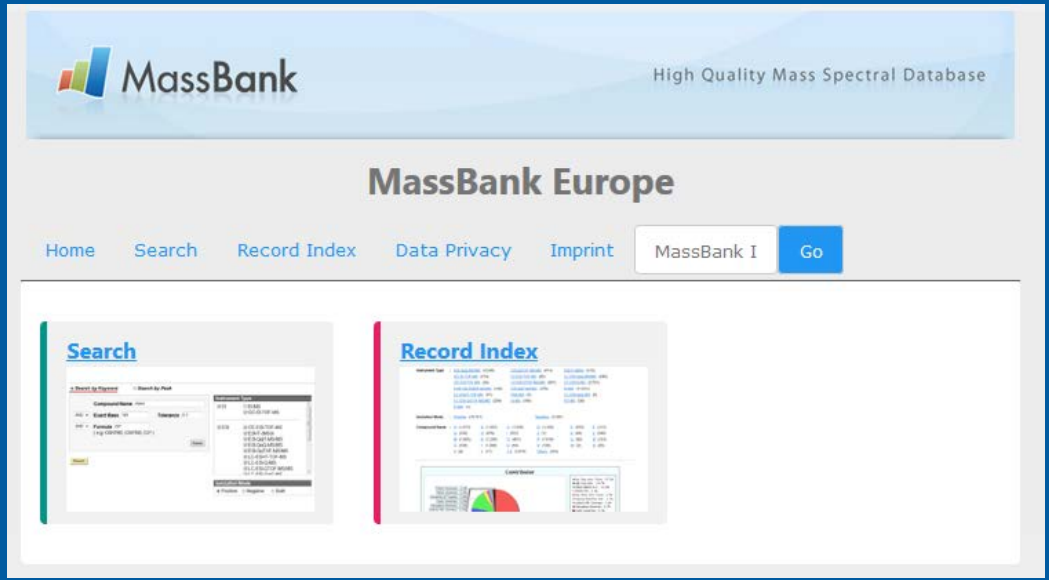
MassBank Europe – a success story of community efforts

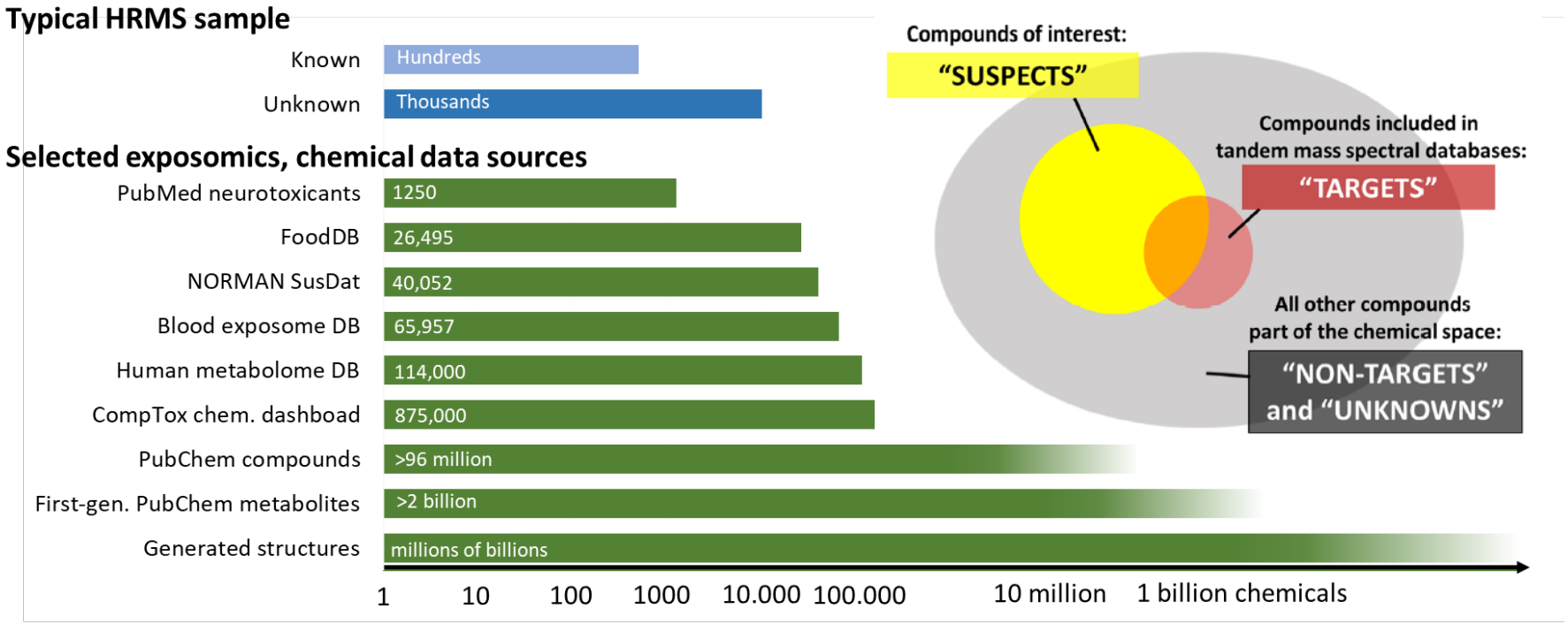
Tobias Schulze, UFZ, Germany
Rene Meier, IPB, Germany
Steffen Neumann, IPB, Germany
Herbert Oberacher, MUI, Austria
Emma Schymanski, LCSB, Luxembourg

