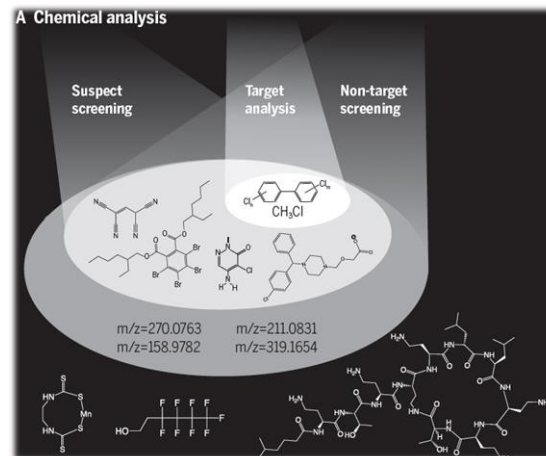
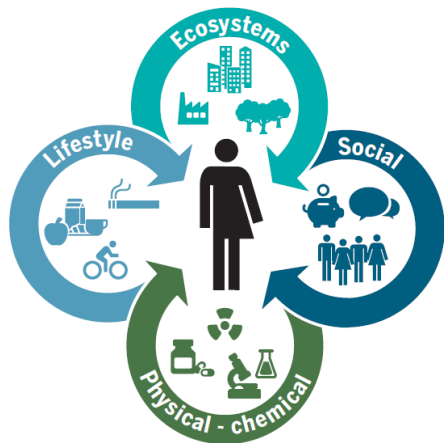


Compound Identification and Exposomics: DIY Databases?



