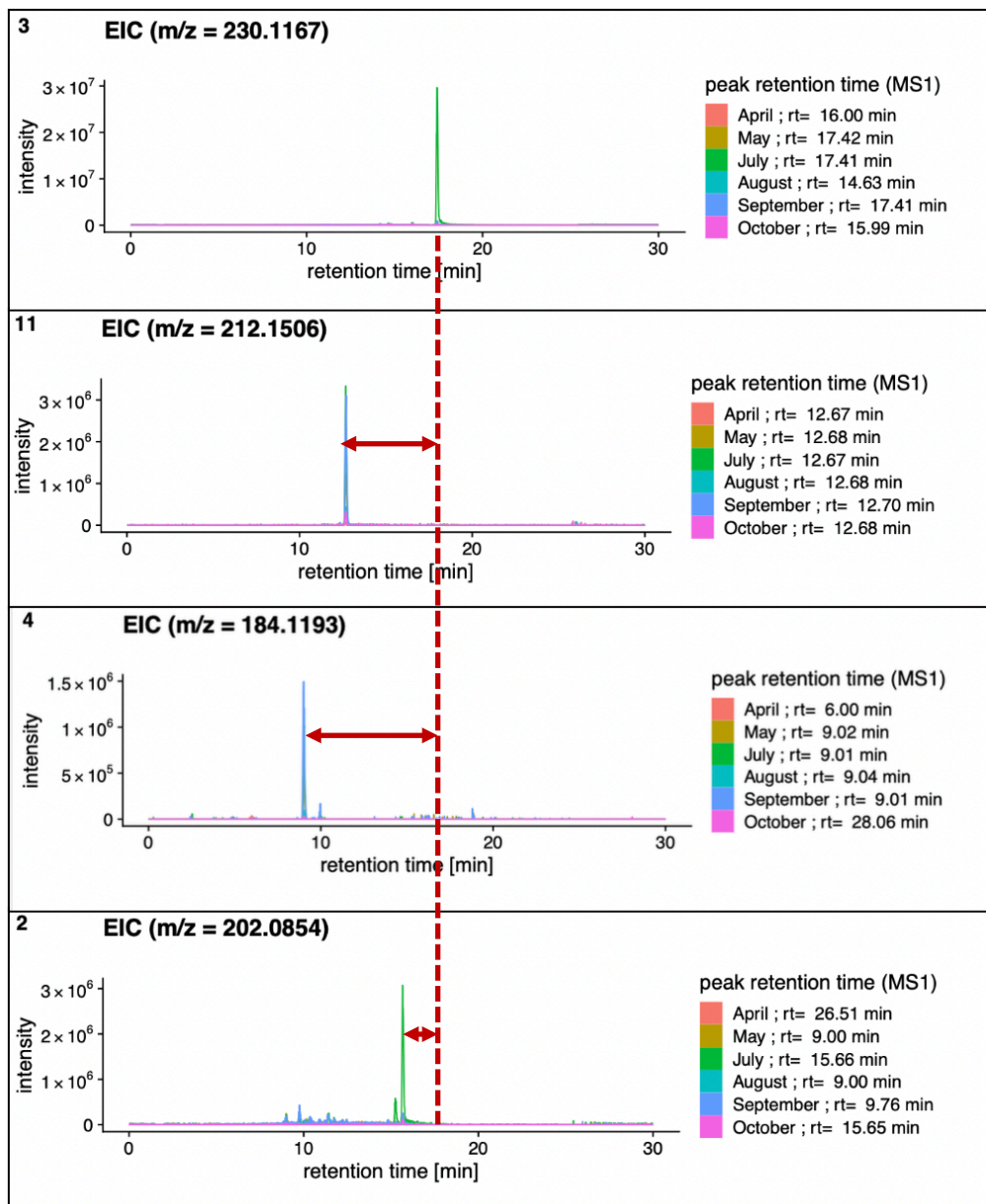
19 items View More Rows & Details

Download

SORT BY Please Choose One

Predecessor Image	Predecessor Name	Transformation	Successor Image	Successor Name	Evidence DOI
	Terbutryn	Mammalian metabolism		2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	10.1002/bms.12000506
	Terbutryn	Mammalian metabolism		2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	10.5281/zenodo.38274

LuxPest – Terbutylazine and (tentative) TPs



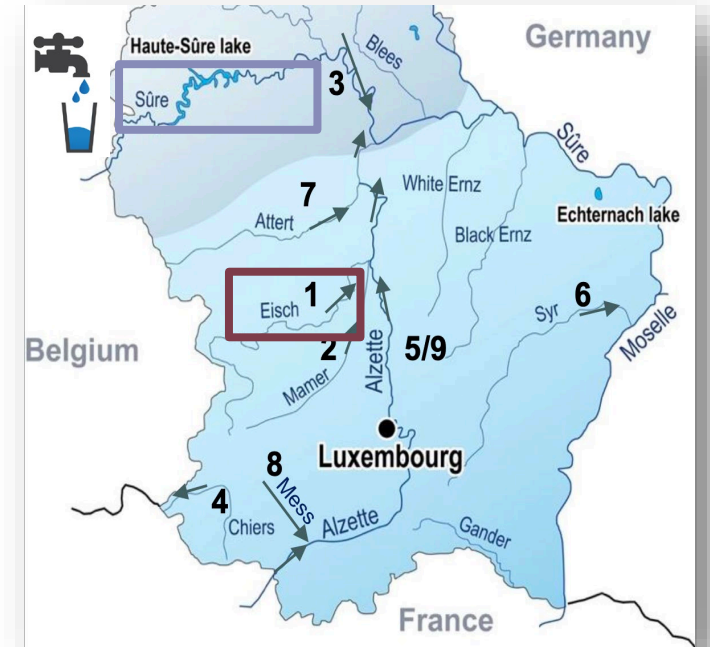
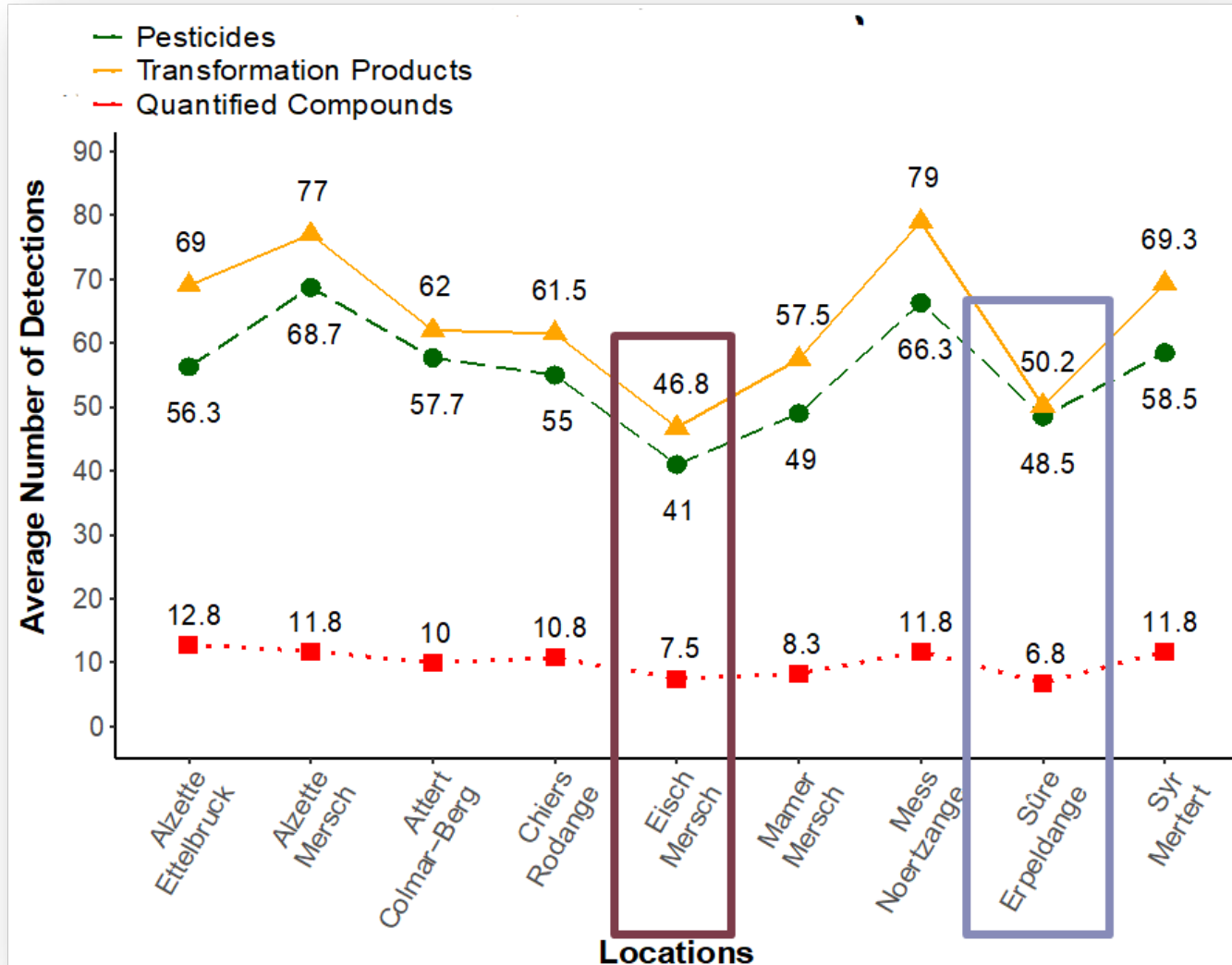
LuxPest – Verification and Quantification



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



LuxPest – Spatial Distribution



LuxPest – preprint out; manuscript in review!



