Non-target Screening for Holistic Chemical Monitoring and Compound Discovery: Open Science, Real-time and Retrospective Approaches

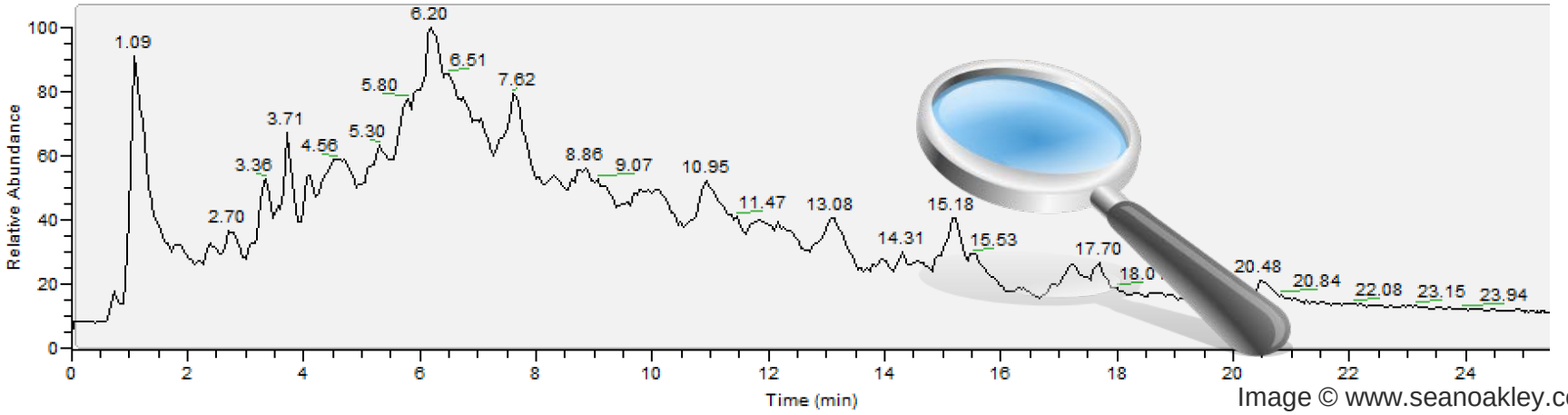
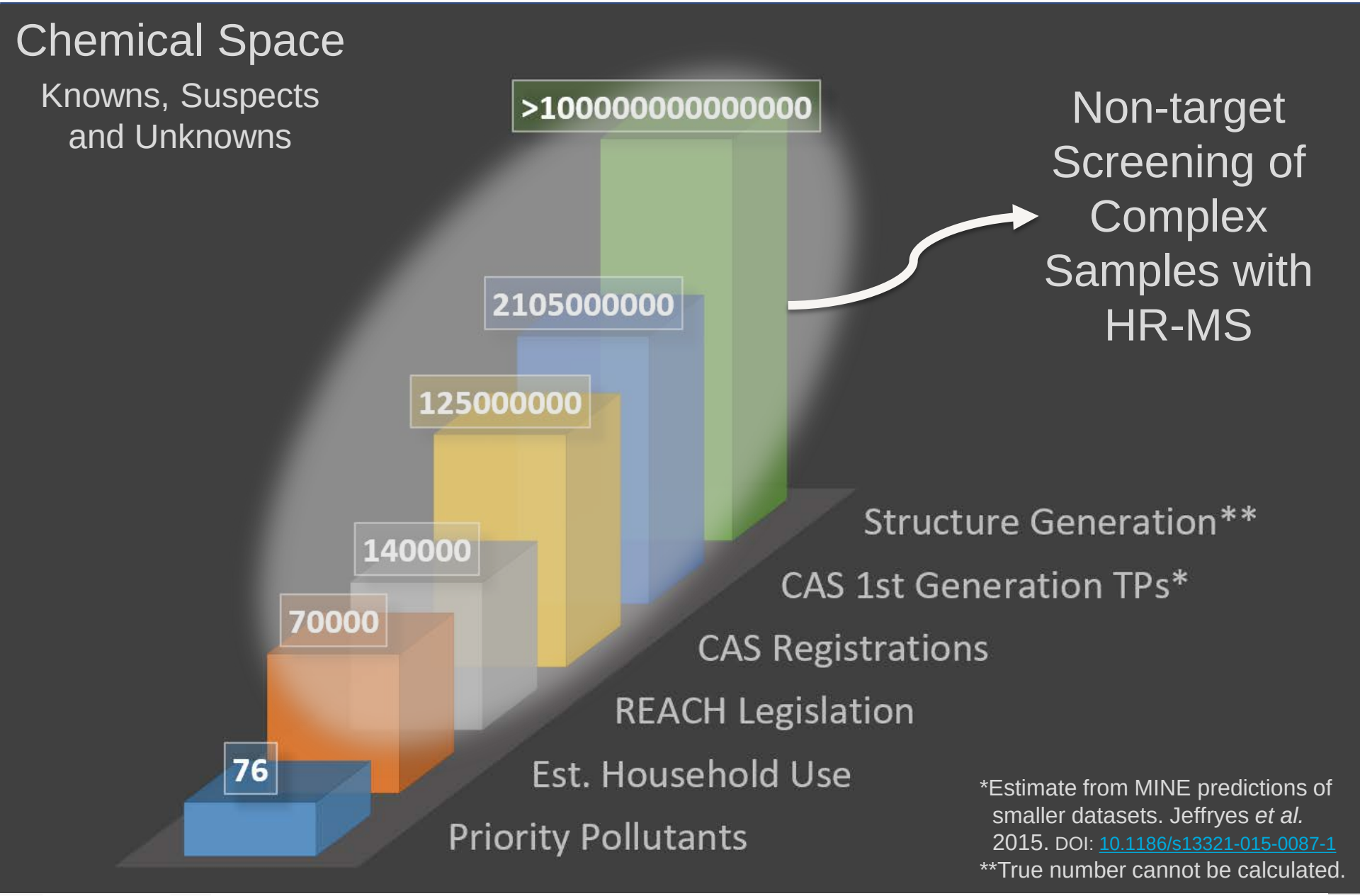


Image © www.seanoakley.com/

Emma Schymanski
Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg.
Email: emma.schymanski@uni.lu

Reza Aalizadeh, Nikiforos Aligizakis, Juliane Hollender, Martin Krauss, Tobias Schulze, Jaroslav Slobodnik, Nikolaos S. Thomaidis, Antony J. Williams

What chemicals are out there? What to monitor?