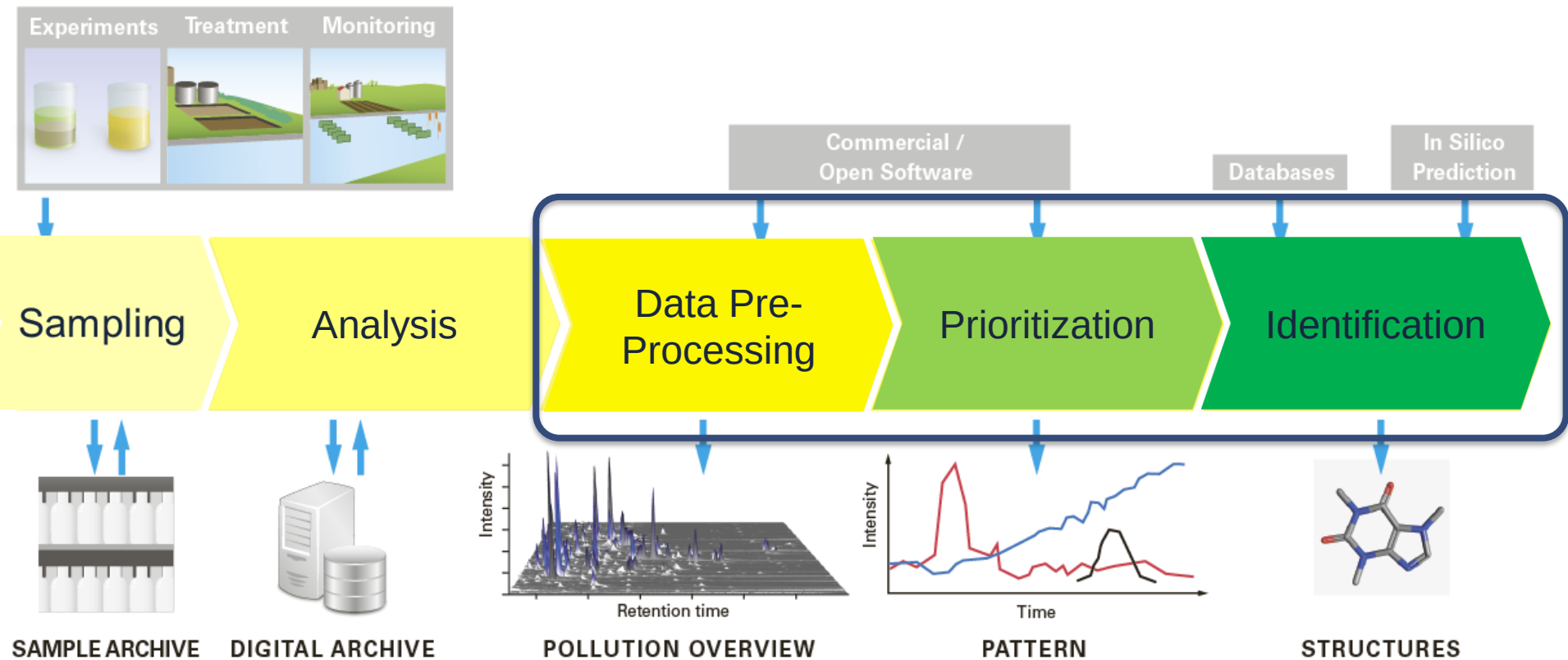
Assoc. Prof. Emma Schymanski

+ the ECI, NORMAN & PubChem teams + more

LCSB, University of Luxembourg

@ESchymanski DOI: [10.5281/zenodo.3742890](https://doi.org/10.5281/zenodo.3742890)

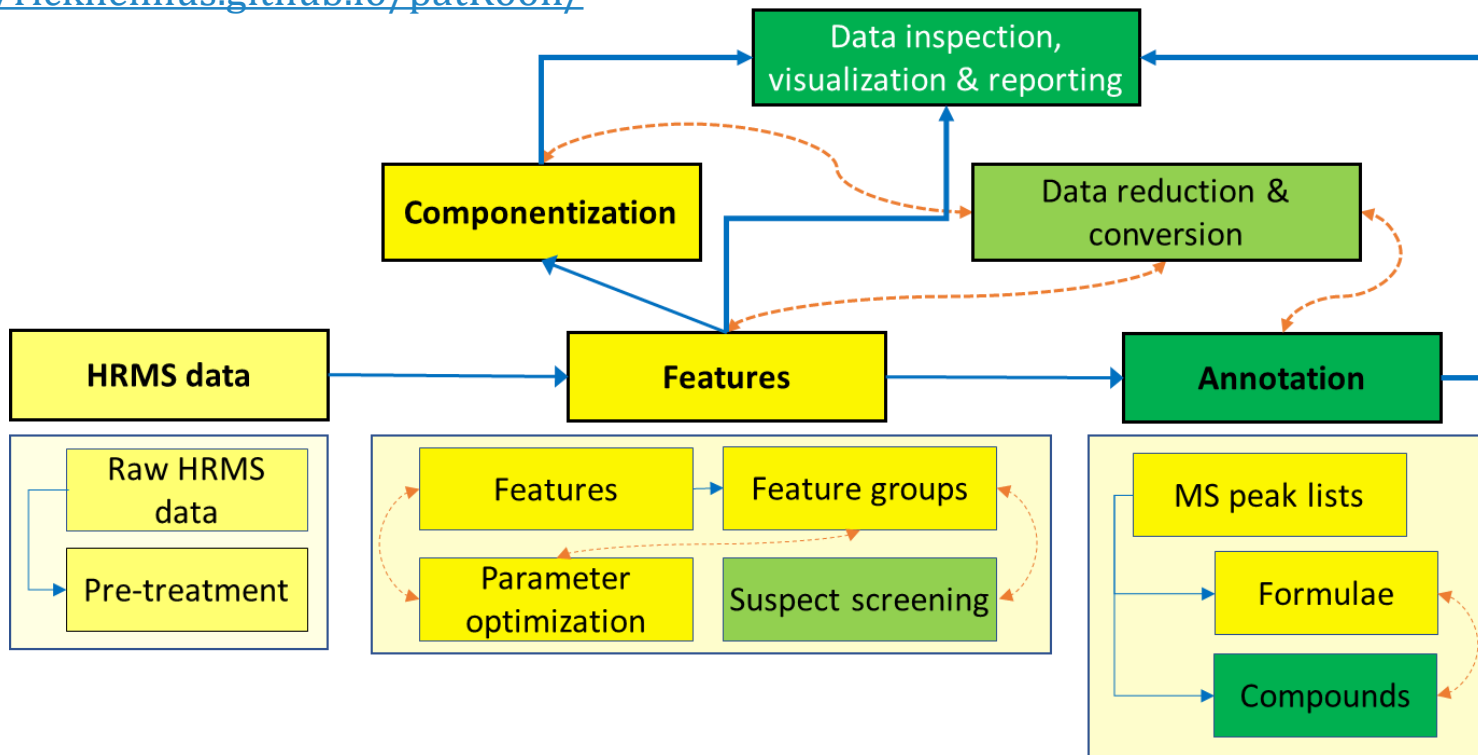
(1) Non-target High Resolution Mass Spectrometry