Research Square

Browse Tools & Services

This is a preprint, a preliminary version of a manuscript that has not completed peer review at a journal. Research Square does not conduct peer review prior to posting preprints. The posting of a preprint on this server should not be interpreted as an endorsement of its validity or suitability for dissemination as established information or for guiding clinical practice.

RESEARCH ARTICLE

Discovering Pesticides and their Transformation Products in Luxembourg Waters using Open Cheminformatics Approaches

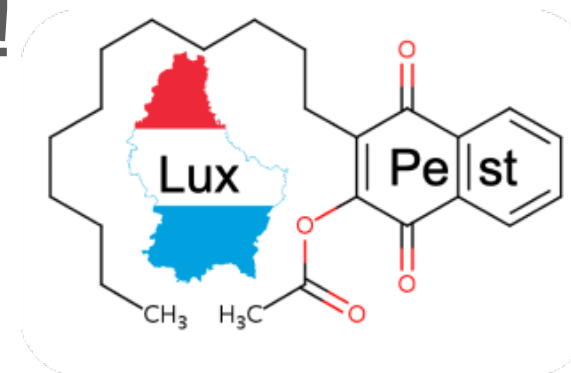
> Jessy Krier, Randolph R. Singh, Todor Kondic, Adelene Lai, Philippe Diderich, Jian Zhang, Paul A. Thiessen, Evan E. Bolton, Emma L. Schymanski

DOI: 10.21203/rs.3.rs-478324/v1 [Download PDF](#)

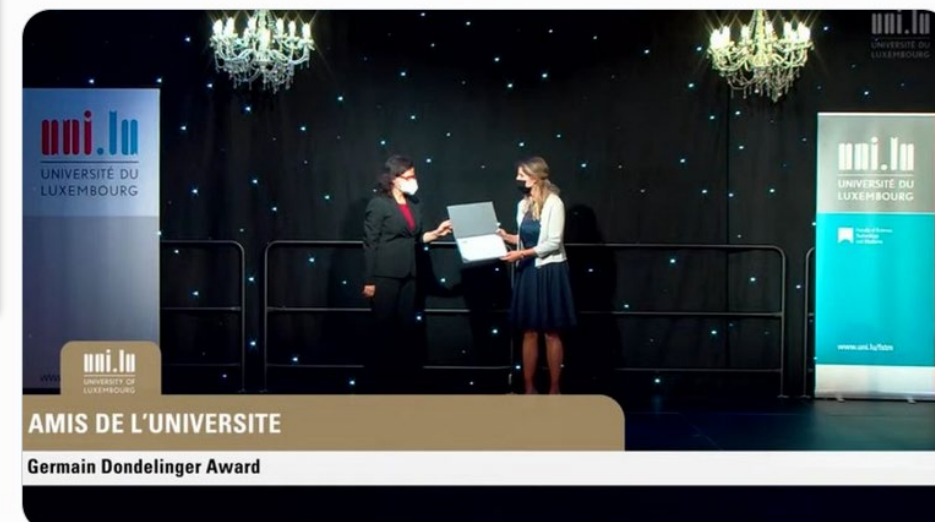
LICENSE: This work is licensed under a CC BY 4.0 License. [Read Full License](#)

DECLARATIONS: [View author declarations.](#)

<https://www.researchsquare.com/article/rs-478324/v1>



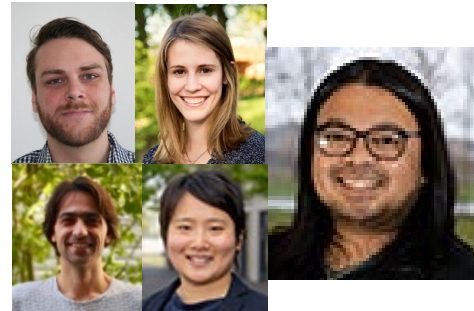
Emma Schymanski @ESchymanski · May 11
Congratulations to the #ECI LCSB @uni_lu #ClassOf2020 including @krije_ @AnjuAnjuraj15 @HibaMohamedTaha and @NarayananMira - special congrats to @krije_ for the Germain Dondelinger Award for her masters thesis! Our first graduates! #ProudPI @FnrLux



LuxPharma – from CNS



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



zenodo

Search [] Upload Communities Log in Sign up

March 6, 2021 Dataset Open Access

S76 | LUXPHARMA | Pharmaceuticals Marketed in Luxembourg

ib Singh, Randolph R
Other(s)
ib Schymanski, Emma

This is the collection associated with list S76 LUXPHARMA Pharmaceuticals Marketed in Luxembourg on the NORMAN Suspect List Exchange.

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

This list contains pharmaceuticals marketed in Luxembourg, as published by d'Gesondheetskeess (CNS, la caisse nationale de santé, www.cns.lu), mapped by name to structures using CompTox by R. Singh et al. (in prep.). List downloaded from <https://cns.public.lu/en/legislations/textes-coordonnes/liste-med-comm.html>. Dataset DOI: 10.5281/zenodo.4587355

Preview

CNS_Numero_National	INPUT	FOUND_BY	DTXSID	PREFERRED_NAME
J05AF06	abacavir	Approved Name	DTXSID4046444	Abacavir
L02BX03	abiraterone	Approved Name	DTXSID80879993	Abiraterone
N07BB03	acamprosate	Approved Name	DTXSID3044259	Acamprosate

LE GOUVERNEMENT DU GRAND-DUCHÉ DE LUXEMBOURG

luxembourg.lu guichet.lu gouvernement.lu crossgov.lu Other sites

www.covid19.lu Find official information on sanitary measures and recommendations, information for travellers, protective measures, sectoral information and thematic FAQ.

