



STOFF-IDENT database – Contents & Data Quality

V. Gronau¹, Wielenbach/D, M. Luthardt³, Weihenstephan/D, T. Placht³, Weihenstephan/D, A. Gilg³,
Weihenstephan/D, F. Leßke³, Weihenstephan/D, M. Sengl², Augsburg/D, M. Letzel¹, Wielenbach/D

¹Bavarian Environment Agency, Demollstraße 31, 82407 Wielenbach

²Bavarian Environment Agency, Bürgermeister-Ulrich-Str. 160, 86179 Augsburg

³University of Applied Sciences Weihenstephan-Triesdorf, Vöttinger Str. 27, 85354 Freising

Bavarian Environment Agency, Demollstraße 31, 82407 Wielenbach, veronika.gronau@lfu.bayern.de

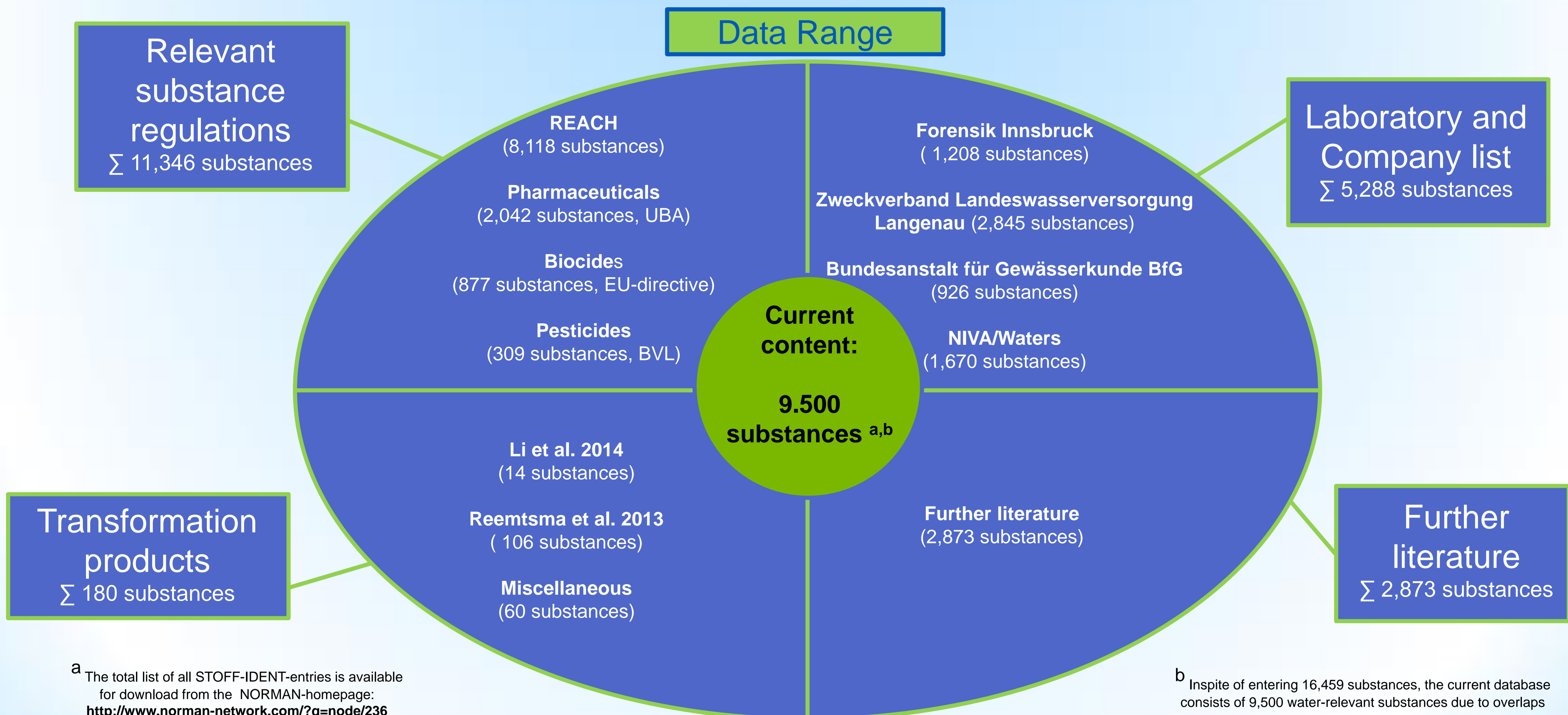


STOFF-IDENT Database

- Tool for identifying previously unknown water-relevant substances using the Non Target and Suspected Target Analytics
- Integration of substance data from respective directives/regulations (REACH, European Biocide Directive 98/8/EC, etc.)
- Integration of STOFF-IDENT into the working platform called FOR-IDENT and with other research tools (MetFrag, MassBank, etc.)
- Link to FOR-IDENT: <http://for-ident.hswt.de>

Scope and Data quality

- Integration of additional water-relevant substances in the database and updating existing substance groups (newly registered substances from regulations)
- Continuous testing and optimization of the data quality
- Error identification and correction (during the project more than 1,500 errors and false information have been identified and corrected)
- An efficient database management system was conceptualized



Rules to ensure data quality with new data

- Substance must have a CAS-number or a SMILES-Code
- Verifying the accuracy of the CAS-number on the digit
- SMILES and formula should not contain a dot or *
- + in the SMILES leads to a separate review of the substances
- 20 < logP < 20

