



## Proposals for NORMAN Joint Programme of Activities 2023

<b>Title</b>	<b>NORMAN Database System (NDS)</b>
<b>Type of activity</b>	Database maintenance and continuous update
<b>Leader</b>	EI
<b>Topic / activities</b>	<p><b>Background / Justification for the proposed activity:</b></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 of 13 integrated databases modules:</p> <ol style="list-style-type: none"> <li>1. Suspect List Exchange - <a href="https://www.norman-network.com/nds/SLE/">https://www.norman-network.com/nds/SLE/</a></li> <li>2. Substance Database - <a href="https://www.norman-network.com/nds/susdat/">https://www.norman-network.com/nds/susdat/</a></li> <li>3. Chemical Occurrence Data (EMPODAT) - <a href="https://www.norman-network.com/nds/empodat/">https://www.norman-network.com/nds/empodat/</a></li> <li>4. Ecotoxicology Database - <a href="https://www.norman-network.com/nds/ecotox/">https://www.norman-network.com/nds/ecotox/</a></li> <li>5. Digital Sample Freezing Platform (DSFP) - <a href="https://norman-data.net/Verification/">https://norman-data.net/Verification/</a></li> <li>6. Substance Factsheets - <a href="https://www.norman-network.com/nds/factsheets/">https://www.norman-network.com/nds/factsheets/</a></li> <li>7. NORMAN MassBank - <a href="https://massbank.eu/MassBank/">https://massbank.eu/MassBank/</a></li> <li>8. Passive Sampling - <a href="https://www.norman-network.com/nds/passive/">https://www.norman-network.com/nds/passive/</a></li> <li>9. Antibiotic Resistance Bacteria/Genes - <a href="https://www.norman-network.com/nds/bacteria/">https://www.norman-network.com/nds/bacteria/</a></li> <li>10. SARS-CoV-2 in sewage - <a href="https://www.norman-network.com/nds/sars_cov_2/">https://www.norman-network.com/nds/sars_cov_2/</a></li> <li>11. Bioassays Monitoring Data - <a href="https://www.norman-network.com/nds/bioassays/">https://www.norman-network.com/nds/bioassays/</a></li> <li>12. Indoor Environment - <a href="https://www.norman-network.com/nds/indoor/">https://www.norman-network.com/nds/indoor/</a></li> <li>13. Prioritisation - <a href="https://www.norman-network.com/nds/prioritisation/">https://www.norman-network.com/nds/prioritisation/</a></li> </ol> <p>All NDS modules can be searched either individually or starting from the module 'Search All Databases' (<a href="https://www.norman-network.com/nds/common/">https://www.norman-network.com/nds/common/</a>), where any substance from SusDat can be searched and displayed with all available data for this substance in any of the database modules. Two new modules, EMPODAT-SUSPECT (see JPA 2024 WG1) and BIOACTIVITY Database (see separate JPA 2024), are under development.</p> <p>An automated prioritisation module in the NDS is available for the target substances archived in EMPODAT. A prototype of the prioritisation module for suspect substances archived in EMPODAT-SUSPECT is currently being developed (see JPA 2024 WG1).</p> <p>A dedicated database and a parallel automated prioritisation module for biota samples has been developed within the EU LIFE APEX project (<a href="https://lifeapex.eu/">https://lifeapex.eu/</a>), containing ca. 1.1 million data entries for 3,123 substances as of December 2023. In 2023, all data have been transferred also to the NDS, however, LIFE APEX database is recommended for a more refined search using biota-specific queries.</p> <p>All NDS modules underwent continuous update in 2023. The EMPODAT database grew from ca. 38 million data entries in 2022 to more than 96 million data in the end of 2023; with monitoring data for 4,453 substances. The Data Collection Templates have been upgraded and their format harmonised for surface water, ground water, wastewater, sewage sludge, sediments, biota, soil, antibiotic resistance bacteria/genes (ARB/ARGs) and passive sampling. A significant number of new data were uploaded in the Indoor environment (&gt;177,000) and Passive sampling (&gt;4,200) database. WG5 suggested to transform the SARS-CoV-2 in sewage database in 2024 and include also pathogens and viruses (cf. separate JPA 2024).</p> <p>The SLE contained in the end of 2023 111 lists (cf. separate JPA 2024). The new lists were merged and uploaded into the SusDat, which grew from 106,737 to 120,522 substances. There is a need to collect all supporting data required for suspect screening (cf. separate JPA on DSFP) and prioritisation (cf. separate JPA WG1) of the new 13,785 substances, which represents a challenge for 2024. Thanks to the significant effort of the NDS Development Team, automated curation tools supported with the manual control resulted in removal of 4,209 duplicate entries based on the name or CAS No. of the substances. There are however, 280 duplicates based on the StdINChIKey, which still need attention and appropriate strategy how to deal with them. Inclusion of more UVCB compounds in the SLE, and thus also in the SusDat, is expected in 2024. A new 'batch conversion of identifiers' feature with a guidance has been made available in SusDat (see new icon!), allowing for seamless conversion among names / CAS. Nos./ StdINChIKey/ NORMAN SusDat ID of any (group of) substances in SusDat. Once NORMAN SusDat IDs are generated, they can be directly used for batch searches, e.g., in EMPODAT or in the prioritisation modules.</p> <p>Substance Factsheets are being updated for latest information from interrelated databases in 3 months interval. In 2023, data on PBT, CMR and ED properties were generated for more than 76,000 substances using the JANUS tool (VEGA platform). Their upload, with consideration of broader strategy prepared by WG1, is scheduled for 2024.</p> <p>The data stored in EMPODAT are being exchanged with the EC IPCHEM on an annual basis. A novel way of formatting the data for upload into the IPCHEM has been agreed at the meeting between the NDS Development Team and IPCHEM team in November 2023. The next 'data exchange' is scheduled for January 2024.</p> <p>Based on the JPA 2023, NORMAN Database Workshop took place in Athens on 26 – 27 October 2023.</p>