FR EN DE

CNS d'Gesondheetskeess

Legislations

Liste des médicaments commercialisés - Triée par code ATC

Publication date: March 6, 2021

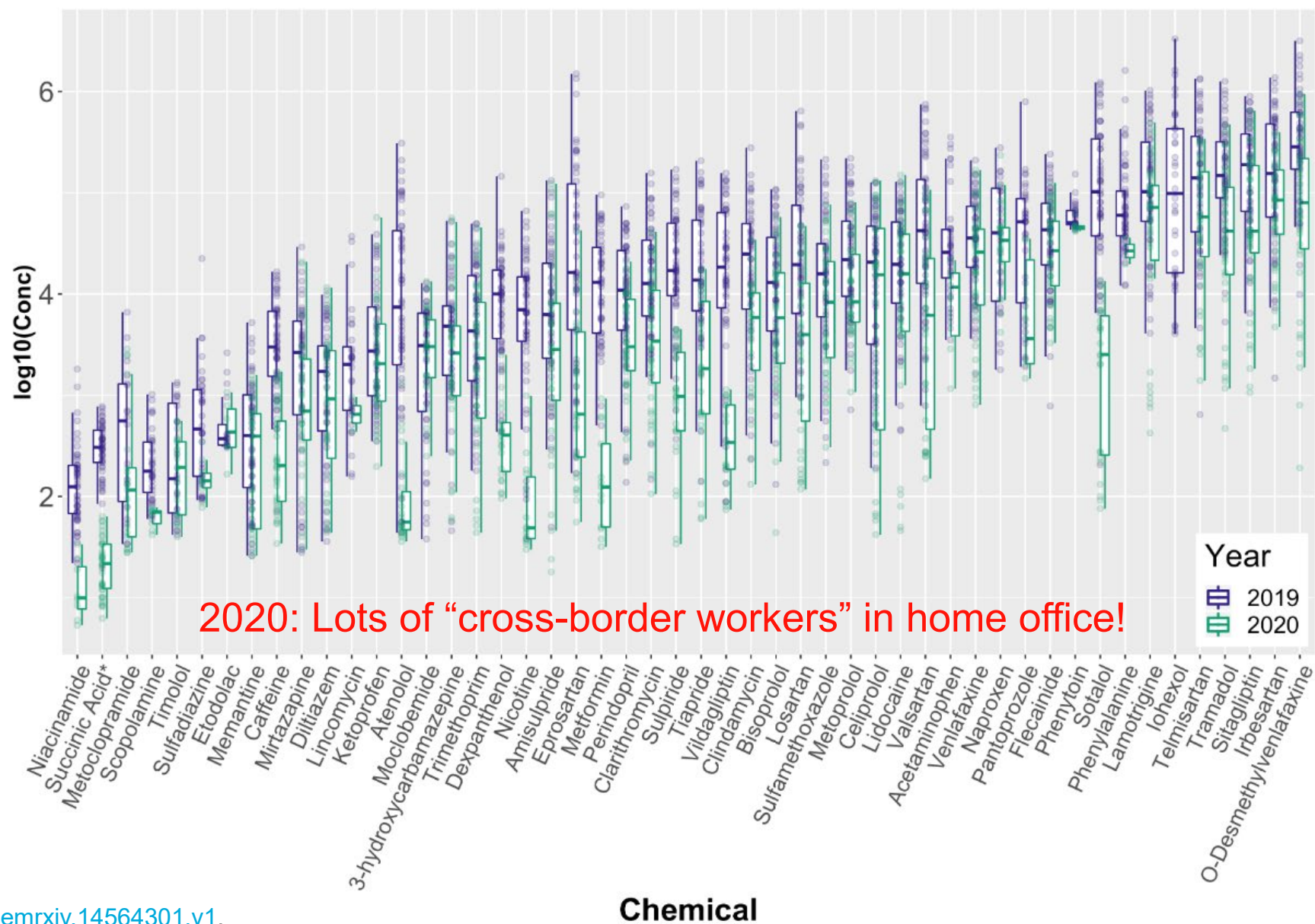
DOI: DOI 10.5281/zenodo.4587356

Keyword(s): Pharmaceuticals

Related identifiers: Part of <https://www.norman-network.com/nds/SLE/>

Communities: LCSB Environmental Cheminformatics Group NORMAN Suspect List Exchange

LuxPharma – 2019 vs 2020 – differences due to COVID?



LuxPharma – preprint also out; manuscript in review...



ChemRxivTMHOW TO SUBMITBROWSEABOUT

Search ChemRxiv

LOG IN

Analytical Chemistry

Occurrence and Distribution of Pharmaceuticals and their Transformation Products in Luxembourgish Surface Waters

WORKING PAPER

Randolph Singh IFREMER (Institut Français de Recherche pour l'Exploitation de la Mer),
Adelene Lai, **Jessy Krier**, **Todor Kondić**, **Philippe Diderich**,
Emma Schymanski

Abstract

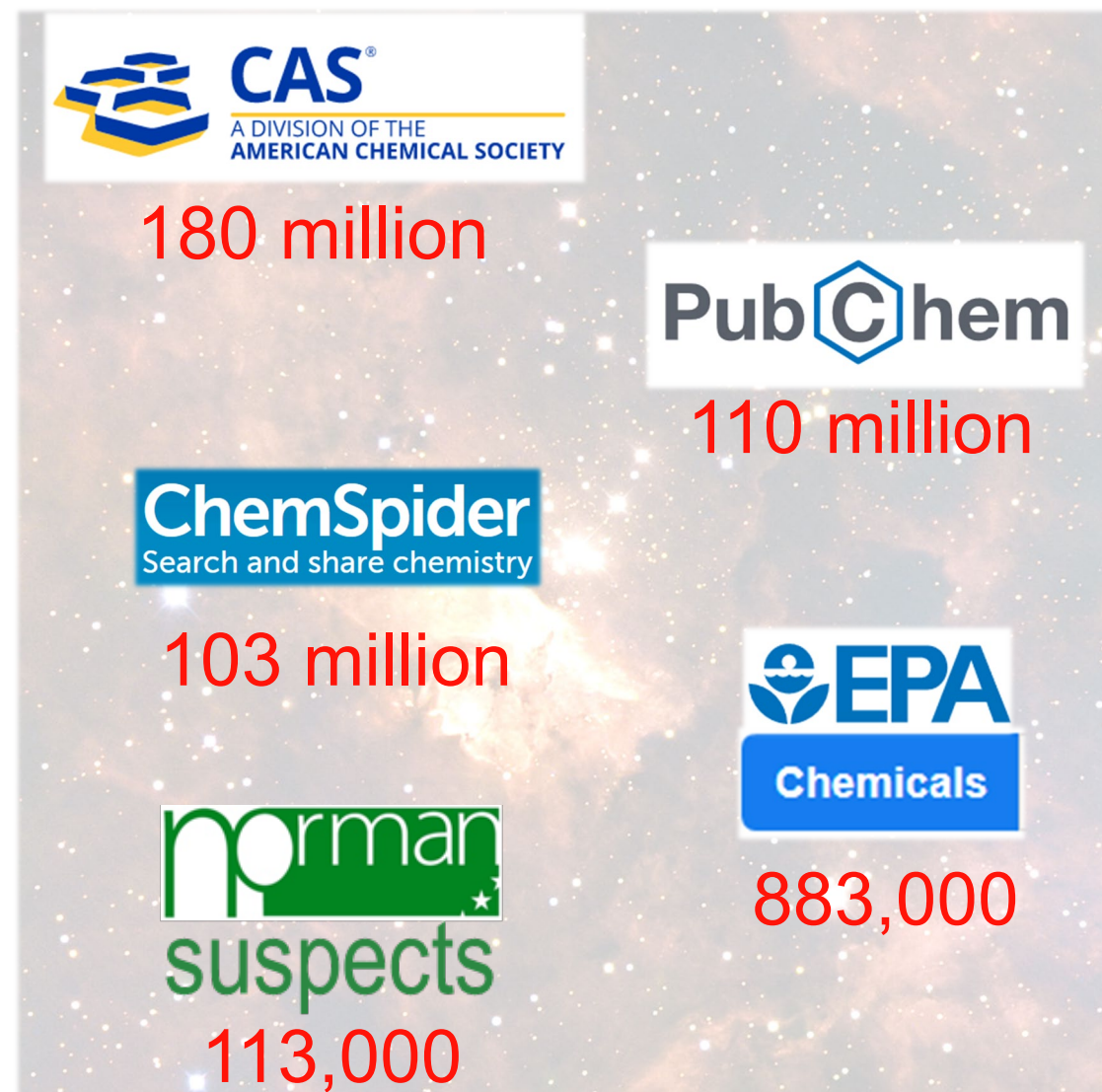
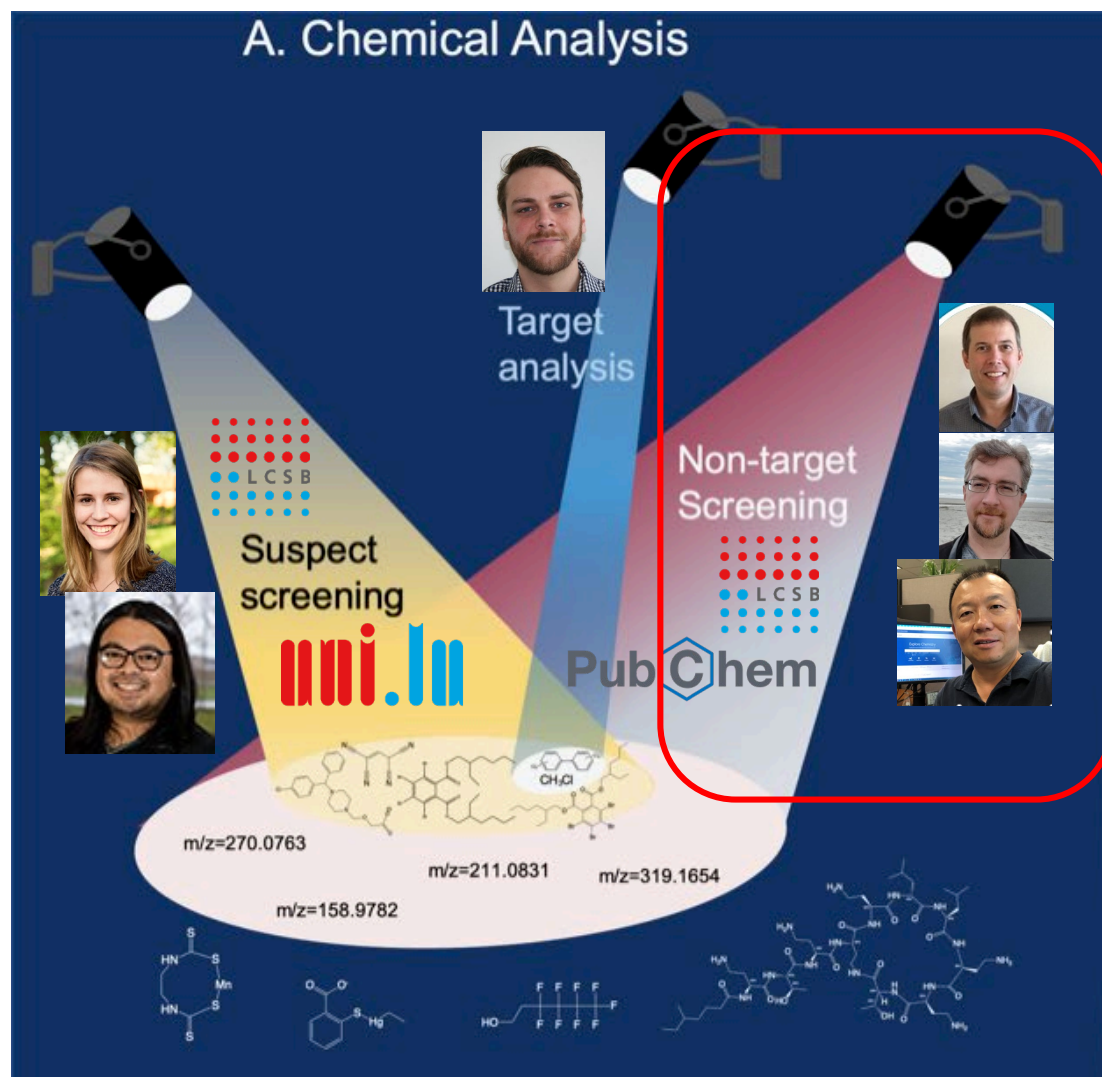
DOWNLOAD