(2) Open Workflows for NT-HRMS: patRoön

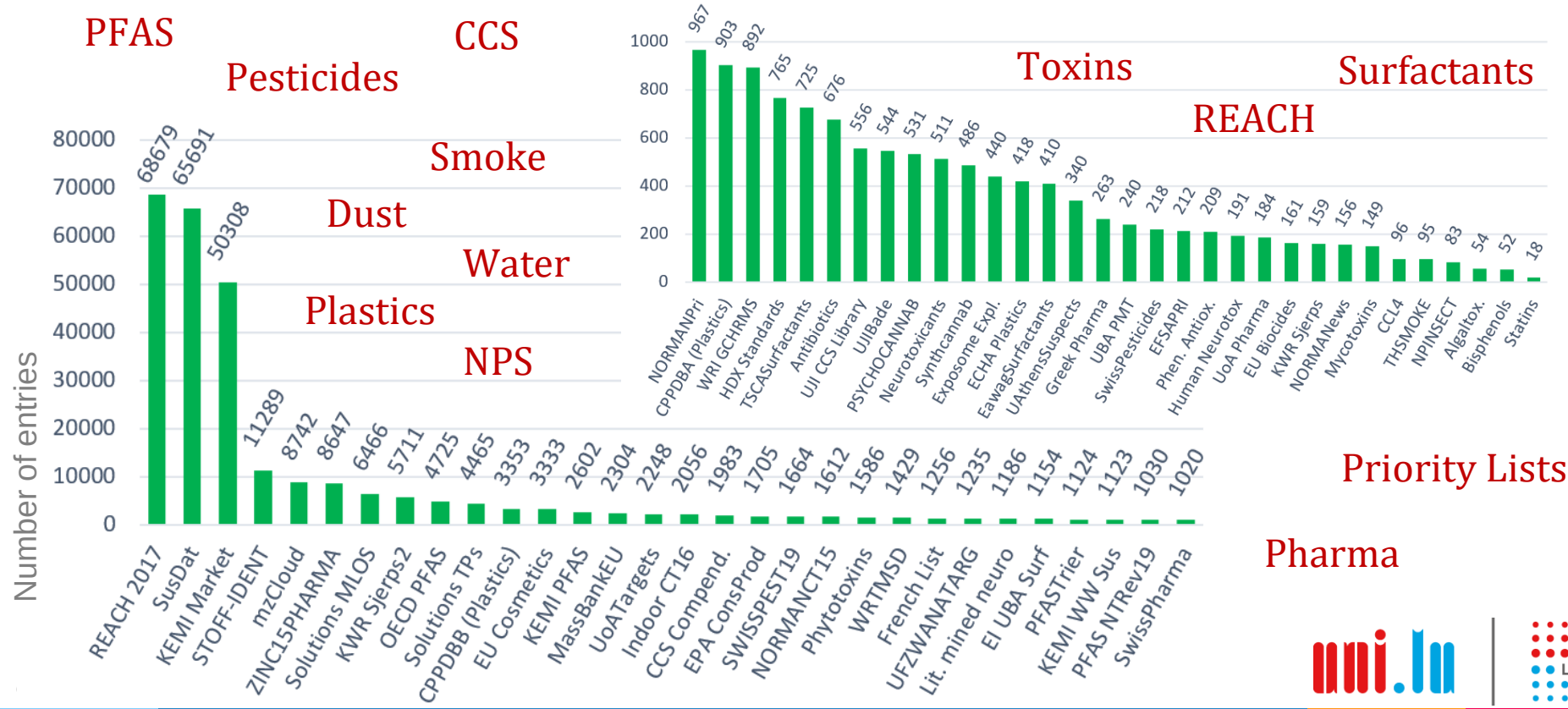
<https://rickhelfmus.github.io/patRoön/>