	<p>The main objective was to discuss strategies for further development of the NDS and to provide a platform for technical discussions among the networks' IT experts. A transition to next generation data science with the use of artificial intelligence, machine learning and advanced statistical and visualisation tools had been discussed. A proposal has been made of (i) a strategy to ensure long-term sustainability and resilience of the NDS and (ii) inclusion of the NDS as part of the European environmental data infrastructure. The discussion was centered towards the following topics:</p> <ul style="list-style-type: none"> <li>• Where are we? What is new?</li> <li>• Data quality and data traceability - novel strategies for data curation</li> <li>• From data to information – towards embedding artificial intelligence and machine learning tools in the NDS</li> <li>• How to become fully FAIR?</li> </ul> <p>For an overview of discussions and major suggestions, see the workshop's report in the internal area of the NORMAN website.</p> <p><b>Description of the proposed activity and expected outcomes for 2024 (and beyond):</b></p> <p>The task for maintenance of the NDS and its continuous upgrade in 2024 include:</p> <ul style="list-style-type: none"> <li>• Critical consideration of the outcomes of the NORMAN Database Workshop. Development of a roadmap/ strategy for further development of the NDS.</li> <li>• Development of API portal allowing for automated data sharing with external databases.</li> <li>• Assigning DOI to all contributed datasets in the NDS.</li> <li>• Further interlinking of all NDS modules and quality check of all input data (it is envisaged that this task could be supported by an appointed QA/QC Expert Group)</li> <li>• NDS Chemical Occurrence Data (EMPODAT): maintenance, upgrading and feeding of new data into the database; sharing the data with IPCHEM;</li> <li>• Cooperation with EEA on gathering data on ARB/ARG;</li> <li>• Development of data mining tools to extract raw data from publications, patents and grey literature and establishment of a workflow for their processing into the 'NORMAN format';</li> <li>• Implementation of visualisation tools in the NDS as agreed in the NORMAN Database Workshop (in collaboration with WG-1);</li> <li>• Continuous upgrade and maintenance of SusDat;</li> <li>• Update of Passive Sampling module with new datasets; programming of Passive Sampling – SUSPECT module;</li> <li>• Upload of new data into the ARBs/ARGs module (JDS4, EMBLAS, The Cruise of Three European Seas, etc.);</li> <li>• Upgrade of Substance Factsheets module – systematic collection of all data needed for prioritisation and data download functions.</li> </ul> <p><b>Added value / Link with other NORMAN activities and / or other projects</b></p> <p>The proposed tasks will benefit all WGs and CWGAs in the NORMAN network.</p>
<b>Participants</b>	EI, all interested members
<b>Proposed in-kind contribution</b>	All – contribution of existing data EI – overall coordination
<b>Contribution needed from NORMAN Association<sup>1</sup></b>	Maintenance and continuous update of the NDS: - EI: 42,000 € Rental of the server hosting the NDS and backup system: - EI: 8,600 € DOI for the datasets contributed to the NDS: - EI: 1500 €

<sup>1</sup> 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.)