Version History
May 11, 2021 Version 1

Metrics

298	Views
107	Content Downloads

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



Problem: Exposomics “Chemical Space” is too big!



180 million



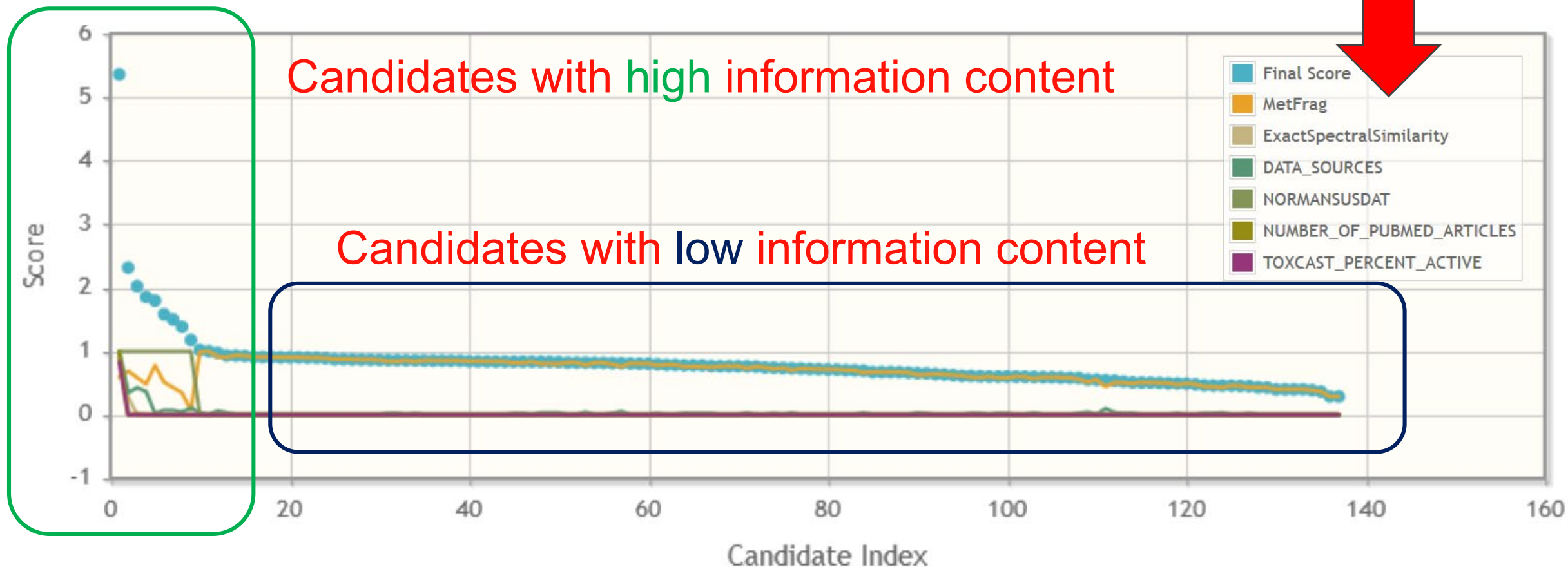
110 million



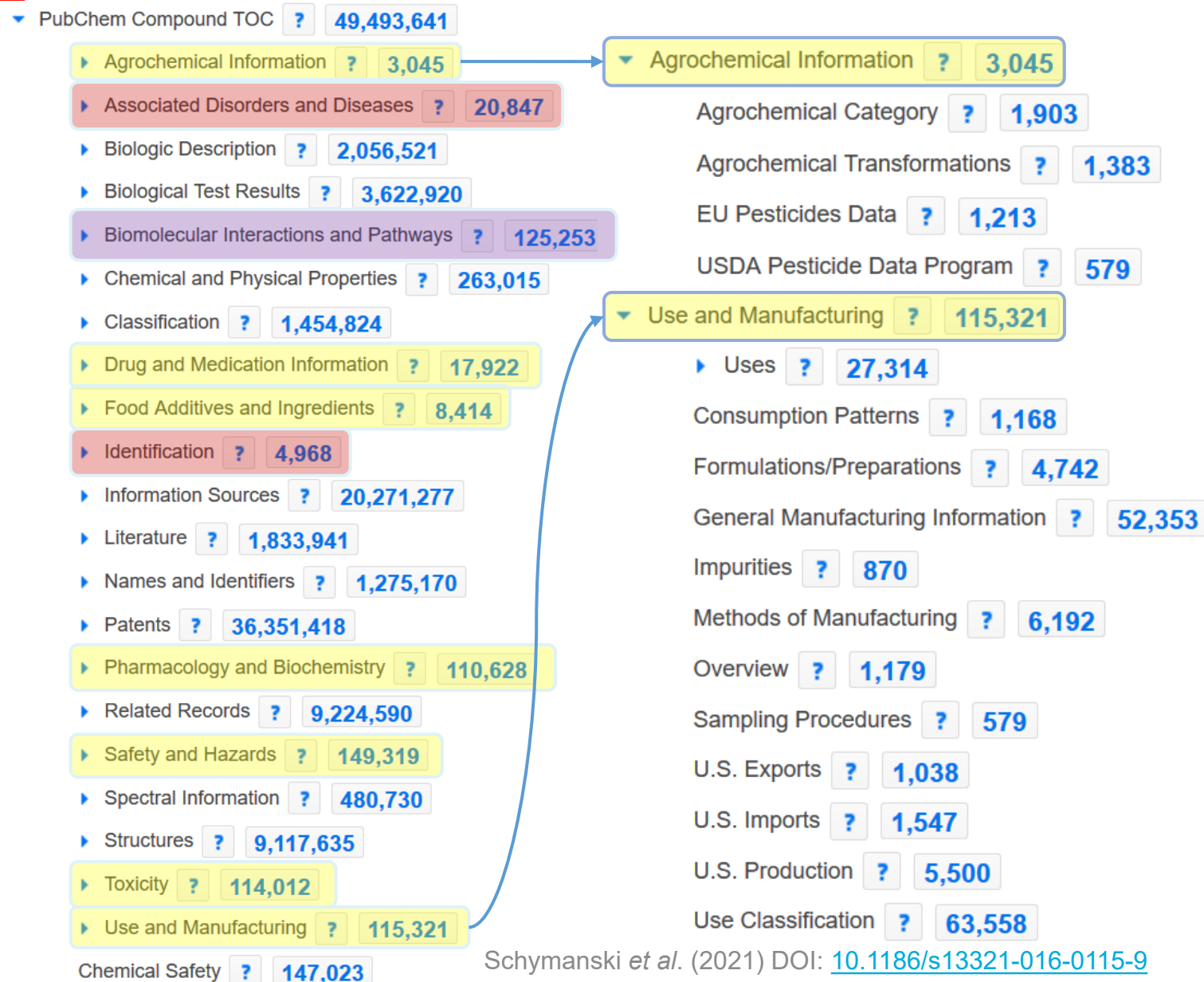
103 million



883,000



Can we break down PubChem into useful bits?