(3) Expert Info: Suspect List Exchange

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

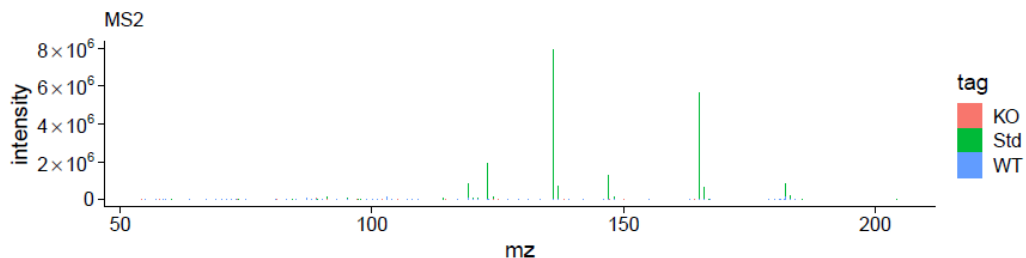
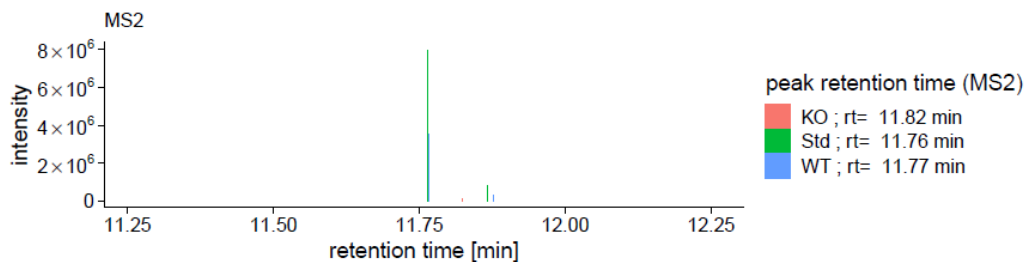
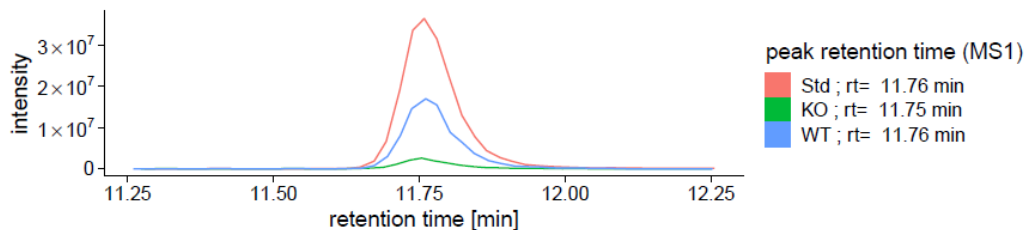
<https://zenodo.org/communities/norman-sle>