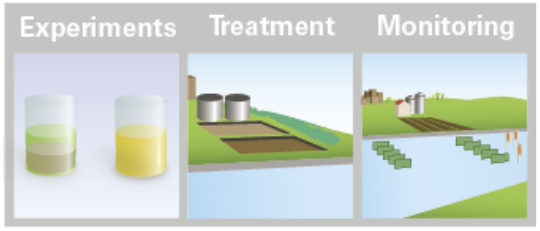


Numbers from Hollender, Schymanski, Singer & Ferguson, 2018, ES&T Feature, 51:20, 11505-11512. DOI: 10.1021/acs.est.7b02184

Non-target Screening for Chemical Monitoring

1) Open Science
2) Real-Time & Retrospective Screening

MassBank.eu | norman suspects | EPA Chemistry Dashboard | MetFrag



Commercial / Open Software | Databases | In Silico Prediction

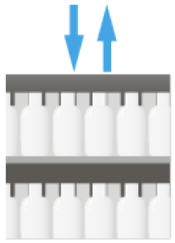
Sampling
Space & Time
Before & After
Hot spots

Analysis
(Sample enrichment)
Chromatography (GC, LC)
Appropriate Ionization
HR-tandem MS

Data Pre-processing
Detect Peaks
Subtract Blind
Replicates & Grouping
Build Profiles, Series

Prioritiziation
Isotopes, Mass Defect
Time/ Spatial Trends
PCA/Clustering
Processes, Effects

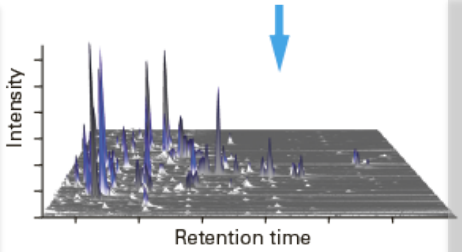
Identification
Database, Suspect Search
Assign Formula, Fragments
Analytical (RT, Ionization,..) & Meta Information



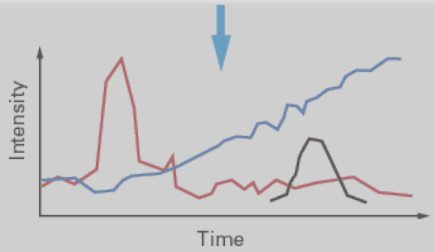
SAMPLE ARCHIVE

DIGITAL ARCHIVE

Retrospective

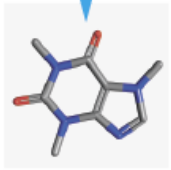


POLLUTION OVERVIEW



PATTERN

Retrospective
Real-Time



STRUCTURES

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

1) Open Science: MassBank EU

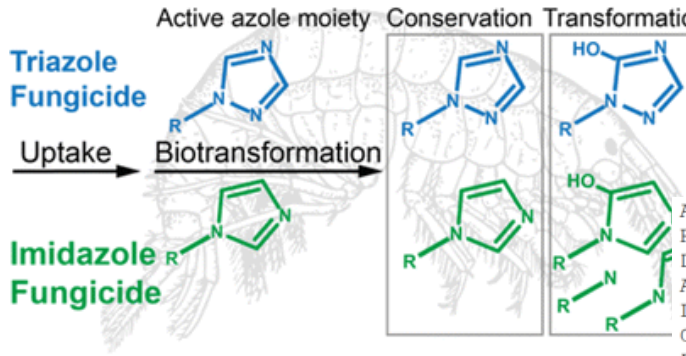
MassBank.eu

http://massbank.eu/MassBank

- MassBank.EU was founded late 2012, hosted at UFZ, Leipzig, Germany
  - **>16,000 MS/MS** spectra; **1,200 substances** from **NORMAN members**
  - MassBank now has >46,000 spectra from 32 contributing institutes!
  - Thorough Github-based modernization ***in progress*** for traceability:

MassBank-data validation status build passing

  - ***Tentative/unknown/literature*** spectra on massbank.eu (not massbank.jp)




```
ACCESSION: ET220001
RECORD_TITLE: Epoxiconazole (EP); LC-ESI-QFT; MS2; CE: 25; R=7
DATE: 2016.01.07
AUTHORS: A. Roesch, E. Schymanski, J. Hollender, Department of
LICENSE: CC BY
COPYRIGHT: Copyright (C) 2015 Eawag, Duebendorf, Switzerland
PUBLICATION: Roesch A, Anliker S, Hollender J, How Biotransfor
COMMENT: CONFIDENCE Parent Substance (Level 1)
COMMENT: INTERNAL_ID 2200
```

```
ACCESSION: ET22060
RECORD_TITLE: EP_M
DATE: 2016.03.01
AUTHORS: A. Roesch
LICENSE: CC BY
COPYRIGHT: Copyright (C) 2015 Eawag, Duebendorf, Switzerland
PUBLICATION: Roesch A, Anliker S, Hollender J, How Biotransformation Influences Toxicokinetics of A
COMMENT: CONFIDENCE Probable structure via diagnostic evidence, tentative identification (Level 2b)
COMMENT: INTERNAL_ID 2206
```


https://github.com/MassBank/MassBank-data; https://github.com/MassBank/MassBank-web/; Rösch *et al* DOI 10.1021/acs.est.5b05186

Creating High-Quality Mass Spectra

RMassBank

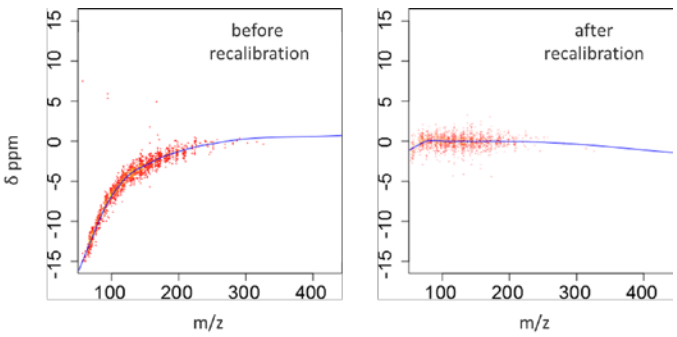
LevelCode	LevelKeyword	COMMENT.CONFIDENCE
1	standard	Reference Standard (Level 1)
1a	standard	Reference Standard (Level 1)
1b	parent	Parent Substance with Reference Standard (Level 1)
1c	confirmed	Identification confirmed with Reference Standard (Level 1)
2	probable	Probable structure, tentative identification (Level 2)
2a	probableLibrary	Probable structure via library match, tentative identification (Level 2a)
2b	probableDiagnostic	Probable structure via diagnostic evidence, tentative identification (Level 2b)
3	tentative	Tentative identification only (Level 3)
3a	tentativeStructure	Tentative identification: most likely structure (Level 3)
3b	tentativeIsomer	Tentative identification: isomers possible (Level 3)
3c	tentativeTPClass	Tentative identification: substance class known (Level 3)
3d	tentativeBestMatch	Tentative identification: best match only (Level 3)
4	formula	Tentative identification: molecular formula only (Level 4)
5	unknown	Tentative identification: structure and formula unknown (Level 5)
5	exactMass	Tentative identification: structure and formula unknown (Level 5)

LC-MS/MS raw data

online resources: CTS, CACTUS

Automatic MS and MS/MS Recalibration and Clean-up Remove interfering peaks

Spectral Annotation with
- Experimental Details
- Compound Information



MassBank records

structure files

norman

MassBank.eu

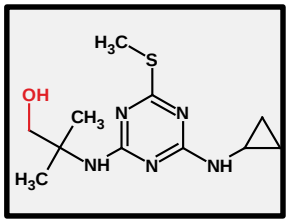
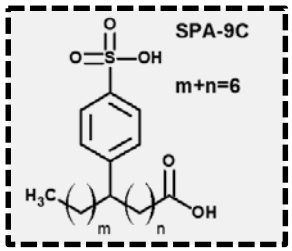
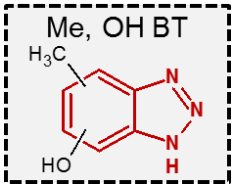
16,004 (61 %*) MS/MS spectra
1,269 (18 %*) substances
*% of ***all open*** LC-MS/MS data

https://github.com/MassBank/RMassBank/
http://bioconductor.org/packages/RMassBank/

