Proposals for NORMAN Joint Programme of Activities 2022

Title	Update and improvement of the NORMAN Substance Database
Type of activity	Database development
Leader	Jaroslav Slobodnik (EI-Environmental Institute) and Nikolaos S. Thomaidis (NKUA-National and Kapodistrian University of Athens)
Topic / activities	***Background / Justification for the proposed activity:*** Numerous activities of the network have been benefited since the introduction of NORMAN Substance Database (SusDat; https://www.norman-network.com/nds/susdat/) in 2016. It is an internationally recognised compound database of environmentally-relevant contaminants of emerging concern (CECs). As of December 2022, it contains 106,660 compounds accompanied with essential information on their unique chemical identifiers, chemical curation level, adduct forms, chemical properties, retention time index (RTI), *in-silico* predicted no-effect concentrations (P-PNECs), use category information, etc. SusDat is at the core of many other database modules in the NORMAN Database System (NDS): EMPODAT, Ecotoxicology Database, Substance Factsheets Passive Sampling Database and Digital Sample Freezing Platform (DSFP). SusDat represents the chemical space that is of interest to the network. SusDat is the most frequently visited module of the NDS. A continuous update, maintenance, curation and improvement of quality of information in SusDat is of importance for the NORMAN network. This proposal aims to continue with (i) upload of individual substance lists contained in the Suspect List Exchange (SLE), (ii) their merging and chemical curation, and (iii) adding missing and new information. ***Description of the proposed activity and expected outcomes for 2023 (and beyond):*** **Task 1**: Derivation of predicted physico-chemical properties (using VEGA platform and other tools) and addition of the generated information in the database **Task 2**: Improvement of classification of compounds in use categories (connection with PARC and PubChem cooperation) **Task 3**: Addition of *in-silico* predicted RTI values for electrospray ionization (both positive and negative) and P-PNECs for new substances in SusDat. Update of RTI and P-PNECs for substances with an updated chemical structure. **Task 4**: Application of the automated curation workflow developed in previous years to update the *validation level* for all substances. Adding missing *validation level* information for newly added substances. Addition of any missing information for existing compounds. **Task 5**: Addition of preferable analytical methods for each compound (i.e. GC-EI, LC-(+/-)ESI etc.). **Taks 6**: Investigation of potential deployment of the automated curation workflow. **Task 7**: Addition of toxicity threshold data for terrestrial environment and human health. **Task 8:** Addition of consensus models for logBCF, $logK_{oa}$, $logK_{oc}$ and $logK_{ow}$ and ionization efficiency (logIE) values. ***Added value / Link with other NORMAN activities and / or other projects*** - Integration in Cross-Action Working Group CTS. - Integration in NORMAN Digital Sample Freezing Platform. - Interlink with other working groups within NORMAN - Interlink with WG1 Prioritisation.
Participants	Any interested parties
Proposed in-kind contribution	EI: Tasks 1, 2, 6 and integration of all new data/information in SusDat UBA: Evaluation of the quality of toxicological data NKUA: Tasks 3, 4, 5, 7, 8
Contribution needed from NORMAN Association	Task 7, 8: NKUA 4,000 € -Addition of properties for newly added compounds Task 6: EI 8,000 € - Programming of the interlink between SLE and SusDat - Improvement/programming of the automated curation algorithm for removal of duplicates (names, CAS Nos., InChIKeys) and manual check/curation of the flagged 'problematic' compounds *Total contribution required: 12,000 €*