(4) ShinyScreen: Auto Screening & Quality Control

100 EIC (m/z = 182.0816)

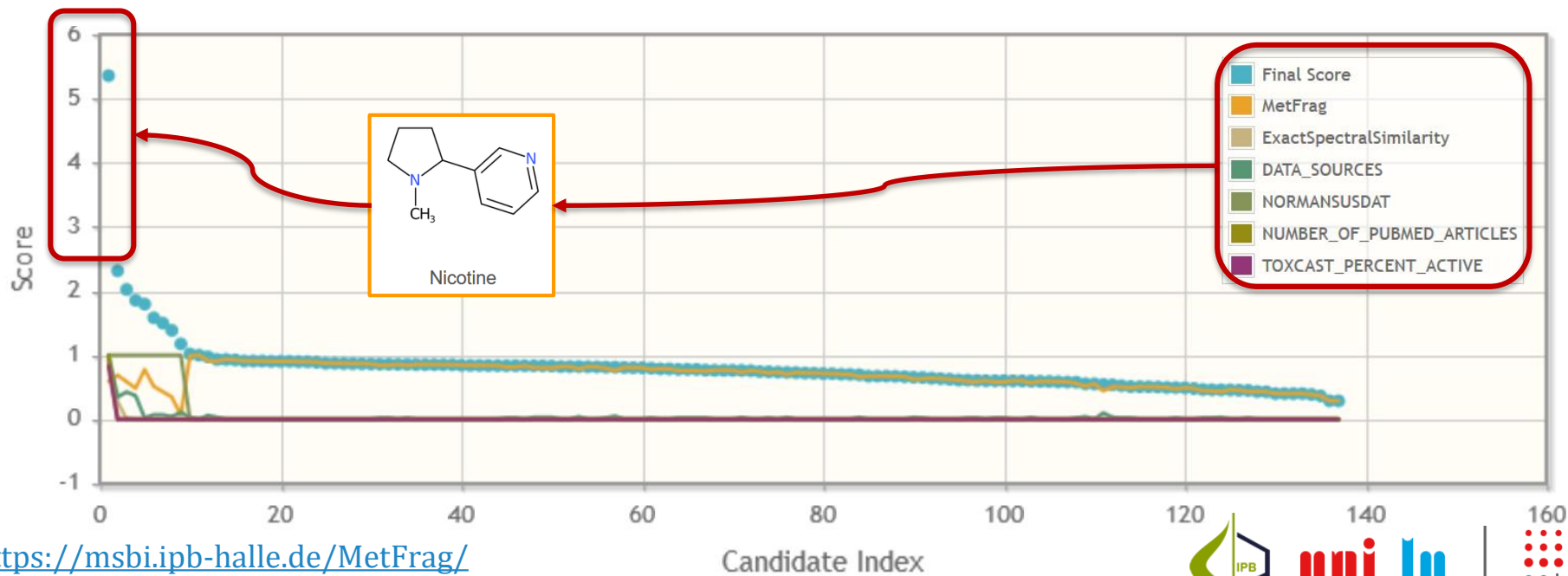
<https://git-r3lab.uni.lu/eci/shinyscreen>