Stravs, Schymanski, Singer and Hollender, 2013, Journal of Mass Spectrometry, 48, 89–99. DOI: 10.1002/jms.3131

Confidence Levels for Tentative Structures

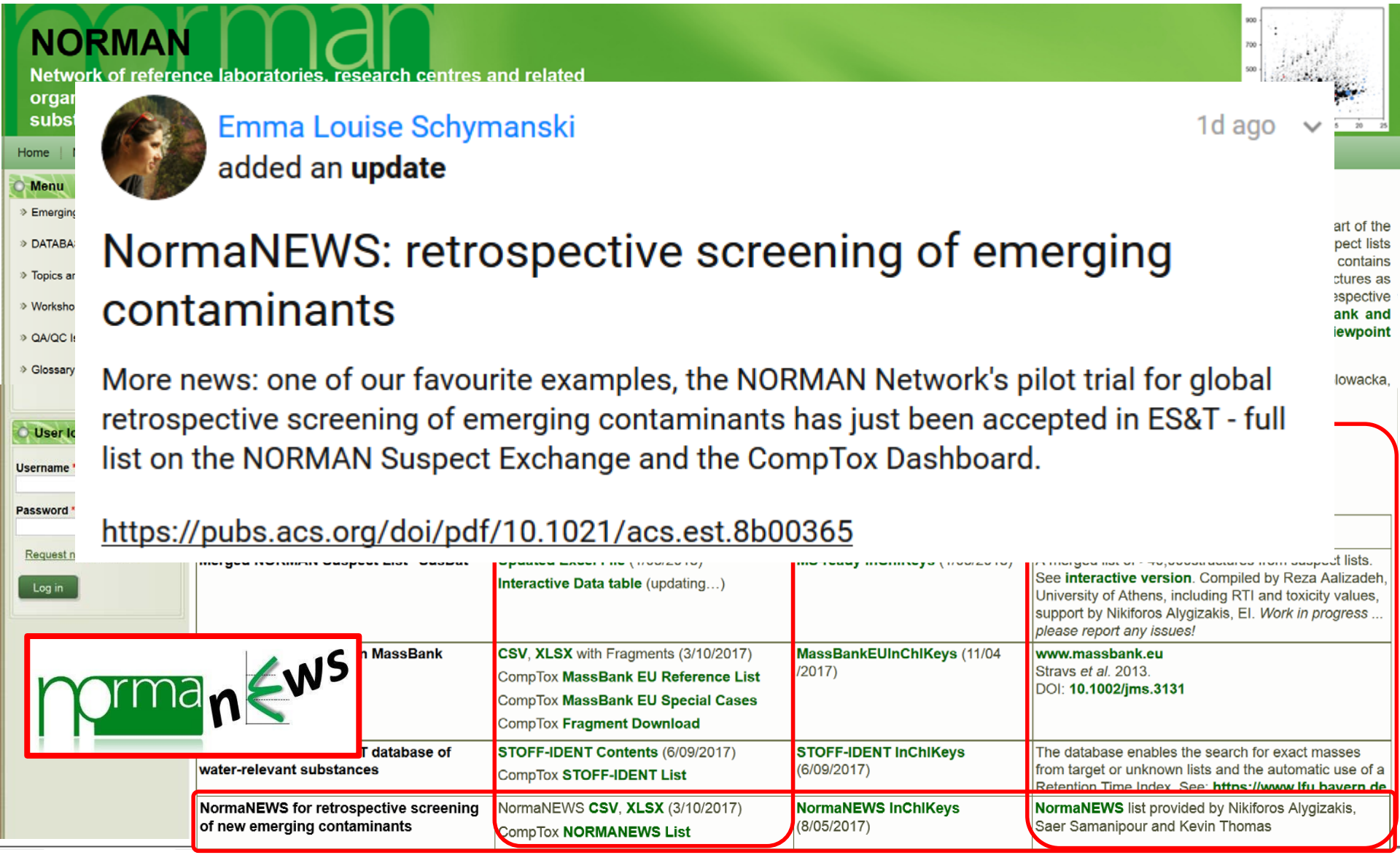
- Annotation is the key to communicating information

Example	*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, MS^2, Library MS^2 MS, MS^2, Exp. data
 	***Level 3: Tentative candidate(s)*** structure, substituent, class	MS, MS^2, Exp. data
$C_6H_5N_3O_4$	***Level 4: Unequivocal molecular formula***	MS isotope/adduct
192.0757	***Level 5: Exact mass*** of interest	MS

Schymanski, Jeon, Gulde, Fenner, Ruff, Singer & Hollender (2014) ES&T, 48 (4), 2097-2098. DOI: 10.1021/es5002105

NORMAN Suspect List Exchange

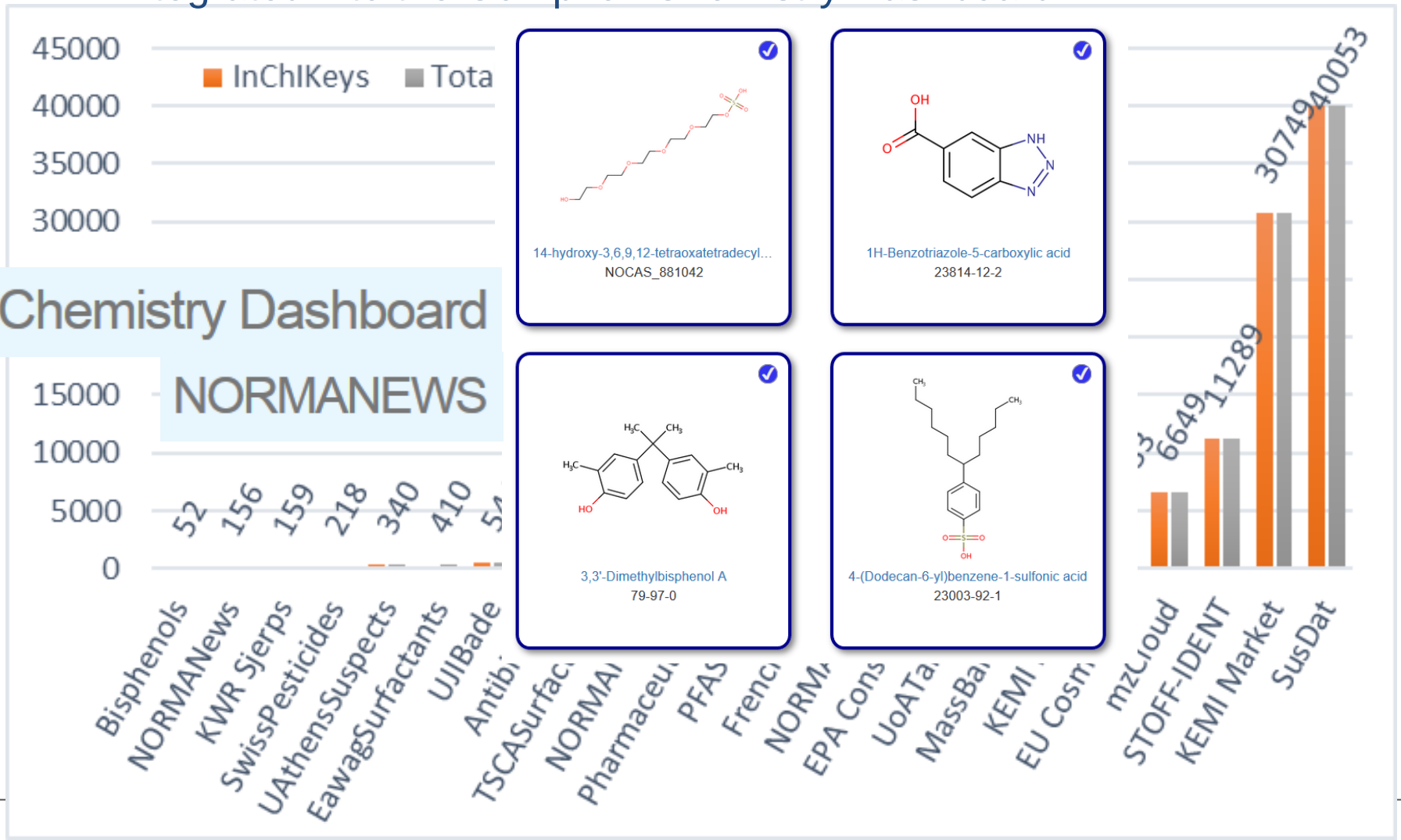
- http://www.norman-network.com/?q=node/236