ASMS 2020 Reboot, June 3, 2020
Live Webinar Workshop on Compound
Identfication by Mass Spectral Library Searching



DOI: 10.5281/zenodo.3865489

Our (community) challenge: Identifying compounds



DOI: 10.1126/science.aay3166 (adapted)
DOI: 10.1186/s12302-020-00314-9

MassBank – library coverage

Record Index

Home | Search | Record Index | Data Privacy | Imprint | MassBank I | Go

Contributor :	AAFC (950)	Env_Anal_Chem_U_Tuebingen (128)	Kazusa (273)	Nihon_Univ (706)
	Athens_Univ (5252)	Fac_Eng_Univ_Tokyo (12379)	Keio_Univ (4780)	Osaka_MCHRI (20)
	BGC_Munich (903)	Fiocruz (898)	Kyoto_Univ (184)	Osaka_Univ (449)
	BS (1318)	Fukuyama_Univ (340)	Literature_Specs (39)	PFOS_research_group (41
	Boise_State_Univ (4)	GL_Sciences_Inc (174)	MPI_for_Chemical_Ecology (691)	RIKEN (11935)
	CASMI_2012 (26)	HBM4EU (1925)	MSSJ (328)	RIKEN_IMS (1140)
	CASMI_2016 (622)	IPB_Halle (677)	MetaboLights (58)	RIKEN_NPDepo (1956)
	Chubu_Univ (2563)	ISAS_Dortmund (339)	Metabolon (149)	RIKEN_ReSpect (4642)
	Eawag (11191)	JEOL_Ltd (44)	NAIST (621)	Tottori_Univ (16)
	Eawag_Additional_Specs (895)	KWR (207)	NaToxAq (3756)	UFZ (1242)
Instrument Type :	APCI-ITFT (69)	ESI-ITTOF (4)	FI-B (1)	LC-APPI-QQ (287)
	APCI-ITTOF (1)	ESI-QIT (1)	GC-APCI-QTOF (94)	LC-ESI-IT (664)
	APCI-Q (3)	ESI-QTOF (252)	GC-EI-Q (45)	LC-ESI-ITFT (16884)
	CE-ESI-TOF (20)	ESI-TOF (128)	GC-EI-QQ (19)	LC-ESI-ITTOF (262)
	CI-B (796)	FAB-B (26)	GC-EI-TOF (954)	LC-ESI-Q (2738)
	CI-Q (8)	FAB-BE (15)	GC-FI-TOF (6)	LC-ESI-QFT (7136)
	EI-B (11810)	FAB-EB (5)	LC-APCI-ITFT (50)	LC-ESI-QIT (413)
	EI-EBEB (12)	FAB-EBEB (172)	LC-APCI-Q (12)	LC-ESI-QQ (8867)
	ESI-ITFT (62)	FD-B (41)	LC-APCI-QTOF (633)	LC-ESI-QQQ (1956)
MS Type :	MS (19982)	MS2 (59680)	MS3 (929)	MS4 (70)
Ion Mode :	NEGATIVE (23266)	POSITIVE (57395)		

