GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Optimierung und Sicherung der Datenqualität von STOFF-IDENT

V. Gronau[1], Wielenbach/D, M. Luthardt[3], Weihenstephan/D, T. Placht[3], Weihenstephan/D, A. Gilg[3], Weihenstephan/D, F. Leßke[3], Weihenstephan/D, M. Sengl[2], Augsburg/D, M. Letzel[1], Wielenbach/D

[1]Bayerisches Landesamt für Umwelt, Demollstraße 31, 82407 Wielenbach

[2]Bayerisches Landesamt für Umwelt, Bürgermeister-Ulrich-Str. 160, 86179 Augsburg

[3]Hochschule Weihenstephan-Triesdorf, Vöttinger Str. 27, 85354 Freising

Bayerisches Landesamt für Umwelt, Demollstraße 31, 82407 Wielenbach, veronika.gronau@lfu.bayern.de

FOR IDENT

STOFF-IDENT- Datenbank

- Hilfsmittel zur Identifizierung bislang unbekannter gewässerrelevanter Stoffe mithilfe der Non- und Suspected-Target Analytik
- Integration der Stoffdaten aus Stoffvollzügen (REACH, Biozid-Richtlinie…)
- Einbindung von STOFF-IDENT in die Arbeitsplattform FOR-IDENT zur Verknüpfung mit anderen Recherche-Tools (MetFrag, MassBank etc.)

Datenqualität und Datenumfang

- Integration weiterer gewässerrelevanter Stoffe in die Datenbank sowie die Aktualisierung bereits vorhandener Stoffgruppen (neu registrierte Substanzen aus Stoffvollzügen)
- Kontinuierliche Prüfung und Optimierung der Datenqualität
- Fehlersuche und –korrektur (seit Projektbeginn Umwandlung von >1.500 Fehlern und falschen Zusatzinformationen in Datensätzen)
- Konzept eines effizienten Datenmanagementsystems

Datenumfang *

Stoffvollzüge
∑ 11.346 Substanzen

- **REACH** (8118 Substanzen)
- **Pharmazeutika** (2042 Substanzen, UBA)
- **Biozide** (877 Substanzen, EU-Richtlinie)
- **PSM** (309 Substanzen, BVL)

Laborlisten/ Firmenlisten
∑ 5.288 Substanzen

- **Forensik Innsbruck** (1208 Substanzen)
- **Zweckverband Landeswasserversorgung Langenau** (2845 Substanzen)
- **Bundesanstalt für Gewässerkunde BfG** (926 Substanzen)
- **NIVA/Waters** (1670 Substanzen)

Aktueller Datenbestand:
9.500 Substanzen **

Transformations-produkte
∑ 180 Substanzen

- **Li et al. 2014** (14 Substanzen)
- **Reemtsma et al. 2013** (106 Substanzen)
- **Sonstige** (60 Substanzen)

Sonstige Literatur
∑ 2.873 Substanzen

- **Sonstige Literatur** (2873 Substanzen)

* Die Gesamtliste aller STOFF-IDENT-Einträge ist auf der NORMAN-Homepage zum Download abrufbar :
http://www.norman-network.com/?q=node/236

** Aufgrund von Überschneidungen liegt der Datenbestand trotz 16.459 eingelesener Substanzen aktuell bei 9.500 gewässerrelevanten Stoffen

Prüfregeln zur Sicherung der Datenqualität bei neuen Daten

- Substanz muss CAS-Nummer oder SMILES-Code haben
- Richtigkeit der CAS-Nummer wird über deren Prüfziffer geprüft
- SMILES und Summenformel dürfen keinen Punkt und kein * enthalten
- + im SMILES führt zu gesonderter Überprüfung der Substanzen
- -20 < logP < 20

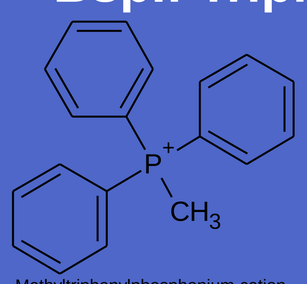
Automatisierte Fehlererkennung im **SI-Crawler**, dem Daten-Einlesetool

