# Proposals for NORMAN Joint Programme of Activities 2026

| | |
|---|---|
| **Title** | **NORMAN Database System (NDS)** |
| **Type of activity** | Database maintenance and continuous update |
| **Leader** | EI |
| **Topic / activities** | ***Background / Justification for the proposed activity:***<br><br>Since the start of the NORMAN project in 2005 and then network in 2009, the NORMAN Database System (NDS) is at the core of the NORMAN activities, providing data and tools to fulfil its goals and visions. In many aspects it also supports the vision of the EU Common Data Platform where reliable data of transparent origin are to be used at the risk assessment of chemical substances in the environment. Also, it copes with the concept of One Substance One Assessment (OSOA) envisioned by the EU Chemicals Strategy.<br>In 2025, the NDS recorded **13 661 975 visits**. Ecotoxicology Database (16 059), SLE (10 724), Substance Database (10 613) and EMPODAT (7 924) being the most sought-for modules considering the visiting unique IP addresses. Notably, most of the visits came via API protocols targeting mainly Ecotoxicology Database and EMPODAT (>13 million). DSFP recorded increased interest (ca. 40 000 vs 30 000 visits in 2024) with highest numbers of visitors from USA, China, Singapore and EU (in that order).<br><br>In 2025, the database system was updated (see JPA 2025) to **NORMAN Database System 2.0** (**https://norman-databases.org**) up the level of latest developments in the field of database management and in the standardised framework. The source code is published on GitHub, and thus all members of the network can contribute to its further development. The database includes the Docker Service, PostgreSQL Database, connection to GITHUB, PHP/LARAVEL framework with professional API capabilities, and SSL certificate for secure browsing (https) and periodic management and server health-checks. The NDS 2.0 was presented at the NORMAN General Assembly in Gdansk in December 2025. In a short interim period both the 'old' NDS (https://www.norman-network.net/nds) and NDS 2.0 will co-exist. **Since February 2026 only NDS 2.0 will be accessible.**<br><br>The NDS consists of 13 integrated databases modules:<br>1. Suspect List Exchange - https://norman-databases.org/sle/sources<br>2. Substance Database - https://norman-databases.org/susdat/substances/<br>3. Chemical Occurrence Data (EMPODAT) - https://norman-databases.org/empodat/home<br>4. Ecotoxicology Database - https://norman-databases.org/ecotox/ecotoxhome<br>5. Digital Sample Freezing Platform (DSFP) - https://dsfp.norman-data.eu<br>6. MassBank Europe - https://massbank.eu/MassBank/<br>7. Antibiotic Resistance Bacteria/Genes - https://norman-databases.org/arbg/arbghome<br>8. Indoor Environment - https://norman-databases.org/indoor/indoorhome<br>9. Passive Sampling - https://norman-databases.org/passive/passivehome<br>10. Bioassays Monitoring Data - https://norman-databases.org/bioassays/bioassayhome<br>11. SARS-CoV-2 in sewage - https://norman-databases.org/sars/home<br>12. Substance Factsheets - https://norman-databases.org/factsheets/search/filter<br>13. Prioritisation - https://norman-databases.org/prioritisation/prioritisationhome<br><br>The backbone of the NDS is Substance Database, based on SLE Database. If the substance is in SusDat, it can be searched and displayed with all available data for this substance in any of the database modules. The NORMAN SusDat ID and StdInChIKey are used as unique identifiers for each substance, allowing for interlinking of the NDS modules. **All contributors to the NDS in general, and SLE in particular, are encouraged to assign NORMAN SusDat ID to the contributed substances** in any of the NDS modules. Once NORMAN SusDat IDs are generated (by using **Batch Conversion of Identifiers**; https://norman-databases.org/susdat/batch), they can be directly used for batch searches, e.g., in EMPODAT, DSFP or in the prioritisation modules.<br><br>The SLE Database underwent a significant development in 2025 and the new version will be accessible in 2026 (see separate JPA 2026 proposal on SLE). The SLE contained in the end of 2025 132 lists. There is a need to add substances from the newly added lists in 2024/2025 into SusDat, which contained 120,978 substances in December 2025. Additionally, it is necessary to collect all supporting data required for suspect screening (cf. separate JPA on DSFP) and prioritisation (cf. separate JPA WG-1) of the new substances. Also, a strong link between SLE and SusDat is to be developed in 2026, reflecting any of the curation changes on both sides.<br><br>A success story of cooperation with PARC in 2025 was extension of EMPODAT for PFAS compounds provided by all EU Member States resulting in 2,291,944 new entries in various matrices. Despite being successful after a laborious data curation, the exercise revealed serious deficiencies in current data interoperability (missing essential metadata, widely varying formats of data archiving) among the national official repositories of EU Member States.<br><br>A massive influx of new data on wide-scope target substances, non-target screening, suspect screening, passive sampling and bioassays from JDS5 is expected by the end of January 2026. All data must be curated and made available for further prioritisation (WG-1) and support of WG-2 (Bioactivity Database, |

Bioassays Monitoring Database) and WG-4 (Water Reuse).

A significant effort has been put into implementation of the automated protocol requested by the EC JRC for transfer of EMPODAT into IPCHEM on an annual basis.