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

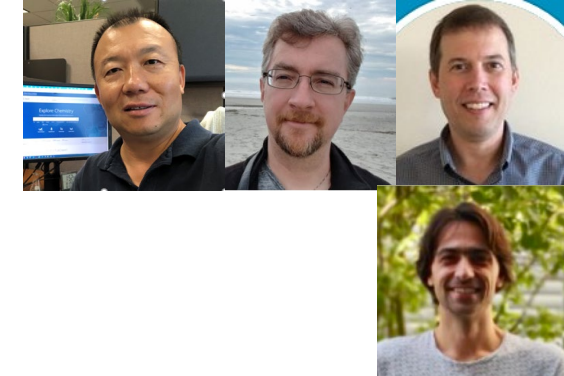
7.3 EU Pesticides Data



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

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

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

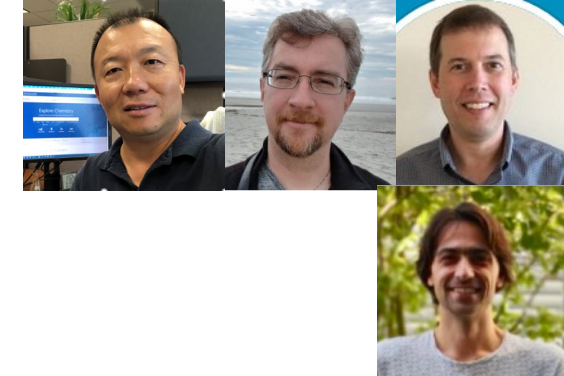
PubChemLite
EXPOSOMICS

~370,000 entries “small”

Schymanski *et al.* (2021) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)

Introducing ...

PubChemLite EXPOSOMICS



Collapsed by InChIKey First Block (skeleton)
and by presence of annotation content

~370,000 entries “small”

PubChem

Compounds (6) Substances (2)

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text searched. [Read More...](#)

6 results SORT BY



2,3-dihydroxy-2,3-dihydrobenzoic Acid; 5,6-dihydroxycyclohexa-1,3-diene-1-carboxylic acid; 100459-00-5; 2,3-dihydro-2,3-dihydroxybenzoic Acid; ACM

Compound CID (sort by): 3

MF: C₇H₈O₄ MW: 156.14g/mol

InChIKey: [INCSWKICYAHB-UHFFFAOYSA-N](#)

IUPAC Name: 5,6-dihydroxycyclohexa-1,3-diene-1-carboxylic acid

Create Date: 2004-09-16

[Summary](#) [Similar Structures Search](#) [Related Records](#)



(2S,3S)-2,3-dihydroxy-2,3-dihydrobenzoic Acid; 176487-06-2; (2S,3S)-2,3-dihydroxybenzoate; (5S,6S)-5,6-dihydroxycyclohexa-1,3-diene-1-carboxylic acid; (2S,3S)-2,3-Dihydro-2,3-dihydroxybenzoate; ...

Compound CID (sort by): 9964159

MF: C₇H₈O₄ MW: 156.14g/mol

InChIKey: [INCSWKICYAHB-WDSKDSINSA-N](#)

IUPAC Name: (5S,6S)-5,6-dihydroxycyclohexa-1,3-diene-1-carboxylic acid

Create Date: 2006-10-25

zenodo [Communities](#)

October 31, 2020

[Dataset](#) [Open Access](#)

[Edit](#)

[New version](#)

PubChemLite for Exposomics

[Bolton, Evan](#); [Schymanski, Emma](#); [Kondic, Todor](#); [Thiessen, Paul](#); [Zhang, Jeff](#)

PubChemLite is a subset of PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) selected from major contents page at the PubChem Classification Browser (<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72>). With this release, there is now just one is the former tier1 plus two new categories (Associated Disorders &

is 371,663 compounds (31 Oct 2020) compiled from 10 categories: DrugMedicInfo, FoodRelated, Pharmacolnfo, Info, Identification.

Collapsed by InChIKey first block, reporting the most CIDs. Entries that will be ignored by MetFrag (e.g. transition metals) have been removed on the PubChem FTP site. The "AnnoTy how represented, the subsequent column (name per category) counts the ories available in the next sub-category of the TOC entry.



Database Settings

Database:

Neutral Mass: Search ppm:

Formula:

Identifiers:

[Retrieve Candidates](#)

[4 Candidates](#)

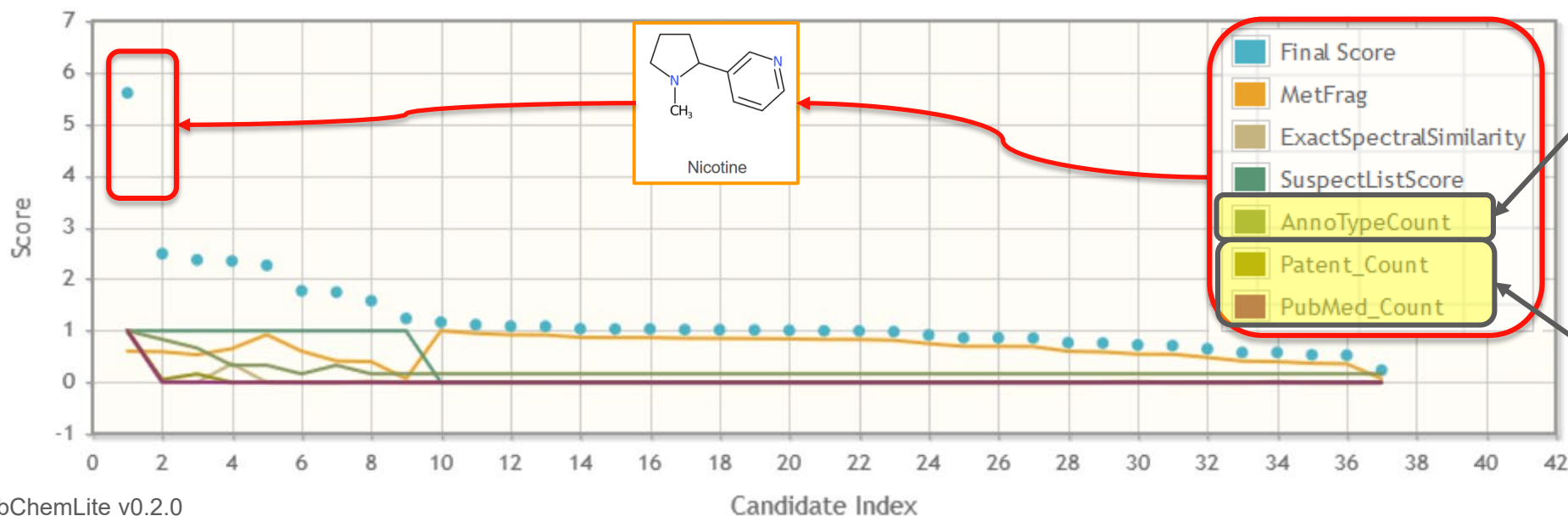
Schymanski *et al.* (2021) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)

MetFragRL + PubChemLite: tailor-made database + metadata



Statistics

Candidate Score Distribution



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
Blomolecular 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 v0.2.0

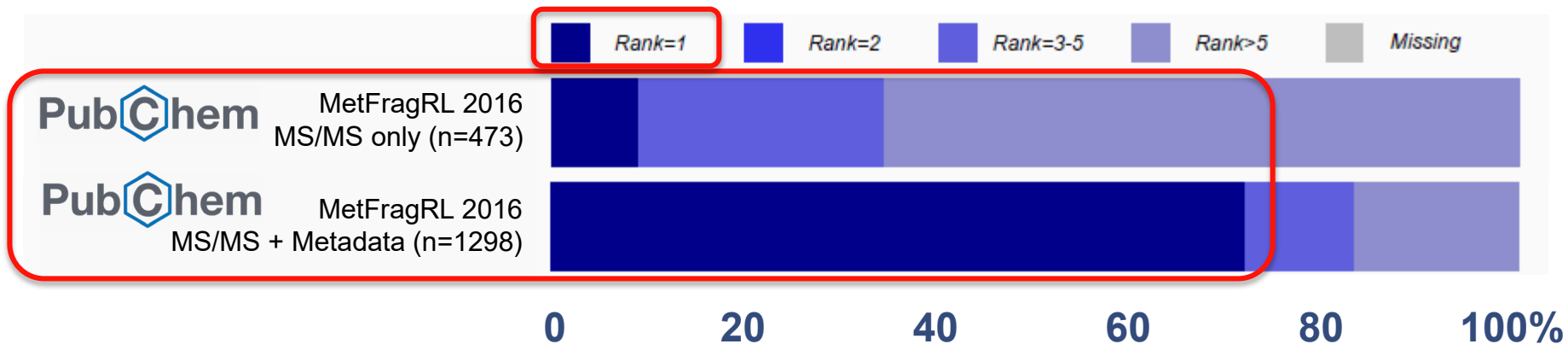
Candidate Index

MetFrag+PubChemLite+Formula+MoNA+SusDat+Pat+Refs+Anno + <https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag>