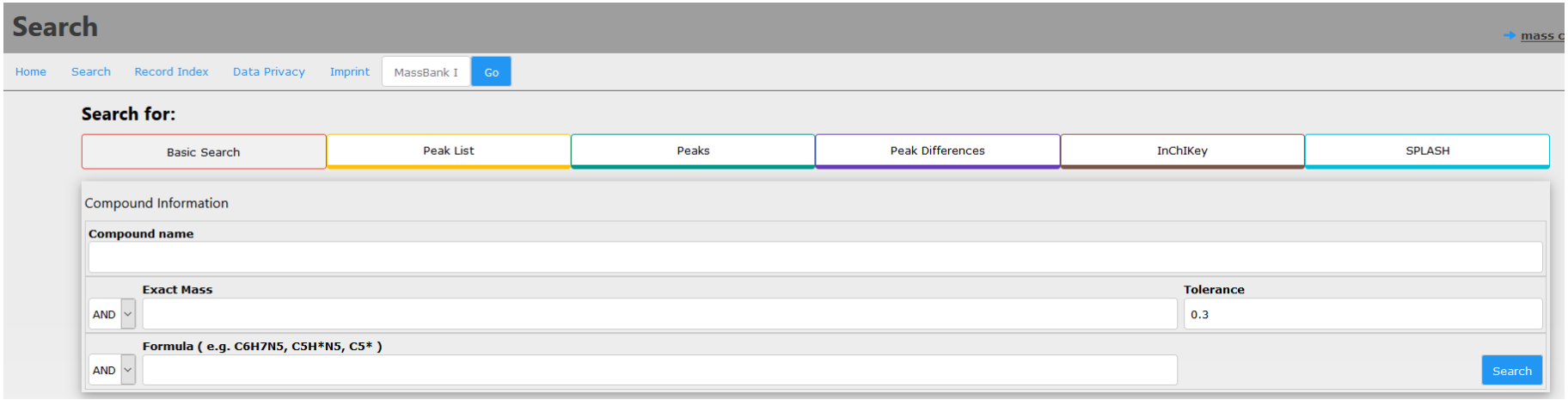
80661 unique spectra (MS2, MS1, some MSn; ESI, EI, APCI…)

14382 unique compounds (anthropogenic, metabolites, TPs)