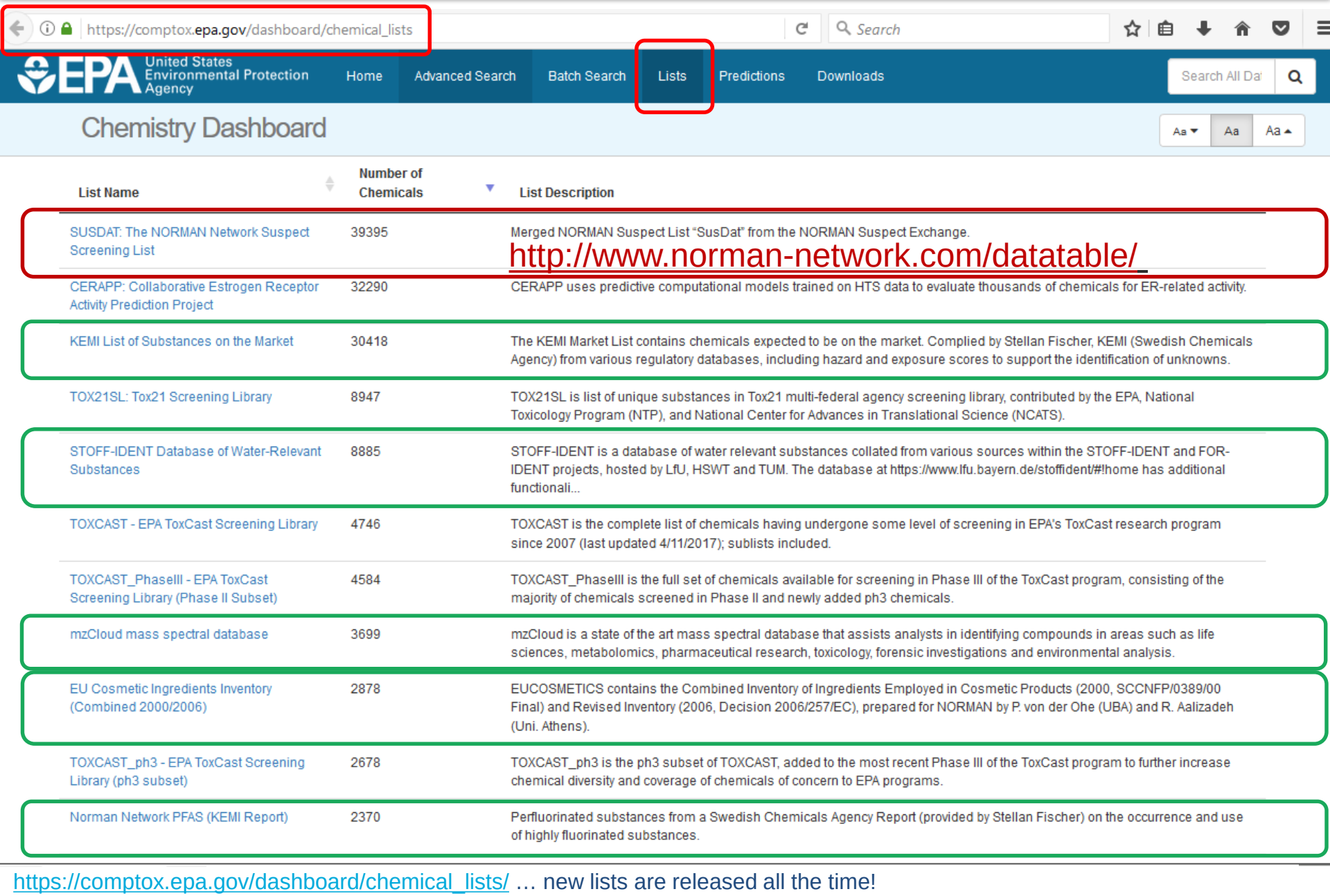
Schymanski, Aalizadeh *et al*. in prep; https://www.researchgate.net/project/Supporting-Mass-Spectrometry-Through-Cheminformatics

NORMAN Suspect Exchange Lists

- Now 21 lists available online … from small to large!
  - Specialist collections (e.g. NormaNEWS) to market lists
  - Integrated into the CompTox Chemistry Dashboard



CompTox Chemistry Dashboard



List Name	Number of Chemicals	List Description
SUSDAT: The NORMAN Network Suspect Screening List	39395	Merged NORMAN Suspect List "SusDat" from the NORMAN Suspect Exchange. http://www.norman-network.com/datatable/
CERAPP: Collaborative Estrogen Receptor Activity Prediction Project	32290	CERAPP uses predictive computational models trained on HTS data to evaluate thousands of chemicals for ER-related activity.
KEMI List of Substances on the Market	30418	The KEMI Market List contains chemicals expected to be on the market. Complied by Stellan Fischer, KEMI (Swedish Chemicals Agency) from various regulatory databases, including hazard and exposure scores to support the identification of unknowns.
TOX21SL: Tox21 Screening Library	8947	TOX21SL is list of unique substances in Tox21 multi-federal agency screening library, contributed by the EPA, National Toxicology Program (NTP), and National Center for Advances in Translational Science (NCATS).
STOFF-IDENT Database of Water-Relevant Substances	8885	STOFF-IDENT is a database of water relevant substances collated from various sources within the STOFF-IDENT and FOR-IDENT projects, hosted by LfU, HSWT and TUM. The database at https://www.lfu.bayern.de/stoffident/#!home has additional functionali...
TOXCAST - EPA ToxCast Screening Library	4746	TOXCAST is the complete list of chemicals having undergone some level of screening in EPA's ToxCast research program since 2007 (last updated 4/11/2017); sublists included.
TOXCAST_PhaseIII - EPA ToxCast Screening Library (Phase II Subset)	4584	TOXCAST_PhaseIII is the full set of chemicals available for screening in Phase III of the ToxCast program, consisting of the majority of chemicals screened in Phase II and newly added ph3 chemicals.
mzCloud mass spectral database	3699	mzCloud is a state of the art mass spectral database that assists analysts in identifying compounds in areas such as life sciences, metabolomics, pharmaceutical research, toxicology, forensic investigations and environmental analysis.
EU Cosmetic Ingredients Inventory (Combined 2000/2006)	2878	EUCOSMETICS contains the Combined Inventory of Ingredients Employed in Cosmetic Products (2000, SCCNFP/0389/00 Final) and Revised Inventory (2006, Decision 2006/257/EC), prepared for NORMAN by P. von der Ohe (UBA) and R. Aalizadeh (Uni. Athens).
TOXCAST_ph3 - EPA ToxCast Screening Library (ph3 subset)	2678	TOXCAST_ph3 is the ph3 subset of TOXCAST, added to the most recent Phase III of the ToxCast program to further increase chemical diversity and coverage of chemicals of concern to EPA programs.
Norman Network PFAS (KEMI Report)	2370	Perfluorinated substances from a Swedish Chemicals Agency Report (provided by Stellan Fischer) on the occurrence and use of highly fluorinated substances.

https://comptox.epa.gov/dashboard/chemical_lists/ … new lists are released all the time!