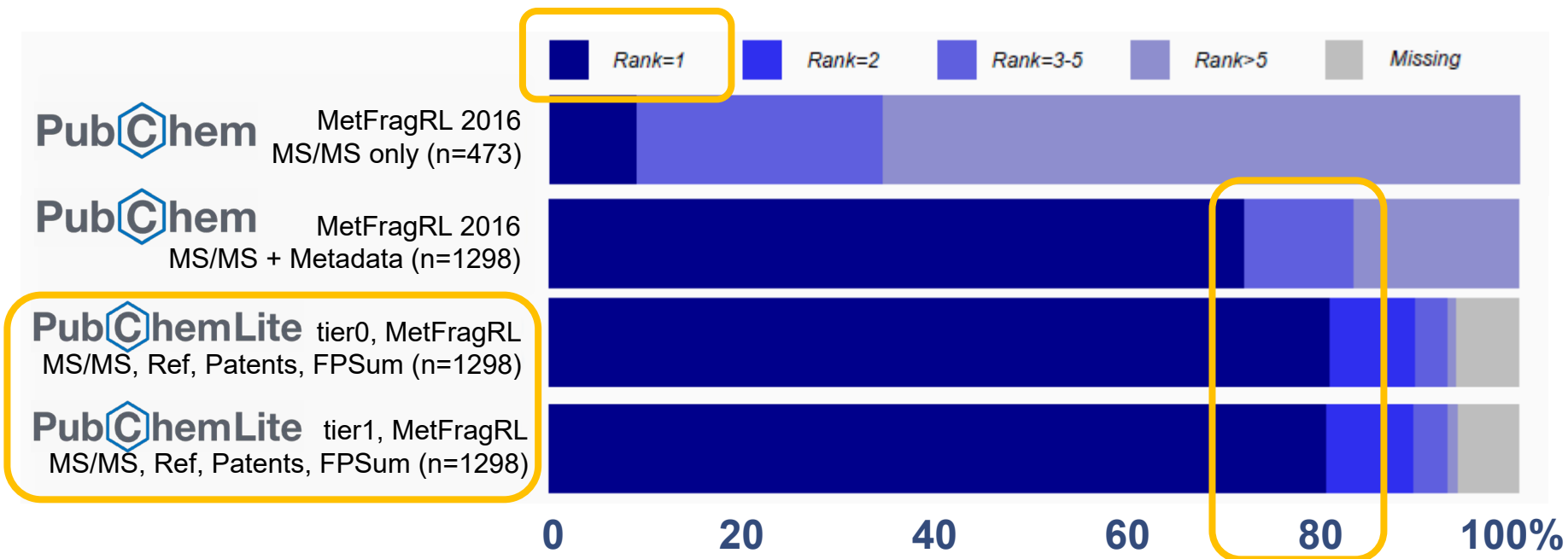
How does PubChemLite perform?

- Earlier PubChem / MetFrag evaluations:



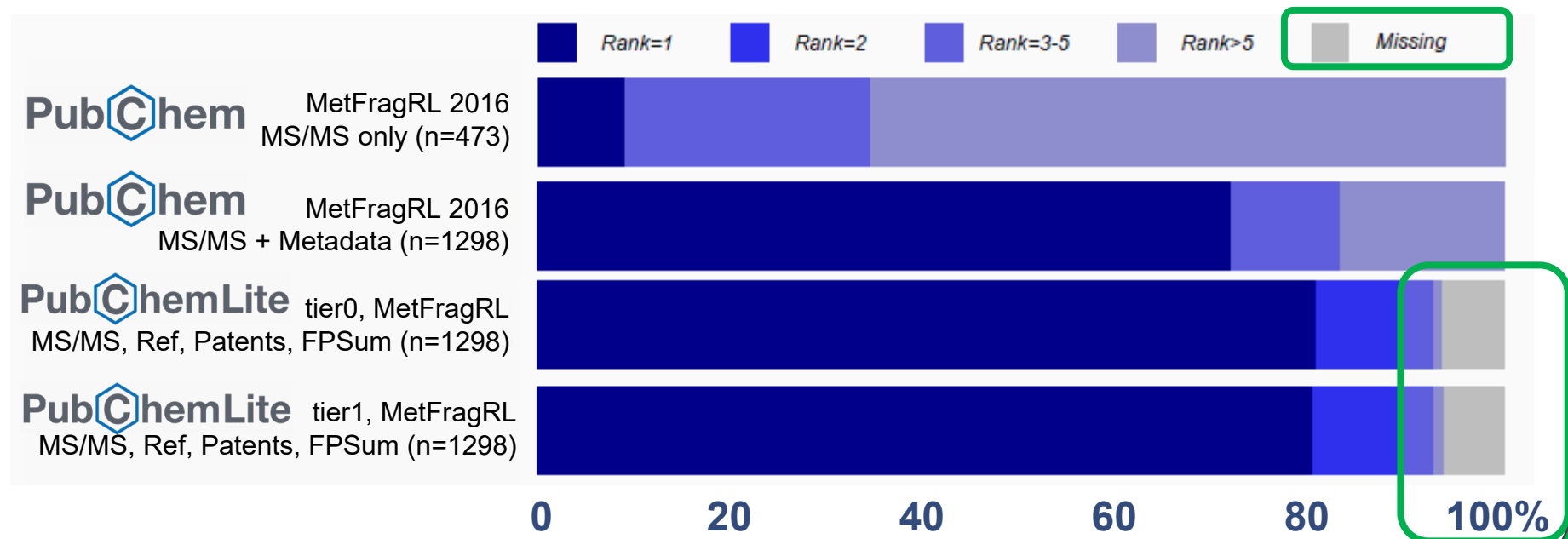
How does PubChemLite perform?

- ~110 M => ~300 K ... how does this influence performance?



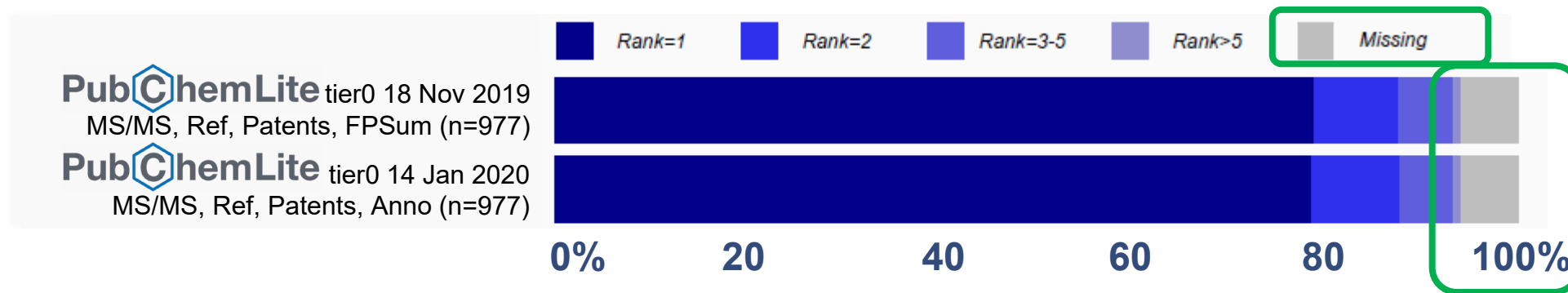
How does PubChemLite perform?

- ~110 M => ~300 K ... how does this influence performance?



norman
suspects

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

norman
suspects

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

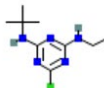
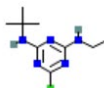
PubChem Terbutylazine (Compound)

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
	Terbutylazine	Mammalian metabolism		6-Chloro-1,3,5-triazine-2,4-diamine	10.5281/zenodo.382
	Terbutylazine	Deethylation		Terbutylazine-desethyl	10.1007/s13361-017-

Transformation Products: Filling the Data Gaps!

PubChem Terbutylazine (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](https://doi.org/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](https://doi.org/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.