Accessed 29 May 2020

https://massbank.eu/MassBank/RecordIndex

MassBank - search



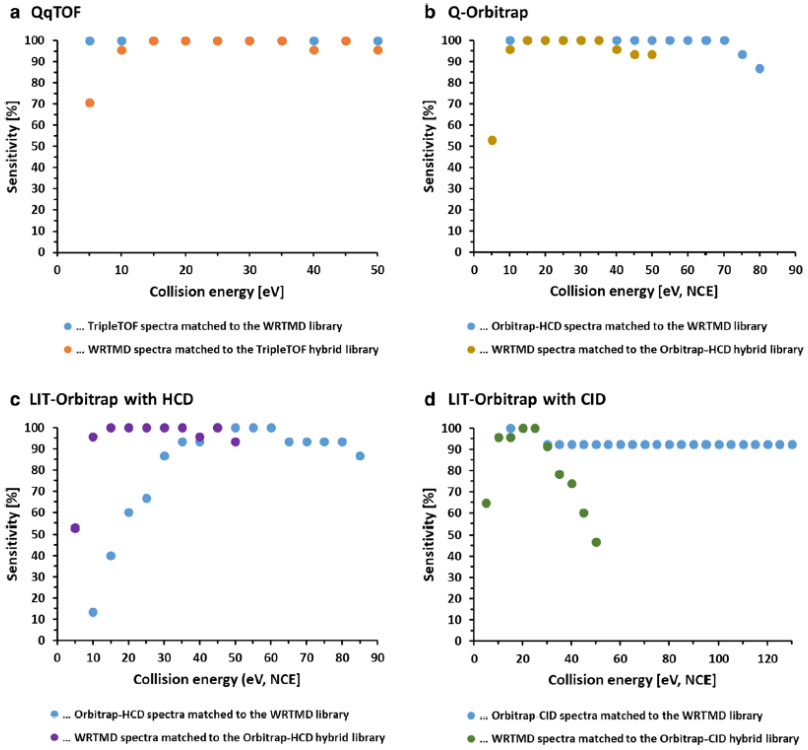
Similarity search:

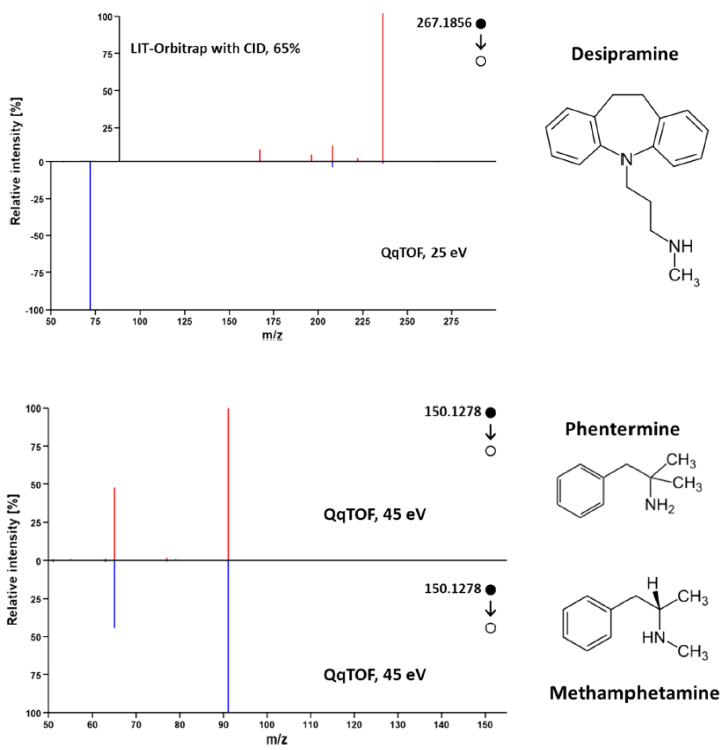
$$W_i = [\text{Intensity of peak}_i]^m [m/z \text{ of peak}_i]^n$$

DOI: 10.1002/jms.1777
DOI: 10.1016/1044-0305(94)87009-8
DOI: 10.1038/nbt.3689

Integration in tools such as MetFrag, MASST, MZmine, NIST MS Search possible by download from MoNA