…All Combined in MetFrag

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

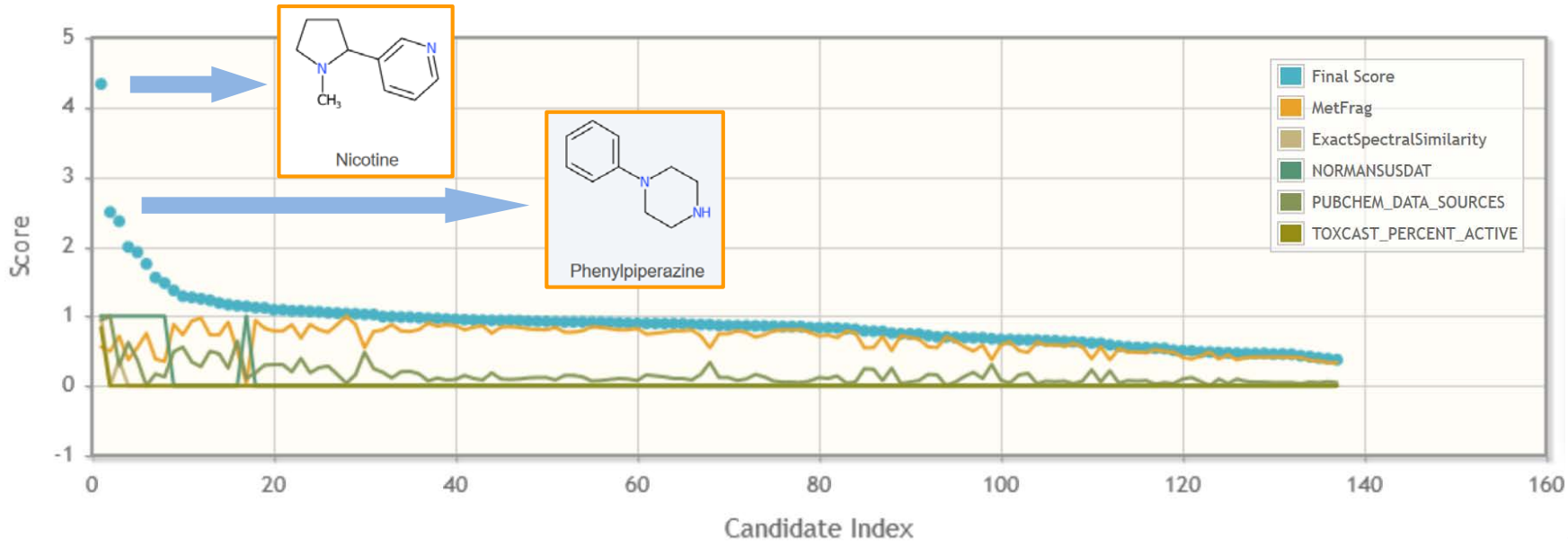
Search

Results

Weights		
MetFrag (1st)	100	%
ExactSpectralSimilarity (2nd)	100	%
NORMANSUSDAT (3rd)	100	%
PUBCHEM_DATA_SOURCES (4th)	100	%
TOXCAST_PERCENT_ACTIVE (5th)	100	%

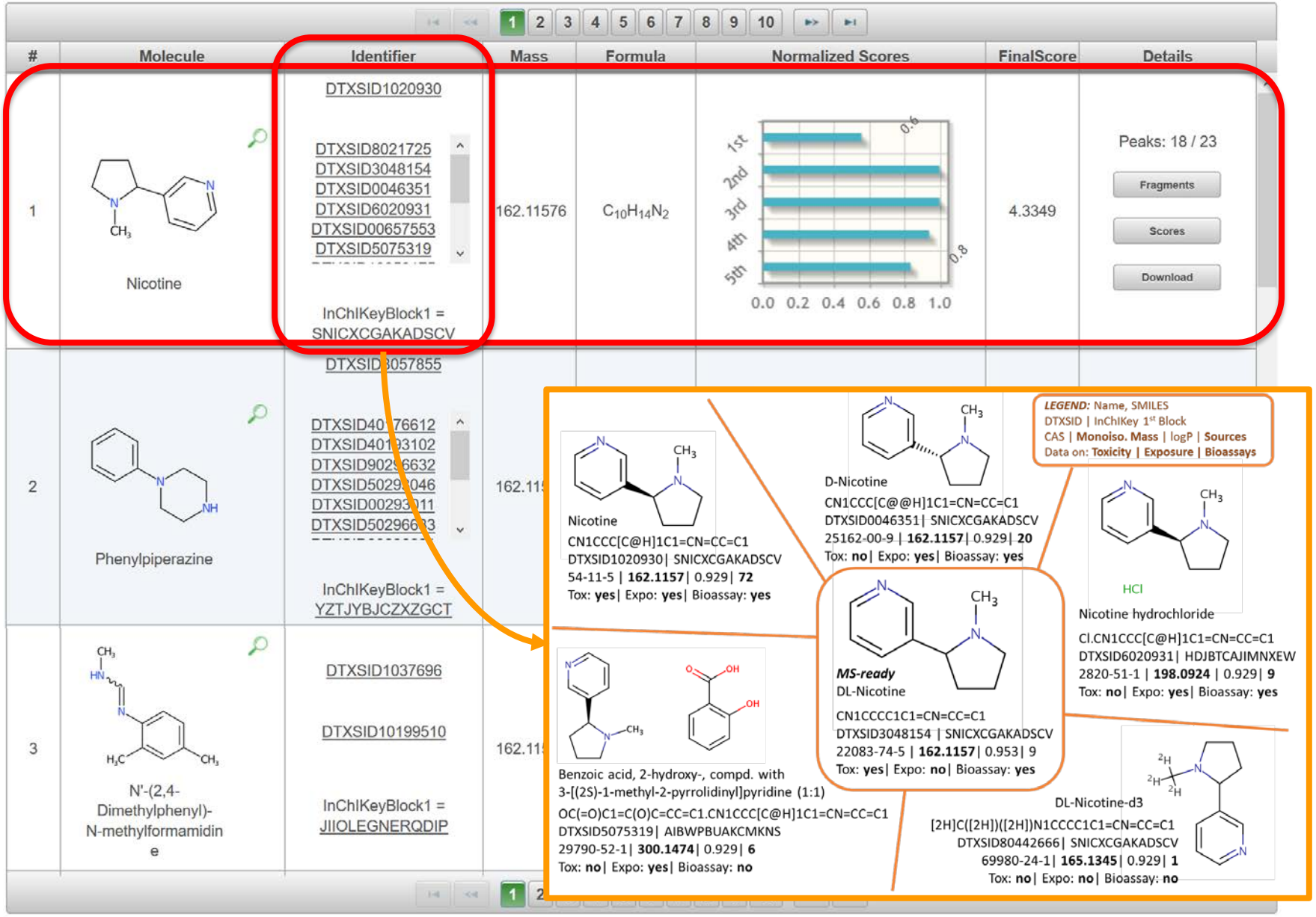
Statistics

Candidate Score Distribution



https://msbi.ipb-halle.de/MetFragBeta/ AND https://comptox.epa.gov/dashboard/dsstoxdb/batch_search (MetFrag Export)