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



(5) MetFrag: Connecting Knowledge for Annotation

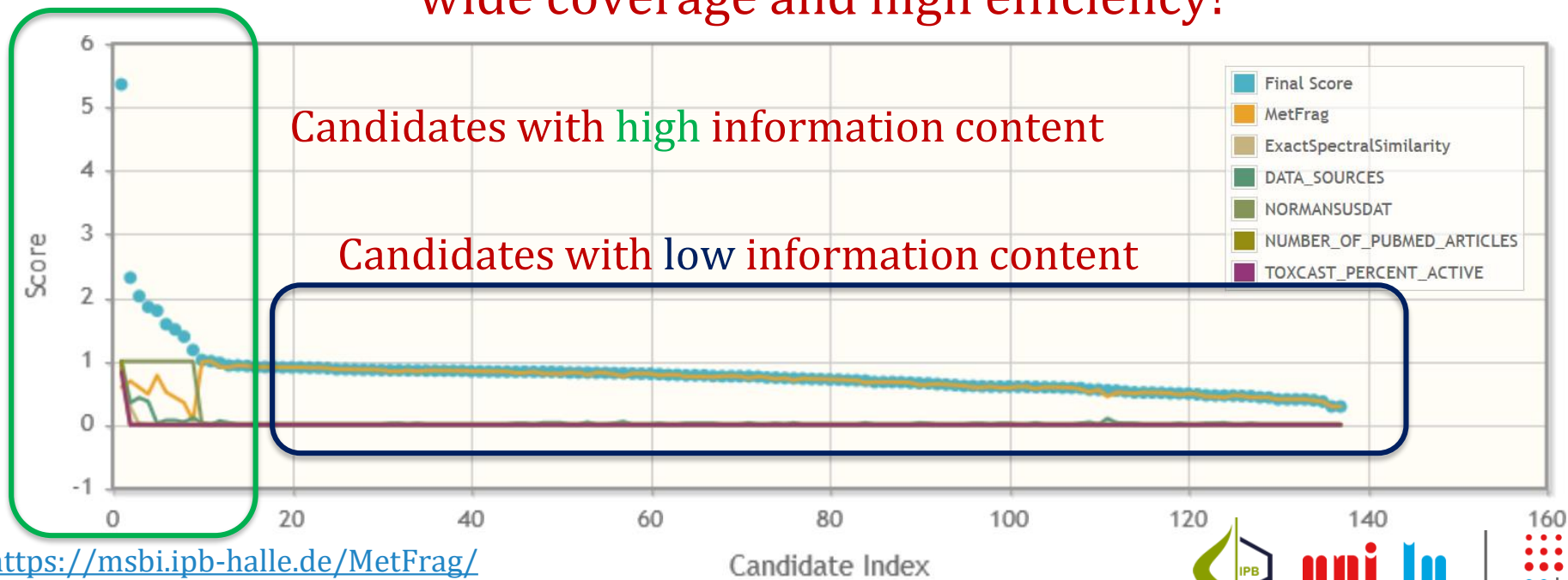


<https://msbi.ipb-halle.de/MetFrag/>
Ruttkies *et al.* (2016) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)



(5) MetFrag: Connecting Knowledge for Annotation

Challenge: the growing number of candidates ...
High throughput exposomics needs both
wide coverage and high efficiency!



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

Ruttkies *et al.* (2016) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)



(6) The 103 M PubChem Challenge



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

PubChem Compound TOC ? 33,765,953

Agrochemical Information ?	2,002
Biologic Description ?	1,539,532
Biological Test Results ?	3,467,416
Biomolecular Interactions and Pathways ?	109,610
Chemical and Physical Properties ?	237,729
Classification ?	18,569,627
Diseases ?	
Drug and Medication Information ?	15,955
Food Additives and Ingredients ?	7,447
Identification ?	5,746
Information Sources ?	20,654,780
Literature ?	1,668,437
Names and Identifiers ?	1,310,169
Patents ?	22,144,888
Pharmacology and Biochemistry ?	130,367
Related Records ?	5,297,096
Safety and Hazards ?	125,607
Spectral Information ?	761,478
Structures ?	5,926,225
Toxicity ?	114,554
Use and Manufacturing ?	108,745
Chemical Safety ?	122,739

103 million ... OR ...
the most relevant / annotated?

PubChemLite tier0: 316 K

PubChemLite tier1: 360 K



Database Settings

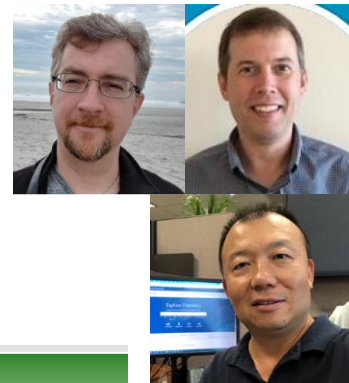
Database: PubChemLite_14Jan2020

Neutral Mass: 162.11576 Search ppm: 5

Formula: C10H14N2

Identifiers:

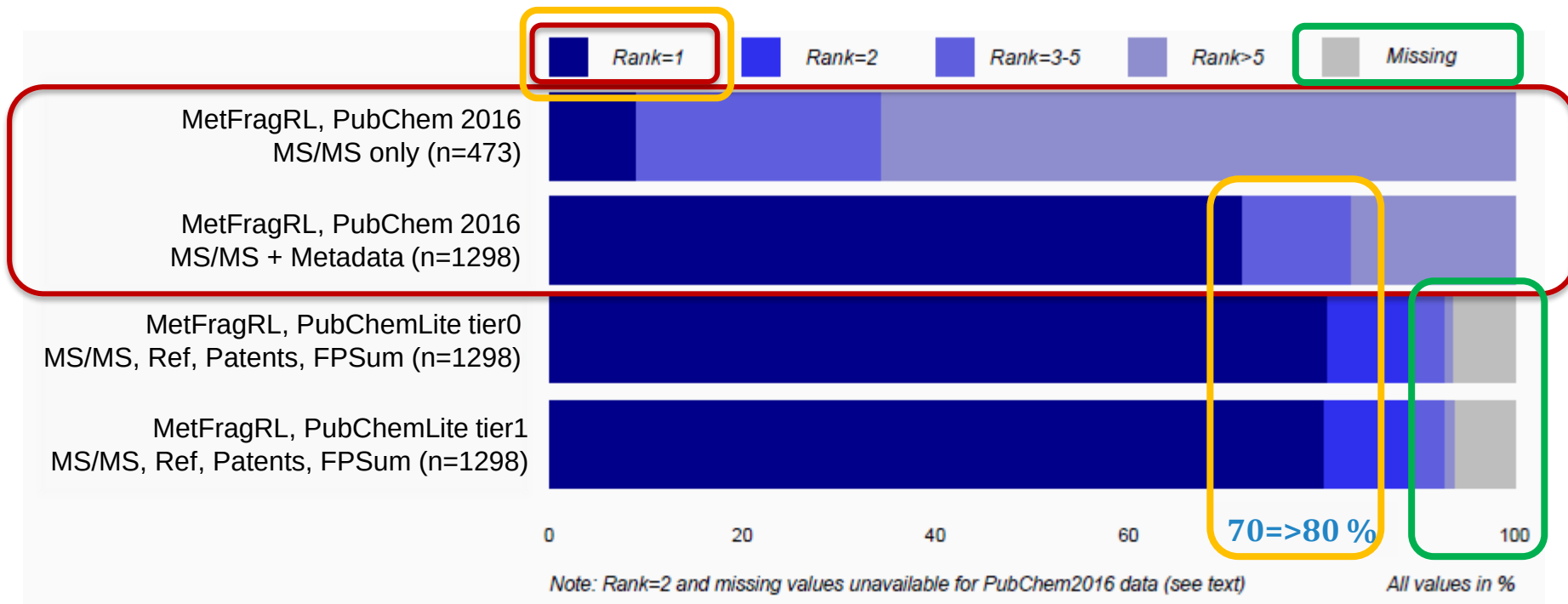
Retrieve Candidates 37 Candidates



Bolton & Schymanski (2020). PubChemLite tier0 and tier1
(Version PubChemLite.0.2.0) DOI: [10.5281/zenodo.3611238](https://doi.org/10.5281/zenodo.3611238)

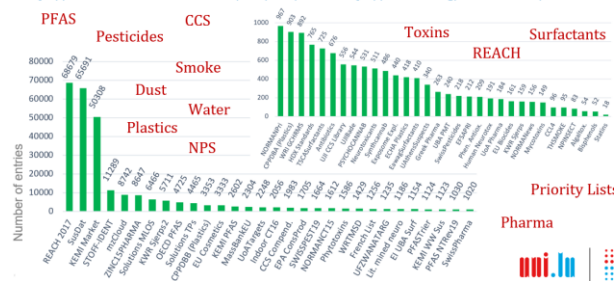


(7) Evaluating PubChemLite



MetFragRL: Ruttkies *et al.* (2016) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)
Bolton, Schymanski *et al.*, *in prep.* + Bolton & Schymanski (2020).
PubChemLite tier0 and tier1, DOI: [10.5281/zenodo.3611238](https://doi.org/10.5281/zenodo.3611238)





(8) DIY => Filling Annotation Gaps

Select classification

NORMAN Suspect List Exchange

Search selected classification by

Keyword Enter desired search term

Classification description (from NORMAN Suspect List Exchange)

The NORMAN Suspect List Exchange (NORMAN-SLE) is a central access point for NORMAN members (and others) to find su

