Session: CINF: Enhance Discovery: Sharing Chemical Structures

Curating and sharing structures and spectra for the environmental community

Emma L. Schymanski[1]* and Antony J. Williams[2]

[1]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, Luxembourg.
[2]National Center for Computational Toxicology, US EPA, Research Triangle Park, Durham, NC, USA.

* emma.schymanski@uni.lu

The increasing popularity of high mass accuracy non-target mass spectrometry methods has yielded extensive identification efforts based on spectral and chemical compound databases in the environmental community and beyond. Increasingly, new methods are relying on open data resources. Candidate structures are often retrieved with either exact mass or molecular formula from large resources such as PubChem, ChemSpider or the EPA CompTox Chemistry Dashboard. Smaller, selective lists of chemicals (also called "suspect lists") can be used to perform more efficient annotation. Mass spectral libraries can then be used to increase the confidence in tentative identification. Additional metadata (e.g. exposure and hazard information, reference and data source information) can be extremely useful to help identify substances of high environmental interest. Exchanging information and "sharing structural linkages" between these resources requires extensive curation to ensure that the information is shared correctly, yet many valuable datasets arise from scientists and regulators with little official cheminformatics training. This talk will cover curation efforts undertaken to map spectral libraries (e.g. MassBank.EU, mzCloud) and suspect lists from the NORMAN Suspect Exchange (http://www.norman-network.com/?q=node/236) to unique chemical identifiers associated with the US EPA CompTox Chemistry Dashboard. The curation workflow takes advantage of years of experience, as well as contact with the original data providers, to enable open access to valuable, curated datasets to support the environmental community and scientists beyond (e.g. https://comptox.epa.gov/dashboard/chemical_lists). This work enables sharing high quality open data with the community for reuse and repurposing. *Note: This abstract does not reflect US EPA policy*.