Reliability of library searches

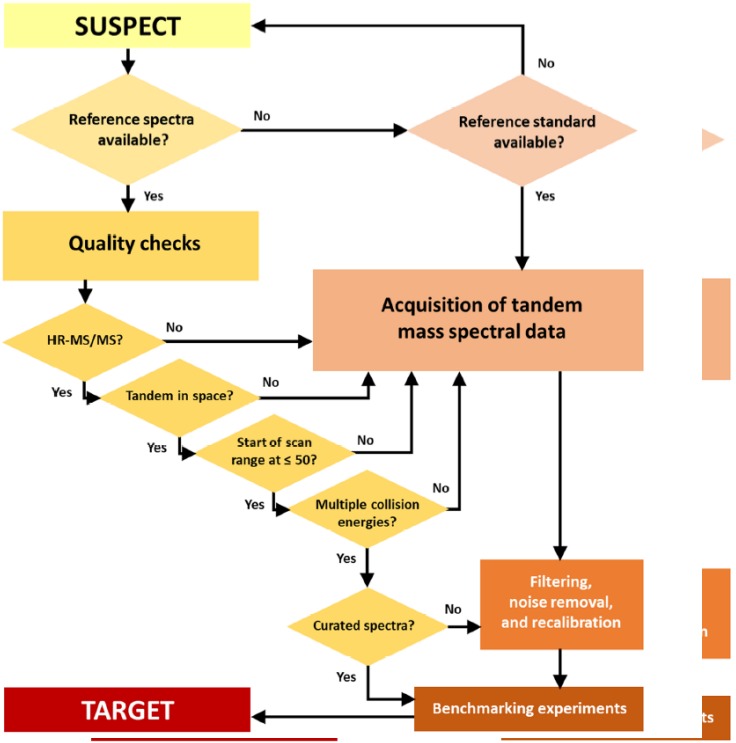




DOI: 10.1186/s12302-020-00314-9

Smart strategy to grow community libraries

- Community MS libraries cannot grow as fast as commercial databases
- Smart collections of suspect lists (priority compounds) could help to focus on relevant compounds for different communities
- Shared MS acquisition will help to populate community MS libraries with relevant content more quickly