Completen...	Suspicious	Name	Source	CAS	Source	SMILES
		Aciclovir	UBA	59277-89-3	UBA	Nc1nc(=O)c2ncn(COCCO)c2[nH]1
		ACIPIMOX	UBA	51037-30-0	Wikipedia	Cc1cncc(C(=O)O)[n+]1[O-]
		ACITRETIN	UBA	55079-83-9	Wikipedia	COc1cc(C)c(\C=C\C(\C)=C\C=C\C(\C
		ACLARUBICIN	UBA	57576-44-0	Wikipedia, chem	CC[C@@]1(O)C[C@H](OC2CC(C(OC3
		Acriflavinium Chloride	UBA	8063-24-9	chemicalbook.cc	[Cl-].Nc1ccc2cc3ccc(N)cc3nc2c1.Cc1c
		ACTINOQUINOL	UBA	15301-40-3	http://www.drug	CCOc1ccc(c2ccccn12)S(O)(=O)=O
		ADAPALENE	UBA	106685-40-9	Wikipedia; chem	COc1ccc(cc1C12CC3CC(CC(C3)C1)C2)
		Adefovirdipivoxil	UBA	142340-99-6	UBA	CC(C)(C)C(=O)OCOP(=O)(COCCn1cn
		Ademetionine disulfate tosylate	UBA	97540-22-2	chemicalbook.cc	OS[[O-]](=O)=O.[O-]S([O-])(=O)=O.C
		ADRENALONE	UBA	99-45-6	chemicalbook.cc	CNCC(=O)c1ccc(O)c(O)c1
		Agomelatin	UBA	138112-76-2	UBA	COc1ccc2cccc(CCNC(C)=O)c2c1
		AJMALICINE	UBA	483-04-5	chemicalbook.cc	COC(=O)C1=C(O[C@@H](C)[C@H]2Cl
		AJMALINE	UBA	4360-12-07	chemicalbook.cc	
		ALATROFLOXACIN	UBA	157605-25-9	wikipedia, chemi	CS(O)(=O)=O.C[C@H](N)C(=O)N[C@
		ALBENDAZOLE	UBA	54965-21-8	Wikipedia	CCCSc1ccc2nc(NC(=O)OC)[nH]c2c1
		ALCLOMETASONE	UBA	66734-13-2	Wikipedia	
		ALCLOXA	UBA	1317-25-5	chemicalbook.cc	

Kategorisierung und Taggen

- Jede in der Datenbank enthaltene Substanz ist einer oder mehreren Kategorien zugeteilt (REACH, Pharmaceuticals, Biocides, PSM, TP‘s)
- Taggen bedeutet, dass eine Substanz mehreren Ursprungslisten zugewiesen werden kann, z.B. zur Spezialsuche in einzelnen Listen: REACH, einzelne Laborlisten, Liste positiv geladener Substanzen, …

Bsp.: Triphenylphosphonium mit M suchen



Methyltriphenylphosphonium cation ($Me\text{-}Ph_3P^+$)

Suche über M [+H] → Kein Ergebnis bei Suche im positiven Ionenmodus, da Masse um +H falsch liegt

Suche über M [±0] → Suche erfolgreich im Ionenmodus [±0] , Substanz wurde gefunden, da positive Grundladung

neu → Automatisierte Gruppensuche aller Substanzen mit positiver Grundladung über M [±0] möglich

Fazit und Ausblick

- Stetig steigende Userzahlen (dadurch weitere Fehlerelimination)
- Verstärkter Fokus auf Transformationsprodukte
- Optimierung der automatischen Fehlererkennung
- Kontinuierliche Erhöhung des Datenbestandes durch Integration weiterer Laborlisten (national und international)

Bayerisches Landesamt für Umwelt · Hochschule Weihenstephan-Triesdorf University of Applied Sciences · Technische Universität München · Zweckverband Landeswasserversorgung · Berliner Wasserbetriebe · Powered by ChemAxon