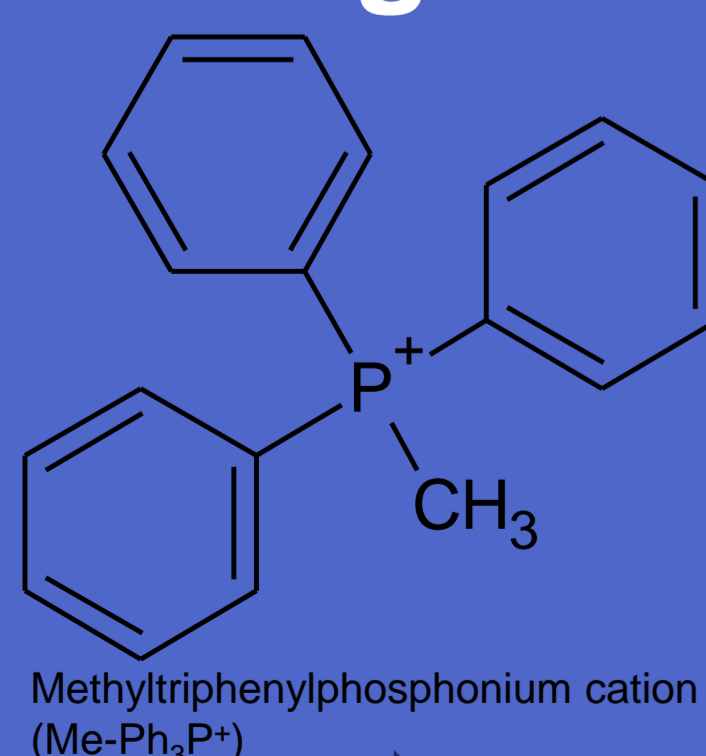
Automated error detection in the
SI-Crawler, the data tool for reading

Completen...	Suspicious	Name	Source	CAS	Source	SMILES
●	●	Acidover	UBA	99277-99-3	UBA	Ndnc(cO)zncm(COCCO)z[H]Hl
●	●	ACIPROX	UBA	51037-20-0	Wikipedia	Cd.nccc(O)O(=O)N=O
●	●	ACIFRETIN	UBA	55979-99-9	Wikipedia	COc1ccc(Cc1C(C)C)C(C)C(C)C(C)C
●	●	ACLARUBICIN	UBA	57576-44-0	Wikipedia, chem	CC(C)C(=O)C(C)C(=O)C(C)C(C)C(C)C
●	●	Acridinium Chloride	UBA	8063-24-9	chemicalbook.cz	[Cl-].[N+]1=CC=CC=C1
●	●	ACTINOQUINOL	UBA	15301-40-3	http://www.drug	CCO1ccc(cc1C2=CC=CC=C2)O1=O
●	●	ADAPALENE	UBA	106685-40-9	Wikipedia, chem	CC1ccc(cc1C2=CC=CC=C2)C(C)C(C)C2
●	●	Adelovirdipivoxil	UBA	142340-99-6	UBA	CC(C)C(C)C(=O)C(C)C(=O)C(C)C(C)C
●	●	Ademetionine disulfate tosylate	UBA	97540-22-2	chemicalbook.cz	OS(=O)(=O)C(=O)C(C)C(=O)C(C)C(C)C
●	●	ADRENALONE	UBA	99-45-6	chemicalbook.cz	CNCC1=O(C)C(C)C1=O
●	●	Agomelatin	UBA	138112-76-2	UBA	COc1ccc(cc1C2=CC=CC=C2)C(C)C(C)C2
●	●	AMALICINE	UBA	483-04-5	chemicalbook.cz	COc1ccc(cc1C2=CC=CC=C2)C(C)C(C)C2
●	●	AMALINE	UBA	486-12-07	chemicalbook.cz	COc1ccc(cc1C2=CC=CC=C2)C(C)C(C)C2
●	●	ALATROFLOXACIN	UBA	157605-25-9	wikipedia, chemi	CS(O)(=O)O.C(C)C(=O)N(C)C(=O)N(C)C
●	●	ALBENDAZOLE	UBA	54865-21-8	Wikipedia	CCc1ccc(cc1N(C)C)C(C)C(C)C(C)C
●	●	ALCLOMETASONE	UBA	66734-13-2	Wikipedia	CC(C)C(C)C(=O)C(C)C(=O)C(C)C(C)C
●	●	ALCLOXA	UBA	1337-25-5	chemicalbook.cz	CC(C)C(C)C(=O)C(C)C(=O)C(C)C(C)C

Categorization und Tagging

- Each substance contained in the database is assigned to one or more categories (REACH, Pharmaceuticals, Biocides, PSM, TP's)
- Tagging means that a substance can be assigned to further source lists, for example special searches in individual lists such as REACH, laboratory lists, list of positively charged substances, ...

E.g.: Triphenylphosphonium Search with M



Search with M [+H]

No result in positive ion mode because the mass is wrong by +H

Search with M [±0]

The search is successful in ion mode [±0]. Substance was found due to positive basic charge.

Automated group search of all substances with positive basic charge on M [±0] is possible

Conclusion and Outlook

- Steadily increasing user numbers, thereby further error elimination (currently 175 users)
- Increased focus on transformation products
- Optimization of automatic error detection
- Continuous increase of the data range by integrating additional laboratory lists (national and international)