International Suspect Screening: NORMAN Suspect Exchange meets the US EPA CompTox Chemistry Dashboard

EMMA SCHYMANSKI[1], ANDREW MCEACHRAN[2], REZA AALIZADEH[3], NIKOLAOS S. THOMAIDIS[3], JAROSLAV SLOBODNIK[4], JON SOBUS[2], JULIANE HOLLENDER[1] AND ANTONY J. WILLIAMS[2].

[1]Eawag: Swiss Federal Institute for Aquatic Science and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland, emma.schymanski@eawag.ch
[2]Oak Ridge Institute for Science and Education Research Participation Program (AM), National Exposure Research Laboratory (JS), National Centre for Computational Toxicity (AJW), Office of Research and Development, U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Research Triangle Park, North Carolina, USA 27711.
[3] National and Kapodistrian University of Athens, Department of Chemistry, Panepistimiopolis Zografou, 157 71 Athens, Greece.
[4]Environmental Institute, s.r.o., Okružná 784/42, 972 41 Koš, Slovak Republic.

Members of the European NORMAN Network of Environmental Laboratories (www.norman-network.com) have many substance lists, including targets, suspects, surfactants, perfluorinated substances and regulated, partially confidential data sets of complex mixtures. The NORMAN Suspect List Exchange (http://www.norman-network.com/?q=node/236) was founded in 2015 following the NORMAN Collaborative Non-target Screening Trial [1] to improve non-targeted screening results via increase the exchange of suspect lists. There are currently 16,473 entries in 13 lists from 9 contributors, with more lists coming. A curation workflow was used to create a single, merged list of 11,922 substances, "NORMAN SusDat" (http://www.norman-network.com/datatable/). Merging datasets of mixed origin and especially those without exact structural information is challenging, yet these lists contain valuable information for suspect screening with mass spectrometry.

The U.S. Environmental Protection Agency Computational Toxicology Program integrates advances in biology, chemistry, and computer science to help prioritize chemicals based on potential human health risks. Their web application, the CompTox Chemistry Dashboard (http://comptox.epa.gov) provides access to data associated with ~740,000 chemical substances, compiled with years of investment in data validation and curation. These data include experimental and predicted physicochemical property data, bioassay screening data associated with ToxCast, product and functional use information and more. Flexible search capabilities using e.g. mass and formula-based searching utilizes rank-ordering of results via functional use statistics to help prioritize detected chemicals for further review, supporting screening studies with significant improvements beyond ChemSpider [2].

The collaboration between NORMAN and the EPA, while in its infancy, has yielded rapid results. Much of the data associated with the Dashboard are freely available (https://comptox.epa.gov/dashboard/downloads) for use (e.g. in the Collaborative Non-target Screening Trial currently run by the EPA) or for integration into other applications (e.g. the identification software MetFrag [3]). Challenges that these partners are actively tackling include how to integrate chemical substances of unknown or variable composition, complex reaction products and biological materials (UVCBs) into their resources.

This abstract does not reflect U.S. EPA policy.

[1] E. Schymanski *et al*. 2015, *Anal Bioanal Chem*, 407: 6237.
[2] A.D. McEachran *et al*. 2017, *Anal Bioanal Chem*, 409: 1729.
[3] C. Ruttkies *et al* 2016, *J Cheminform*, 8 (1): 3.