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

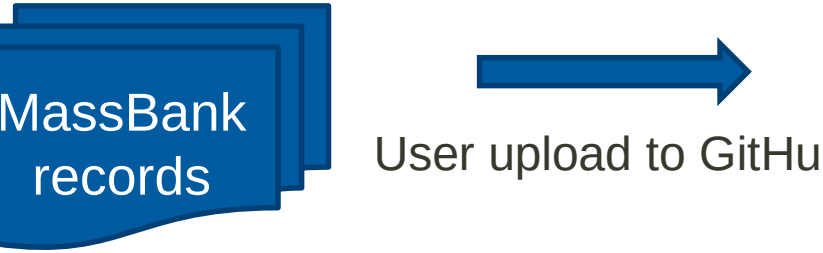
DOI: 10.1186/s12302-020-00314-9

MassBank – data community

GitHub

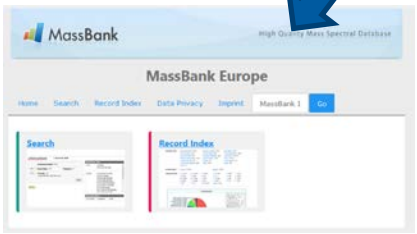
MassBank records

User upload to GitHub



Share with friends

Automated deployment in MassBank



Automated validation and curation with Travis CI

https://www.travis-ci.org

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

MassBank – going FAIR

Full machine readable representation of each record and open access integration


```
"identifier": "AU248503",
"url": "https://massbank.eu/RecordDisplay?id=AU248503",
"name": "(Methoxymethyl)triphenylphosphonium",
"alternateName": ["(Methoxymethyl)triphenylphosphonium", "methoxymethyl(triphenyl)phosphanium"],
"molecularFormula": "[C20H20OP]+",
"monoisotopicMolecularWeight": "307.125176899999985380418365821242332458496093 75",
"inChI": "InChI=1S/C20H20OP/c1-21-17-22(18-11-5-2-6-12-18,19-13-7-3-8-14-19)20-15-9-4-10-16-20/h2-16H,17H2,1H3/q+1",
"smiles": "COC[P+](C1=CC=CC=C1)(C1=CC=CC=C1)C1=CC=CC=C1",
"@context": "http://schema.org",
"@type": "MolecularEntity"
},
{
"identifier": "AU248503",
"url": "https://massbank.eu/RecordDisplay?id=AU248503",
"headline": "(Methoxymethyl)triphenylphosphonium; LC-ESI-QTOF; MS2; CE: 30 eV; R=35000; [M+H]+",
"name": "(Methoxymethyl)triphenylphosphonium",
"datePublished": "2019-04-08",
"license": "https://creativecommons.org/licenses",
"citation": "null",
"comment": ["CONFIDENCE standard compound", "INTERNAL_ID 2485"],
"alternateName": ["(Methoxymethyl)triphenylphosphonium", "methoxymethyl(triphenyl)phosphanium"],
"@context": "http://schema.org",
"@type": "Dataset"
```


Wikidata:Property proposal/MassBank Accession ID

< Wikidata:Property proposal

MassBank Accession ID [edit]

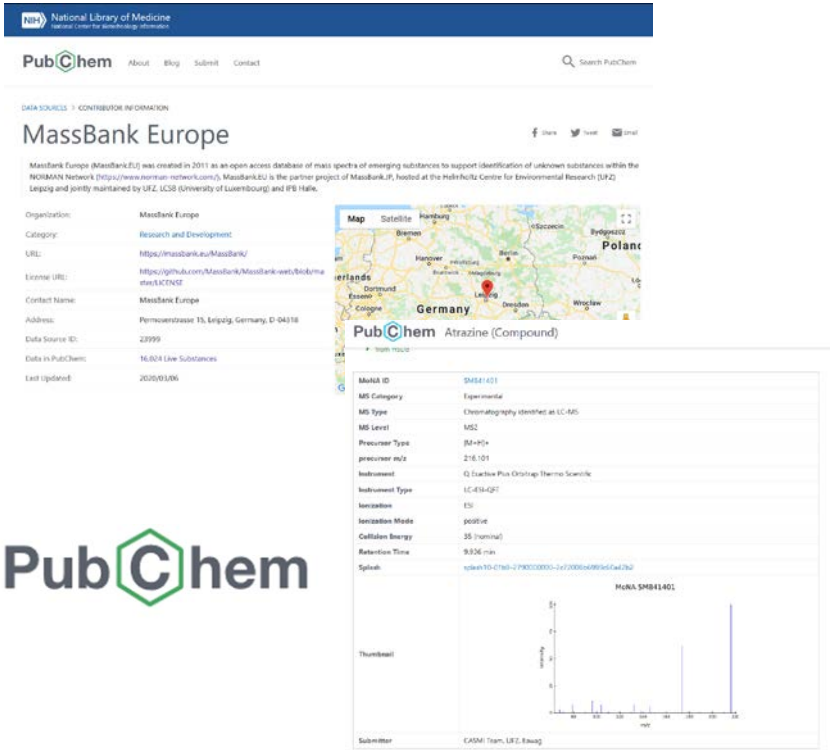
Originally proposed at Wikidata:Property proposal/Natural science

Done: MassBank Accession ID (P6689) (Talk and documentation)