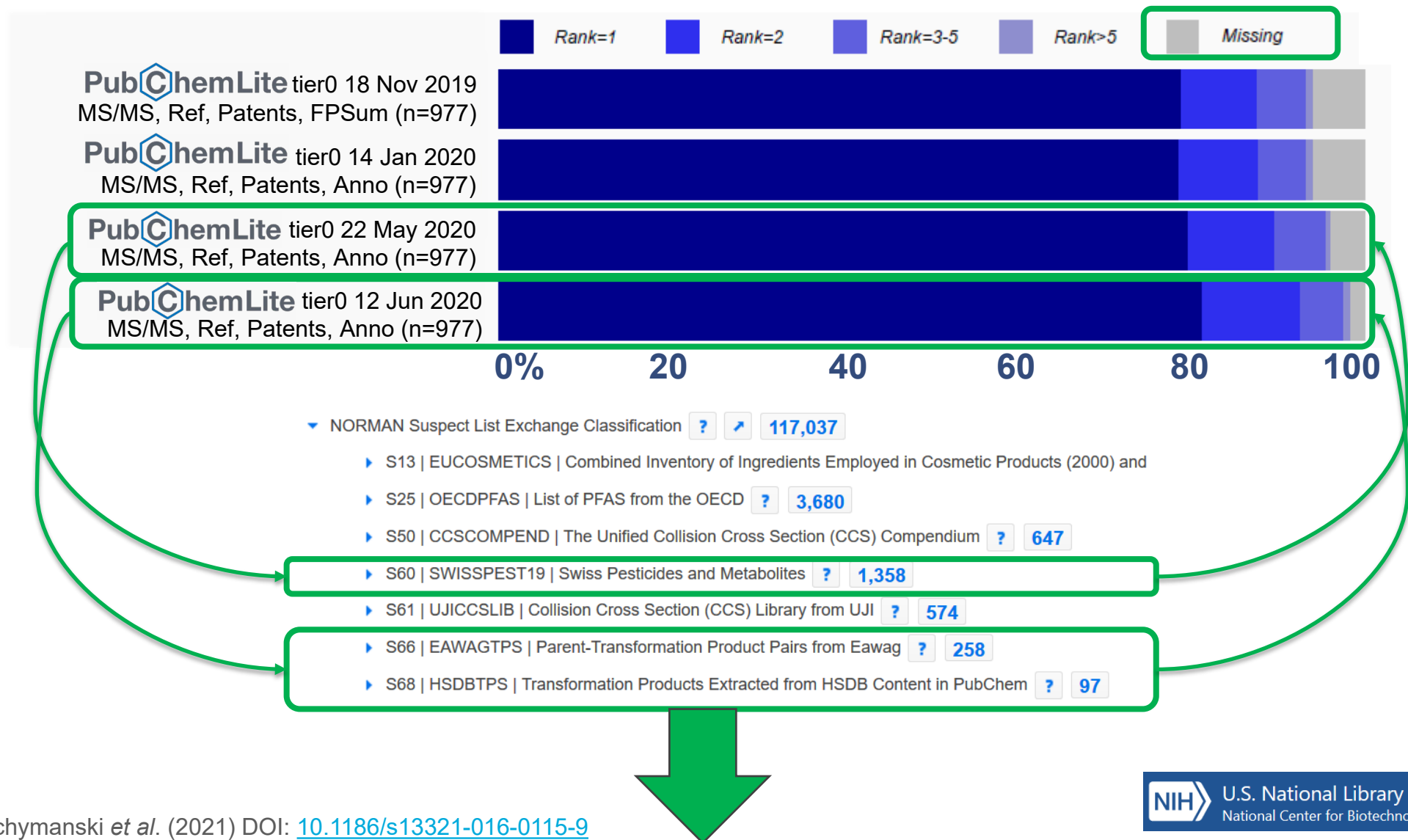
S60 | SWISSEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 | [DOI:10.5281/zenodo.3544759](https://doi.org/10.5281/zenodo.3544759)

► [NORMAN Suspect List Exchange](#)

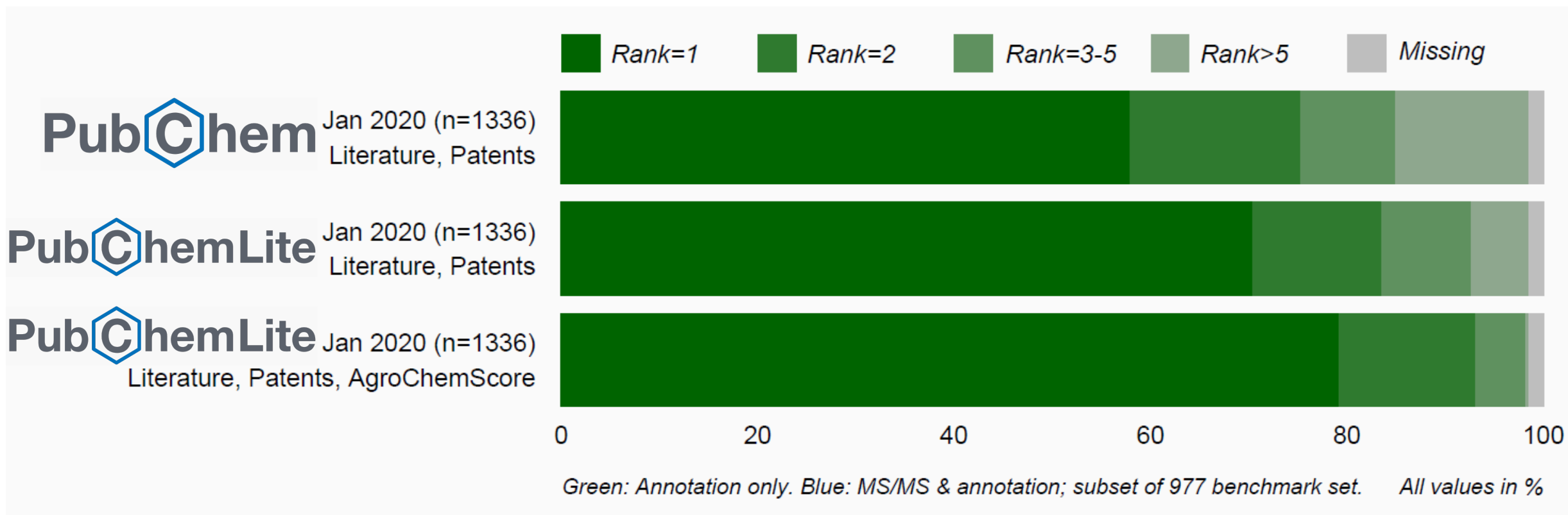
▼ Agrochemical Information	?	3,045
Agrochemical Category	?	1,903
Agrochemical Transformations	?	1,383
EU Pesticides Data	?	1,213
USDA Pesticide Data Program	?	579

PubChem **Lite**
EXPOSOMICS

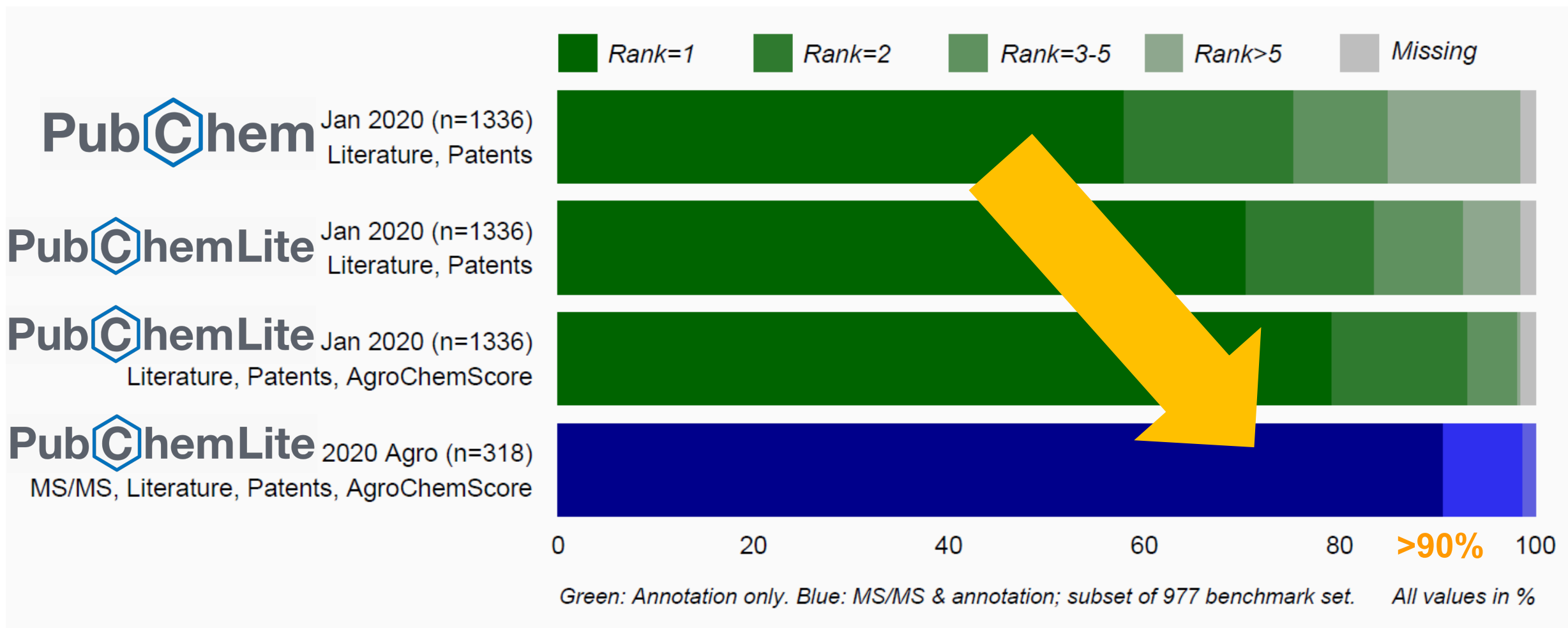
Assessing the Missing Entries in PubChemLite