Data type counts to display Display zero count nodes? Filter by Entrez History

None **Compound** **Yes** **No** Choose one

Browse NORMAN Suspect List Exchange Tree

- NORMAN Suspect List Exchange Classification ? 142,335
 - S25 | OECDPFAS | List of PFAS from the OECD ? 3,680
 - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites ? 1,358
 - Pesticide transformation products (TPs, metabolites, successors) ? 1,107
 - Pesticides (parents, predecessors) ? 267
 - S00 | SUSDAT | Merged NORMAN Suspect List: SusDat ? 66,880
 - S01 | MASSBANK | NORMAN Compounds in MassBank EU ? 16,179
 - S02 | STOFFIDENT | HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances ? 11,924
 - S03 | NORMANCT15 | NORMAN Collaborative Trial Targets and Suspects ? 651
 - S04 | UJIBADE | Target List from UJI used in Bade et al 2015 ? 573
 - S05 | KWRSJERPS | KWR Drinking Water Suspect List ? 164
 - S06 | ITNANTIBIOTIC | Antibiotic List from the ITN MSCA ANSWER ? 747

PubChem NORMAN Suspect List Exchange

124,863 Live Substances

Data in PubChem: 5,038 Annotations

1 Classification

Last Updated: 2020/04/01

Evan Bolton, Jeff (Jian) Zhang, Paul Thiessen and the entire PubChem team!

<https://pubchem.ncbi.nlm.nih.gov/source/23819>



(8) DIY => Filling Annotation Gaps

Select classification **NORMAN Suspect List Exchange** Search selected classification by **Keyword** Enter desired search term

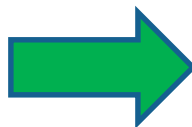
Classification description (from **NORMAN Suspect List Exchange**)
The NORMAN Suspect List Exchange (NORMAN-SLE) is a central access point for NORMAN members (and others) to find su

Data type counts to display Display zero count nodes? Filter by Entrez History

None **Compound** **Yes** **No** Choose one

Browse NORMAN Suspect List Exchange Tree

- NORMAN Suspect List Exchange Classification ? 142,335
 - S25 | OECD PFAS | List of PFAS from the OECD ? 3,680
 - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites ? 1,358
 - Pesticide transformation products (TPs, metabolites, successors) ? 1,107
 - Pesticides (parents, predecessors) ? 267
 - S00 | SUSDAT | Merged NORMAN Suspect List: SusDat ? 66,880
 - S01 | MASSBANK | NORMAN Compounds in MassBank EU ? 16,179
 - S02 | STOFFIDENT | HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances ? 11,924
 - S03 | NORMANCT15 | NORMAN Collaborative Trial Targets and Suspects ? 651
 - S04 | UJIBADE | Target List from UJI used in Bade et al 2015 ? 573
 - S05 | KWRSJERPS | KWR Drinking Water Suspect List ? 164
 - S06 | ITNANTIBIOTIC | Antibiotic List from the ITN MSCA ANSWER ? 747



PubChem Folpet (Compound)

8.7 Transformations

3 items View More Details

Download

SORT BY Please Choose One					
Predecessor Image	Predecessor Name	Transformation	Successor Image	Successor Name	Evidence DOI
	Folpet	Environmental Transformation		Phthalimide	10.1016/j.watres.2019.114972
	Folpet	Environmental Transformation		Phthalamic acid	10.1016/j.watres.2019.114972
	Folpet	Environmental Transformation		Phthalic acid	10.1016/j.watres.2019.114972

Evan Bolton, Jeff (Jian) Zhang, Paul Thiessen
and the entire PubChem team!

Acknowledgements!

patRoon: Rick Helmus

ShinyScreen: Todor Kondic & ECI team

NORMAN-SLE: NORMAN contributors

MetFrag: Steffen Neumann, Christoph Ruttkies

PubChem: Evan Bolton, Jian (Jeff) Zhang,
Paul Thiessen & the whole team

My many other collaborators over the years!



emma.schymanski@uni.lu @ESchymanski

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

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

Join us! <http://emea3.mrted.ly/2g7vc>

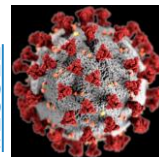
https://covid.pages.uni.lu/map_curation



ECI: Todor, Emma, Anjana, Hiba, Mira*, Corey, Lorenzo, German*, Adelene, Randolph, (in.) Simone, Véro, Rick*, Jessy. *Ext/Alumni



UNIVERSITY OF AMSTERDAM



#23312



Luxembourg National
Research Fund