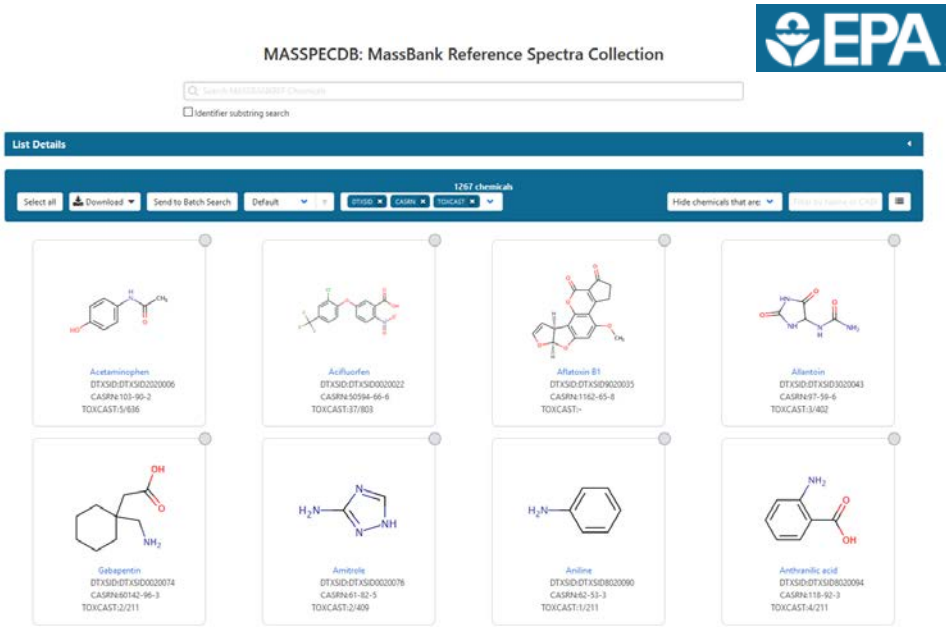
Description	Accession number for entries in the MassBank database (records of mass spectrometry).
Represents	MassBank (Q24088019)
Data type	External identifier
Domain	chemical compound (Q11173)
Allowed values	[A-Z]{2}[A-Z0-9][0-9]{5}
Example 1	caffeine (Q60235) → EA030311
Example 2	Kanamycin A (Q27094615) → SMI00011
Example 3	acetaminophen (Q57055) → AU112601
Source	https://github.com/MassBank/MassBank-web/blob/master/Documentation/MassBankRecordFormat.md#2.1.1
Planned use	Link Wikidata the MassBank (Q24088019) database entries.
Number of IDs in source	53390
Expected completeness	always incomplete (Q21873886)
Formatter URL	https://massbank.eu/MassBank/RecordDisplay.jsp?id=**$1**
Robot and gadget jobs	Accession IDs will be added from a public CCZero dataset (to be compiled) with QuickStatements

https://bioschemas.org
https://www.assembleplus.eu
https://elixir-europe.org/about-us/who-we-are/nodes/germany
https://www.wikidata.org/wiki/Wikidata:Property_proposal/MassBank_Accession_ID

MassBank – compound database integration



PubChem



https://pubchem.ncbi.nlm.nih.gov/source/23999
https://comptox.epa.gov/dashboard/chemical_lists/MASSBANKREF

MassBank – Take home messages

- Open data mass spectral libraries are a great benefit for science
- Creating open data libraries is a big community effort, but they are growing
- Sharing smartly collated suspect lists will help to populate community mass spectral libraries with relevant content for users (in progress)
- Open data and strict implementation of FAIR principles guarantees full interopteratibiliy with external mass spectral libraries, compound databases and tools (in progress)

Acknowledgements

- Hendrik Treudler (IPB)
- Juliane Hollender, Michael Stravs, Heinz Singer (Eawag)
- Werner Brack, Martin Krauss, Erik Müller, Marc Stöhr, and the IT team (UFZ)
- Takaaki Nishioka (MassBank Japan)
- Antony Williams (US EPA)
- Evan Bolton, Jeff Zhang, Paul Thiessen (NCBI/NLM/NIH)
- Egon Willighagen (Maastricht University)
- Sajjan Singh Mehta, Gert Wohlgemuth, Oliver Fiehn (UC Davis)
- Ming Wang, Pieter Dorrestein (UC San Diego)
- David Wishart (University of Alberta)
- Xiaoyu Yang, Steven Stein (NIST)
- MassBank community

Contact:
tobias.schulze@ufz.de
emma.schymanski@uni.lu
sneumann@ipb-halle.de

@tsufz

DOI: 10.5281/zenodo.3865489