

Proposals for NORMAN Joint Programme of Activities 2023

Title	NORMAN Database System (NDS)
Type of activity	Database maintenance and continuous update
Leader	EI
Topic / activities	<p>Background / Justification for the proposed activity:</p> <p>The NORMAN Database System (NDS) is a joint activity of all NORMAN members and at the core of the NORMAN activities, providing data and tools to fulfil its goals and visions. The NDS consists nowadays of 13 integrated databases modules:</p> <ol style="list-style-type: none"> 1. Suspect List Exchange - https://www.norman-network.com/nds/SLE/ 2. Substance Database - https://www.norman-network.com/nds/susdat/ 3. Chemical Occurrence Data (EMPODAT) - https://www.norman-network.com/nds/empodat/ 4. Ecotoxicology Database - https://www.norman-network.com/nds/ecotox/ 5. Digital Sample Freezing Platform (DSFP) - https://norman-data.net/Verification/ 6. Substance Factsheets - https://www.norman-network.com/nds/factsheets/ 7. NORMAN MassBank - https://massbank.eu/MassBank/ 8. Passive Sampling - https://www.norman-network.com/nds/passive/ 9. Antibiotic Resistance Bacteria/Genes - https://www.norman-network.com/nds/bacteria/ 10. SARS-CoV-2 in sewage - https://www.norman-network.com/nds/sars_cov_2/ 11. Bioassays Monitoring Data - https://www.norman-network.com/nds/bioassays/ 12. Indoor Environment - https://www.norman-network.com/nds/indoor/ 13. Prioritisation - https://www.norman-network.com/nds/prioritisation/ <p>A new module EMPODAT-SUSPECT (https://www.norman-network.com/nds/suspect/) has been developed in 2021 in support of the WG1 Prioritisation. Some first testing has been performed in 2022 with more than 26 million data entries. Attempts to speed up the data search queries are on-going. Further work will be done in 2023 in collaboration with WG-1 to improve the current prototype.</p> <p>A new module BIOACTIVITY DATABASE has been developed in 2022. This prototype needs to be revised and optimised according to the practice of datasets that are added. The same goes for the user interface of the BioActivity Database. This work will be carried out in 2023 in collaboration with the KWR and the team involved in the "Bioactivity Factsheet".</p> <p>All NDS modules can be searched either individually or starting from the module 'Search All Databases' (https://www.norman-network.com/nds/common/), where a presence of any substance from SusDat in any of the database modules is shown with all existing data.</p> <p>Automated prioritisation module in the NDS have been developed for the target substances archived in EMPODAT. More recently, a prototype of such prioritisation module for suspect substances archived in EMPODAT-SUSPECT have been developed and tested. However, the prototype does not reflect yet the latest algorithm for categorisation of substances and its further upgrade is expected in 2023 following the instructions and guidance from the WG1 (see WG-1 Factsheet).</p> <p>A database fully compatible with the NDS and a parallel automated prioritisation module for top predators and their prey has been developed within the EU LIFE APEX project (https://lifeapex.eu/). Additional data have been generated within OSPAR CONNECT and HELCOM PreEMP and UBA-HELCOM projects. Altogether ca. 1.1 million data entries on more than 3250 substances in biota samples started to be integrated with the NDS in 2022; the process will be finalised in 2023.</p> <p>Each of the NDS modules is being continuously updated, either in terms of the required structure of the data (Data Collection Templates – DCTs downloadable at each of the modules) or data. The EMPODAT database grew from ca. 20 million to ca. 38 million data entries in 2022 while additional ca. 45 million data are being curated. A capacity of SusDat to support suspect screening has been increased from ca. 65,000 substances to ca. 95,000 substances in each sample stored in DSFP (>3,200). Predicted ecotoxicological threshold values in freshwater (fw), marine water (mw), biota (fw/mw) and sediments (fw/mw) are now available for ca. 93,000 compounds. The SusDat contains information on applicability domain of the stored substances, indicating if the compound can be analysed by GC or LC, and if so, in the EI, positive or negative ionisation mode. The Substance Factsheet database is updated each three months to reflect latest changes in the US EPA CompTox Chemicals Dashboard database. A process to visualise the data in individual NDS modules is on-going.</p> <p>A procedure of harmonising quality checks of the target analyses data between EMPODAT and IPCHEM has been agreed at the meeting in September 2022. The data stored in EMPODAT are being exchanged with IPCHEM on an annual basis.</p> <p>Description of the proposed activity and expected outcomes for 2023 (and beyond):</p> <p>The maintenance of the NDS and its continuous upgrade for new data require a significant effort. The tasks for 2023 include:</p> <ul style="list-style-type: none"> • Continuous upgrade of all NDS modules;

	<ul style="list-style-type: none"> • NDS Chemical Occurrence Data (EMPODAT): maintenance, upgrading and feeding of new data into the database; sharing the data with IPCHEM; • Continuous upgrade of all DCTs for an extended list of NORMAN substances (SusDat), drop-down lists and definitions of obligatory parameters; • Further upgrade of quality control tools for automated quality check of all input data, harmonisation of quality check procedures with IPCHEM; • Further development of data mining tools to extract raw data from external database systems (e.g. database of Norwegian EPA, UBA Germany, ICES, etc.) and establishment of workflows/APIs for their conversion into the 'NORMAN format'; • Further development of the EMPODAT-SUSPECT module (upgrading and testing of the current prototype) • Enhancement of visualization (maps, trends) and data analysis (batch search, statistical evaluations) capabilities of the NDS; • Update of Passive Sampling module, with focus on datasets obtained by dynamic passive sampling technique; design of Passive Sampling – SUSPECT module and upload of the test set of NTS data; in cooperation with the PS-CWG; • Upload of new data into the ARBs/ARGs (WG5), bioassays (WG2 and 3), indoor environment (WG6), and SARS-CoV-2 in sewage (WG5) modules; upload of test water reuse datasets into EMPODAT; • Upgrade of Substance Factsheets module – systematic collection of all data needed for prioritisation and further improvement of data download functions; • Upload of data from wide-scope target screening and suspect screening of chemicals in top predators and their prey carried out in the LIFE APEX project; • Upload of data from OSPAR CONNECT, UBA-HELCOM and PreEMPT projects; • Organisation of the NORMAN Database Workshop. Strategies for further development and technical discussions among IT experts. <p>Added value / Link with other NORMAN activities and / or other projects</p> <p>The proposed tasks will benefit all WGs and CWGAs in the NORMAN network.</p>
Participants	EI, all interested members
Proposed contribution	<p>in-kind</p> <p>All – contribution of existing data EI – overall coordination</p>
Contribution needed from NORMAN Association¹	<p>Maintenance and continuous update of the NDS:</p> <ul style="list-style-type: none"> - EI: 42,000 € <p>Rental of the server hosting the NDS and backup system:</p> <ul style="list-style-type: none"> - EI: 7,200 € <p>NORMAN Database Workshop</p> <ul style="list-style-type: none"> - 3,000 €

¹ Please, provide here a transparent justification of the requested resources and of the in-kind contribution, thereby distinguishing between the costs associated with “person-months” for the organisation, the “travelling costs” for invited speakers and the costs for the logistics (e.g. meals, room rental etc.)