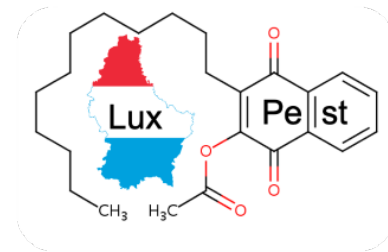
Influence of the Annotation Content in PubChemLite



Influence of the Annotation Content in PubChemLite



Open Source Workflows for Chemical Discovery



norman
suspects

zenodo



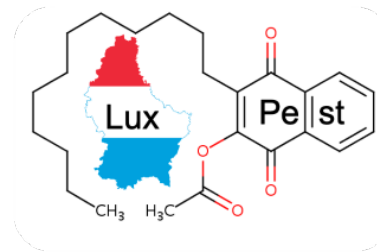
MetFrag

MassBank
High Quality Mass Spectral Database

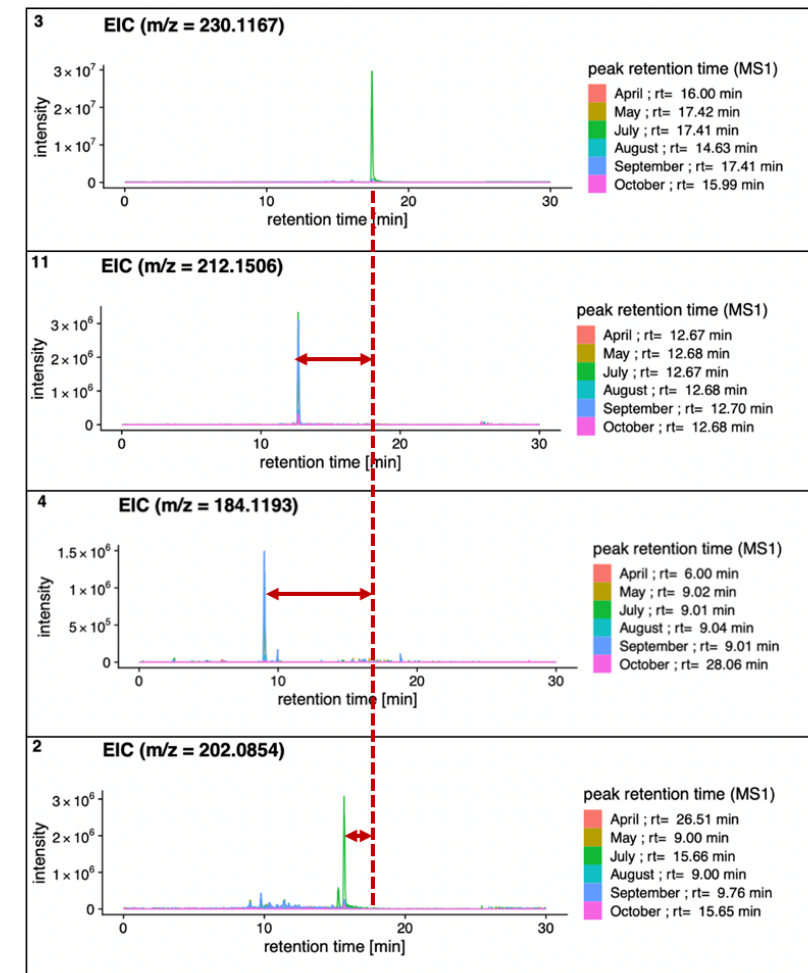


- Open and FAIR Expert Knowledge Exchange
- Open Source pre-screening with auto-QC & manual review
- Comprehensive & open annotation combining MetFrag & MassBank

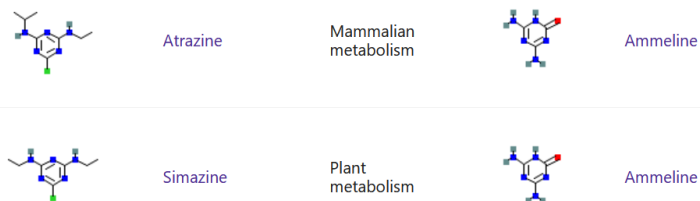
Open Source Workflows for Chemical Discovery



- Extensive transformation product efforts
- Enhanced visualisation and interpretation



PubChem Ammeline (Compound)



PubChem Terbutylazine (Compound)

8.3 Metabolism/Metabolites

Metabolism of terbutylazine 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 Terbutylazine (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 Terbutylazine (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)

Outcomes?



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



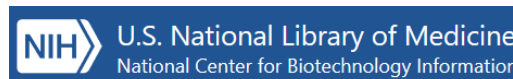
Continued efforts for improved
monitoring of chemicals
(and actions!)
in Luxembourg ...
... and the world!

PubChemLite
EXPOSOMICS

~370,000 entries “small”



Recent Presentation DOI: [10.5281/zenodo.4722507](https://doi.org/10.5281/zenodo.4722507)
Schymanski *et al.* (2021) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)



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

Thank you!



emma.schymanski@uni.lu and [@ESchymanski](https://twitter.com/ESchymanski)

Further Information:

Meet ECI@DOI: [10.5281/zenodo.4596021](https://doi.org/10.5281/zenodo.4596021)

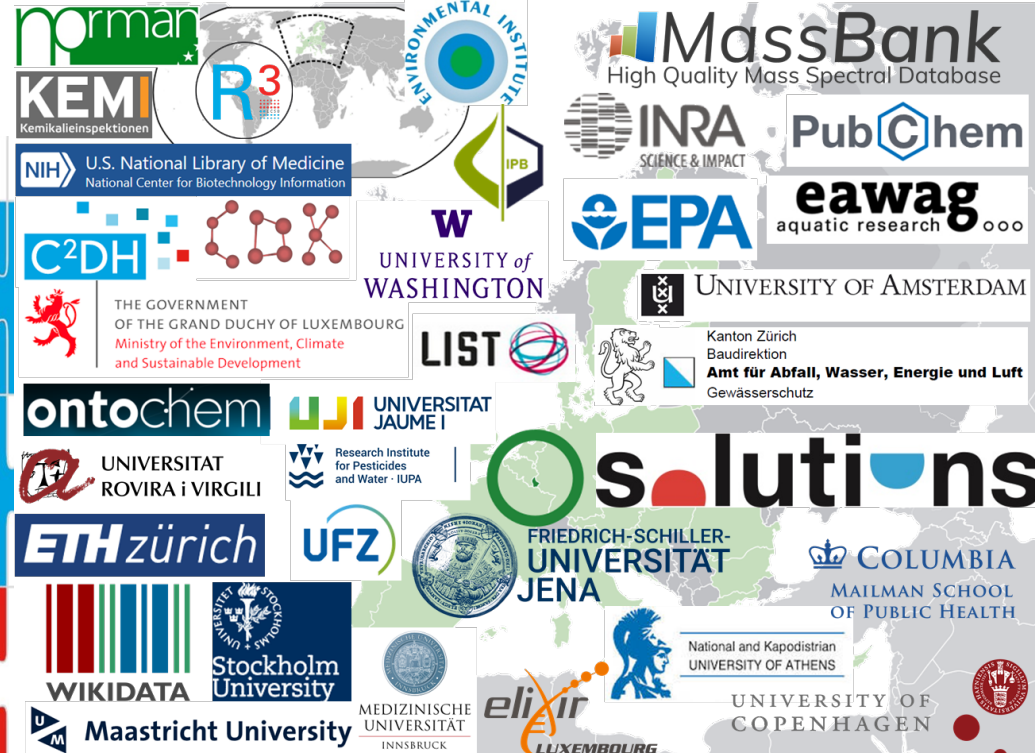
<https://zenodo.org/communities/lcsb-eci/>

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

<https://git-r3lab.uni.lu/eci/>



INSTITUTE FOR ADVANCED STUDIES (IAS)



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

PubChem Met:rag



Slides @ DOI: [10.5281/zenodo.5059829](https://doi.org/10.5281/zenodo.5059829)