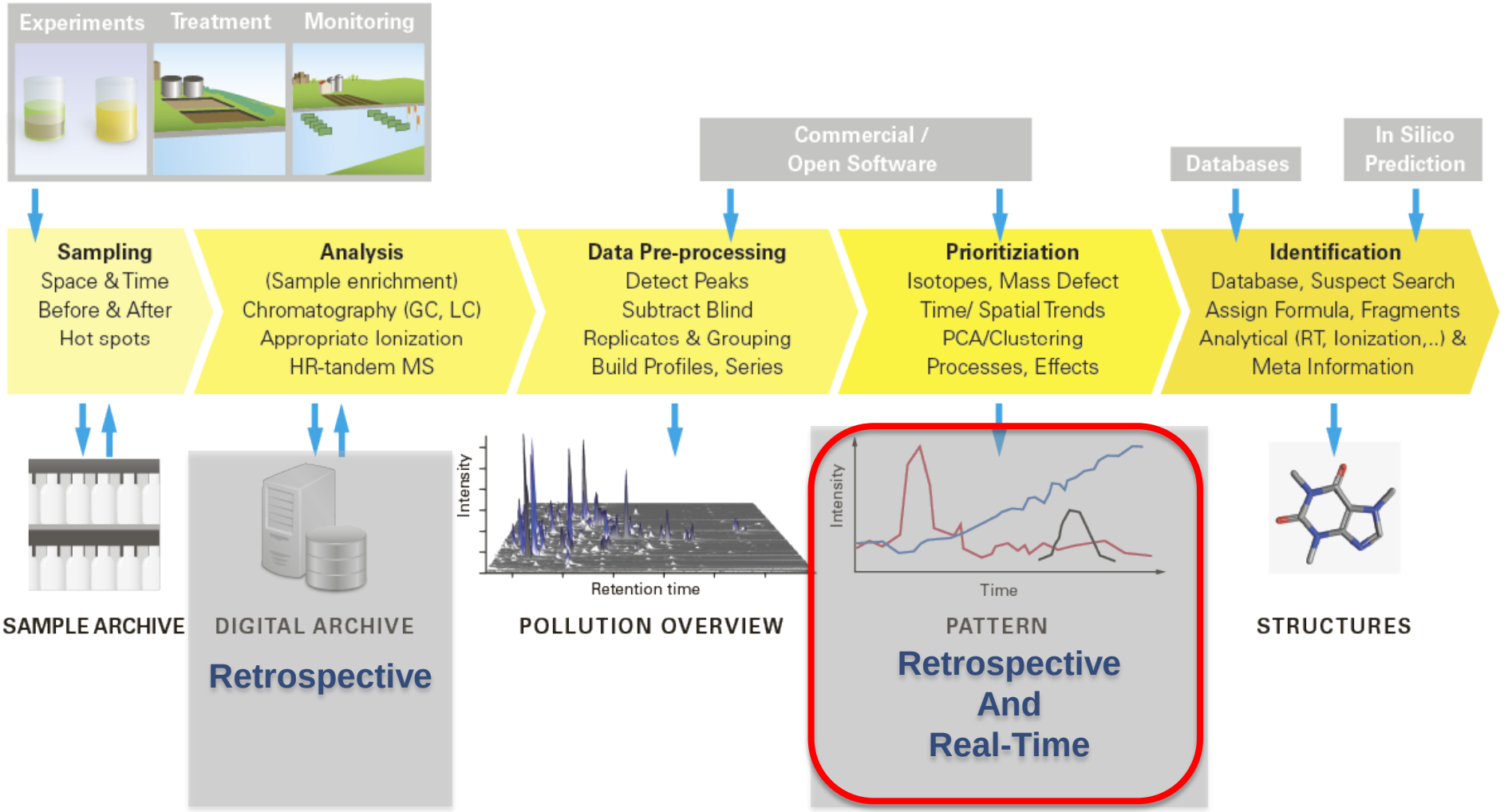
…All Combined in MetFrag



Example from: Schymanski & Williams, 2017, ES&T, 51 (10), pp 5357–5359. DOI: 10.1021/acs.est.7b01908

Non-target Screening for Chemical Monitoring

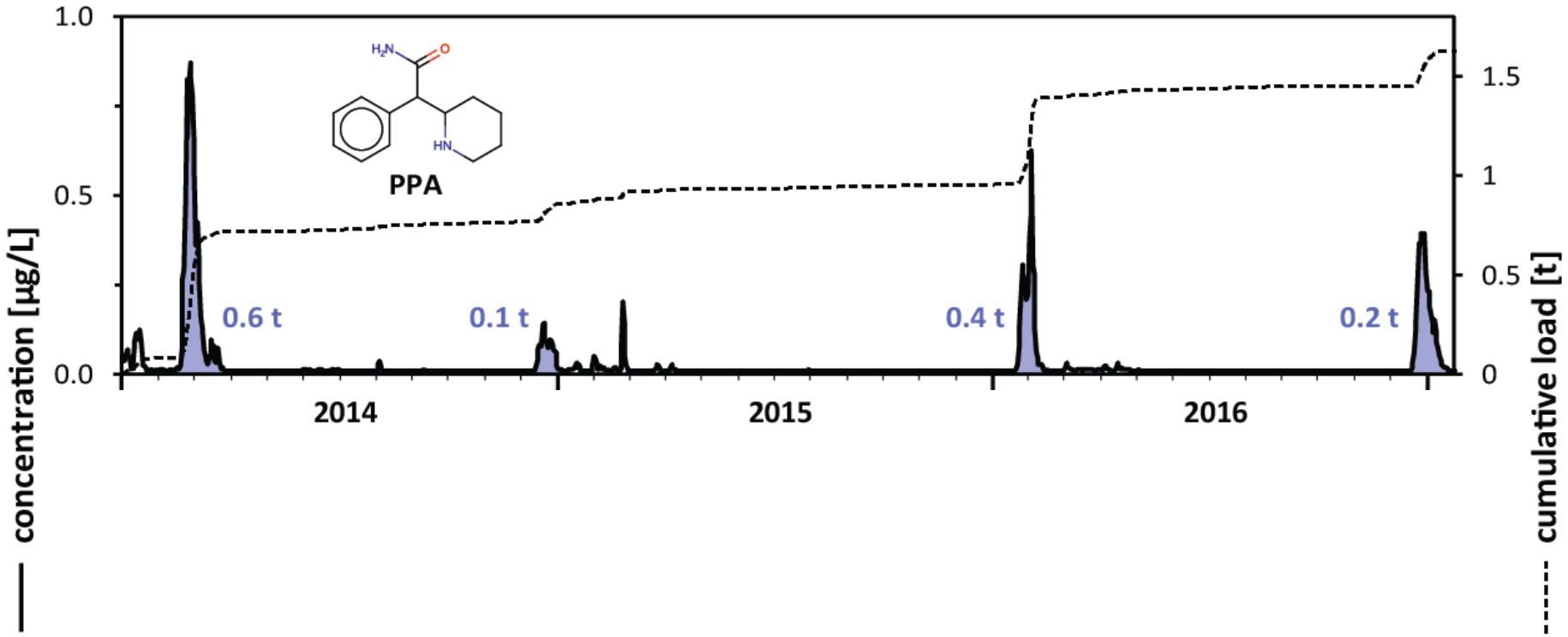
o Part 2: Real-Time and Retrospective Screening



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

Real-time Monitoring of the Rhine River

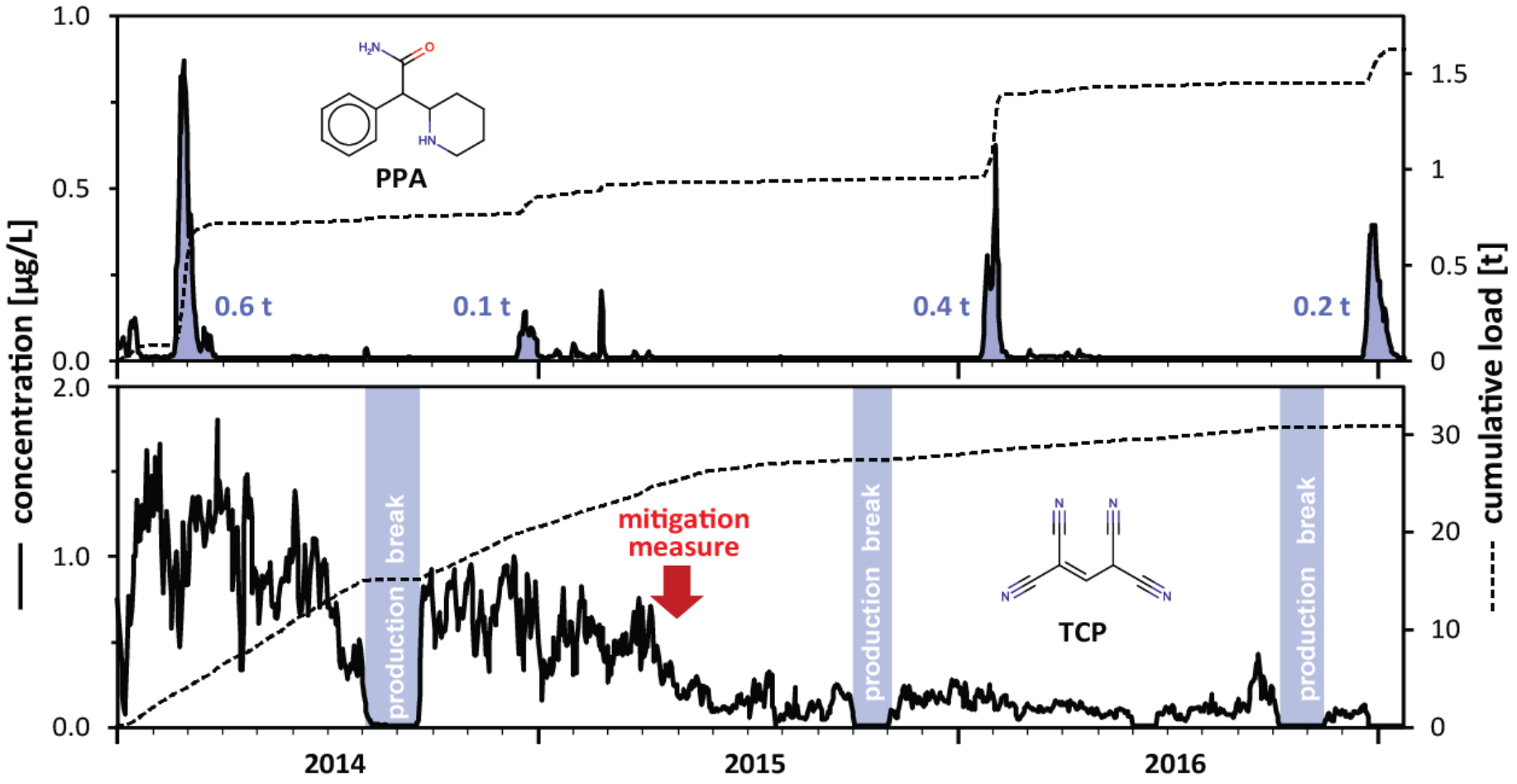
Previously unknown chemicals detected due to “stand-out” patterns



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

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

European (World-)Wide Exchange of Suspects

Tentatively Identified Spectra:
http://goo.gl/0t7jGp
Hits in GNPS MassIVE datasets:
TPs in skin: http://goo.gl/NmO4tx
Surfactants: http://goo.gl/7sY9Pf

MassBank.eu

NORMAN Suspect List Exchange:
http://www.norman-network.com/?q=node/236

Back to main page | Back to status page | Collapse all | Download

Continuous ID Search: MSV000078934 - GNPS_CAICE_CARB_C18_Aerosol_Headspace_Samples_NEGATIVE_POLARITY_Maxis_Impact_LCMS_

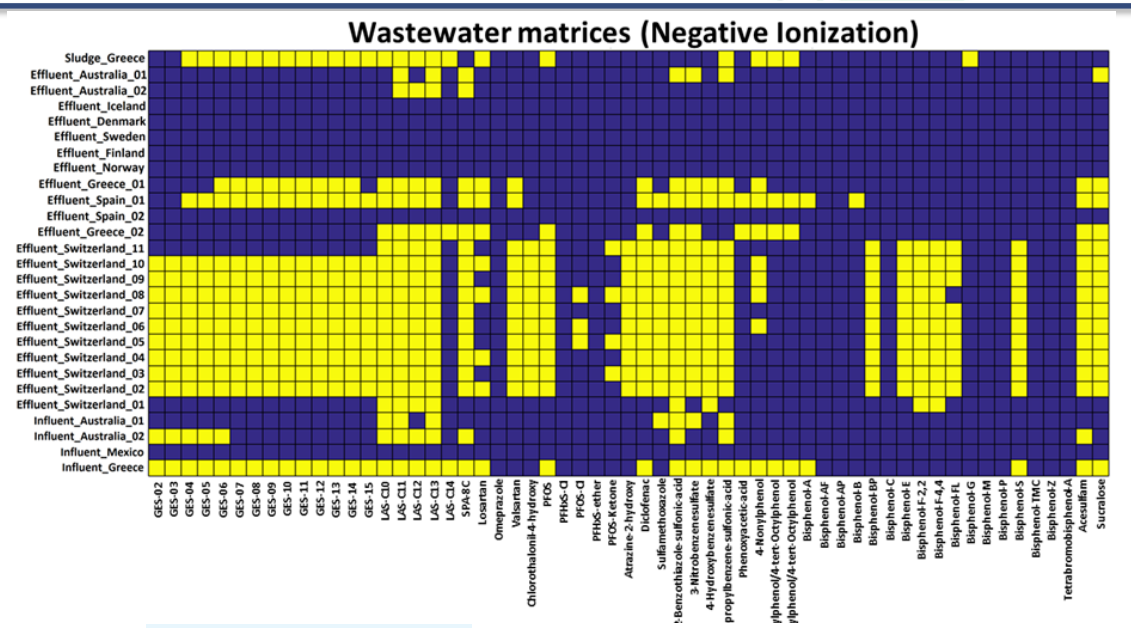
Hits 1 ~ 30 out of 3072 Go to Go

Select columns

Filter	ClusterIndex	NumSpectra	PrecursorMZ	PrecursorInt	RTMean	DefaultGroups	LibraryID
Show Analogs 1	9752	8	311.77600	648800.00000	436	G1,	MassbankEU:ETS00014 C11-LAS (STANDARD MIX)|C11-linear alkylbenzyl sulfonate|4-(undecan-5-yl)benzenesulfonic acid
Show Analogs 2	9776	76	311.16800	18631200.00000	505	G1,	MassbankEU:ETS00014 C11-LAS (STANDARD MIX)|C11-linear alkylbenzyl sulfonate|4-(undecan-5-yl)benzenesulfonic acid
Show Analogs 3	1	26	159.19180	11038200.00000	483	G1,	
Show Analogs 4	4	7	173.11400	5976280.00000	140	G1,	

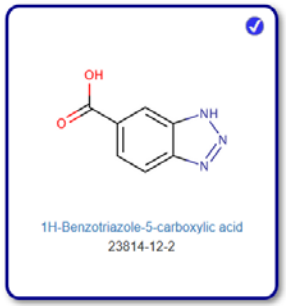
Schymanski et al. 2015, *ABC*, DOI: 10.1007/s00216-015-8681-7; Wang *et al* 2016 *Nature Biotechnology*, DOI: 10.1038/nbt.3597

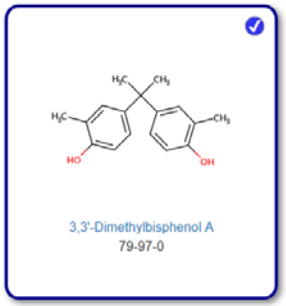
World-Wide Exchange: Emerging Suspects

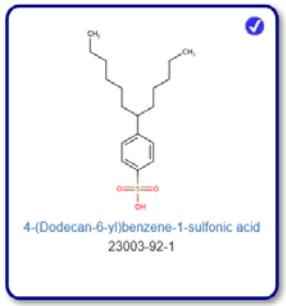


Chemistry Dashboard

NORMANEWS





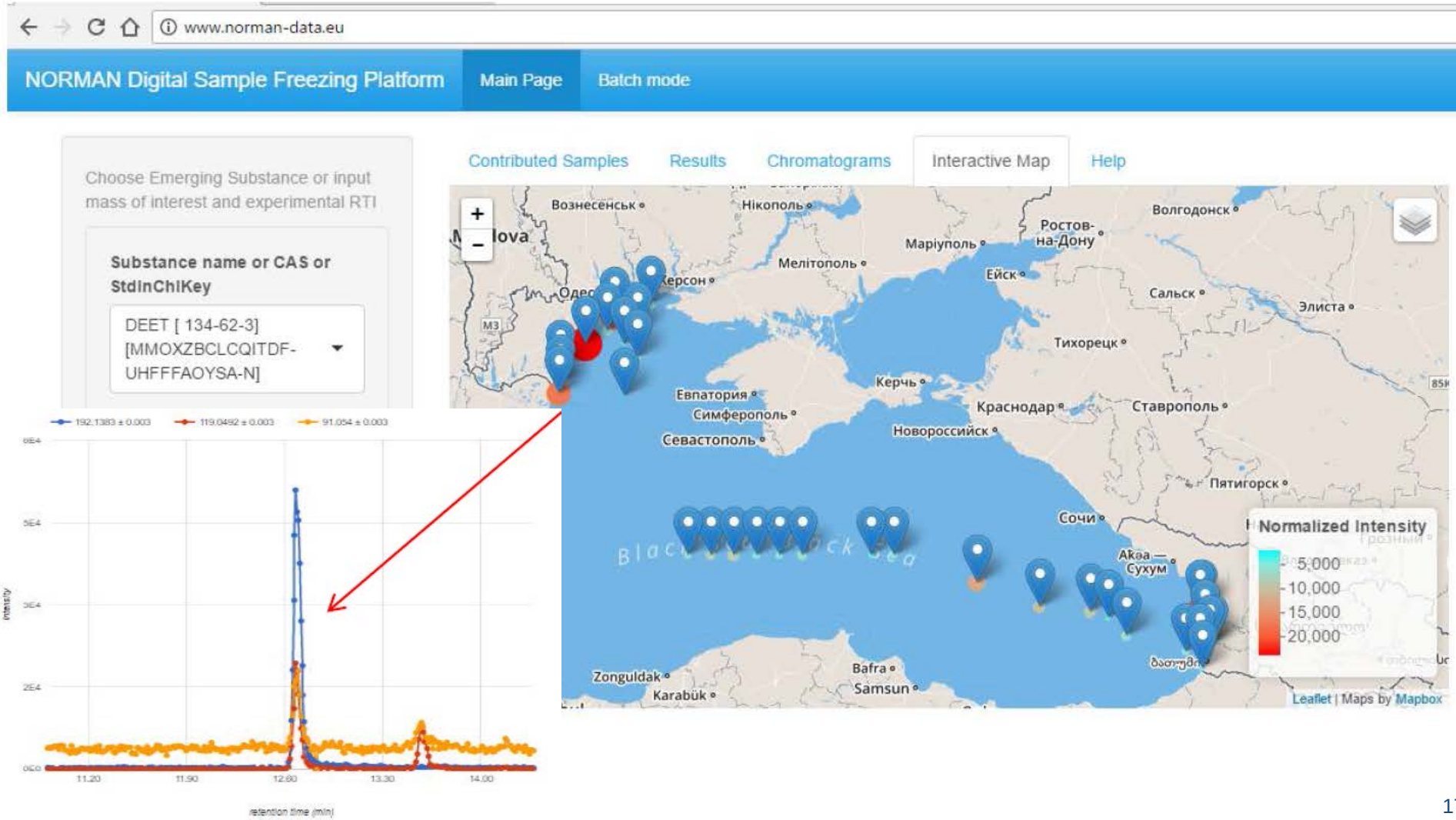


Map images © Google Maps

Alygizakis et al. 2018 ES&T, DOI: 10.1021/acs.est.8b00365. https://comptox.epa.gov/dashboard/chemical_lists/normanews

NORMAN Digital Sample Freezing Platform

"Live" retrospective screening of known and unknown chemicals in European samples (various matrices)



Aligizakis *et al*, in prep.

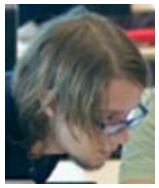
Acknowledgements I






solutions

EU Grant 603437








ENVIRONMENTAL INSTITUTE

MassBank.eu

norman suspects

MetFrag

EPA Chemistry Dashboard

eawag aquatic research

Bayerisches Landesamt für Umwelt

NIVA

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

KEMI Kemikalieinspektionen

National and Kapodistrian UNIVERSITY OF ATHENS

LCSB

uni.lu UNIVERSITÉ DU LUXEMBOURG

emma.schymanski@uni.lu

Further Information:

http://www.norman-network.com/?q=node/236

https://massbank.eu/MassBank/

https://comptox.epa.gov/dashboard/

https://www.researchgate.net/project/Supporting-Mass-Spectrometry-Through-Cheminformatics

https://github.com/MassBank/


Community Efforts!
modelkey
ART2 CURE
MassBank consortium
norman
Nathan Yates
Emma Schymanski
Nuno Bandeira
Pietro Franceschi
Juliane Hollander
Steve Stein
Enrico Davoli
© SCHLOSS DAGSTUHL – LZI GMBH
licensed under Creative Commons License CC BY-NC-ND