In the NORMAN Ecotoxicology Database, **all NORMAN members are permanently invited to contribute with PNECs, especially those based on the experimental values.** Based on the specific request of the WG-8 (marine mammals) and WG-7 (terrestrial environment), a concept for model prediction of toxicity thresholds for mammals has been developed in close cooperation with the HORIZON TerraChem project using ToxAI. The resulting EC50/LC50 values were converted into ug/kg ww values commonly measured in tissues of mammals (e.g. liver) using selected PBTK models (also presented at the GA NORMAN in Gdansk). Lowest PNEC Biota mammals are now available in the Ecotoxicology Database for several hundreds of target substances detected in the LIFE APEX project. The predictions will be available in 2026 for up to 5 000 substances detected by suspect screening in LIFE APEX (see also JPA 2026 WG-1 activities).
The Ecotoxicology Database is facing a serious problem how to include the ever-increasing number of predicted PNECs by various models. In 2024, rat toxicity threshold values predictions were produced for SusDat substances using VEGA QSAR model. ToxAI is forecasting to predict toxicity threshold values for more than 90 specific end points (individual species, terrestrial/marine/mammals/birds etc.; traditionally PNECs in EU legislation are based on freshwater toxicity of algae/fish/daphnia). There is no concept how to deal with these new predictions **and all members of the network are invited to provide a framework** (see also JPA 2026 WG-1 activities). The marine and terrestrial (incl. soil) endpoints are of particular interest to many NORMAN members.

All modules contain API documentation. In order to enhance the FAIRness of NDS, on-line **accounts for assigning DOI to contributed datasets** has been developed using https://doi.datacite.org/. Any contributed dataset to the NDS will therefore has its unique DOI and can be cited in peer-reviewed literature. For more details contact alygizakis@ei.sk.

Two new NDS modules have been developed in 2024-25:

14. Bioactivity Database - https://www.norman-network.net/nds/badb/ (open access)
15. Hazards and Properties Database - https://www.norman-network.net/nds/hazards/ (for login/password contact valeria.dulio@ineris.fr)

Further input into both of the modules will be provided by WG-2 (see separate JPA 2026 proposal on Bioactivity Database) and WG-1 (see JPA 2026 proposal on WG-1 activities).

Beta versions of two new NDS modules have been developed in close cooperation of WG-1, WG-8 experts and HORIZON TerraChem project:

16. EMPODAT Suspect Database - https://norman-databases.org/empodat_suspect/home (for login/password contact slobodnik@ei.sk)
17. Literature Chemical Exposure Database (short Literature Database) - https://norman-databases.org/literature/home (for login/password contact slobodnik@ei.sk)

**EMPODAT Suspect** is currently hosting 34 126 229 data entries obtained by retrospective suspect screening of 92 988 substances in 367 samples (proposed by WG-8) stored in DSFP. Each data entry is accompanied with the IP score, referring to the identification confidence level (Dulio et al, 2024; DOI: 10.1186/s12302-024-00936-3). It is planned to gradually include such 'suspect screening extract data' from most of the DSFP samples (at present ca. 6 000), which may result in database handling/prioritising >550 million data in near future (cf. JPA 2026 WG-1 activities).
**Literature Database** prototype is a product of AI-based data mining in peer-reviewed and grey literature dealing with chemicals in terrestrial environment with focus on top predators. The first datasets (176 296 data entries) came from the University of Leiden, whereas large datasets from NATURALIS (the Netherlands; both non-NORMAN members) are in the pipeline. EI has recently downloaded pesticide monitoring data from the EU LUCAS soil monitoring project. The NDS is sought for as a sustainable repository for such data. Similar data-mining efforts are on-going among many NORMAN members and in PARC (e.g. NIVA-lead efforts). There is a vision of using the literature data as a part of the EU Early-Warning System and benchmarking occurrence of emerging chemicals in environment in areas in which there are no fixed EU legacy monitoring programmes (e.g. terrestrial environment, soil, indoor environment, passive sampling, water reuse, marine environment, etc.). Both modules were presented and discussed at the NORMAN GA 2025 in Gdansk. **All members of the network are invited to participate at the further development of the modules and population with data.**

Two automated prioritisation modules (Dulio et al., 2024, see also JPA WG-1 activities) are being updated/developed in the NDS:

18. Prioritisation tool for target substances archived in EMPODAT (for access, contact valeria.dulio@ineris.fr)
19. Prioritisation tool for suspect screening substances archived in EMPODAT Suspect (for access, contact valeria.dulio@ineris.fr)

The prioritisation tools are not publicly accessible, as the NORMAN network would like to control outcomes of prioritisation efforts labelled as 'NORMAN'. Any local/national/regional prioritisation efforts using these tools should be carried out in close cooperation with WG-1. The prioritisation tool for target substances got updated for link to hazard score (PBMT properties; see Hazards and Properties Database - https://www.norman-network.net/nds/hazards/). Also, biota matrices (fish/molluscs/mammals; available

previously only in LIFE APEX prioritisation tool) has been included. An update for scoring CMR/ED properties (already available using the JANUS tool from VEGA platform for >93 000 substances) is expected in 2026.
The prototype of prioritisation tool for EMPODAT Suspect is under development (see also EMPODAT Suspect Database - https://norman-databases.org/empodat_suspect/home) and will be finalised in 2026. **Contribution of all NORMAN members on its further development is welcome.** In 2026, the prioritisation activities will have a specific focus on WG-8 (marine environment), Joint Danube Survey 5 (freshwater environment) and WG-7 (terrestrial environment). A set of tools allowing for visualisation of prioritised data/indicators will be further developed in 2026.

Extracts of passive samplers (hydrophobic, polar, PFAS) obtained from JDS5 will be analysed by GC- and LC-HRMS non-targeted and suspect screening methodologies. It has been proposed at the NORMAN GA that the NTS data should be stored in DSFP and resulting suspect screening data (ca. 93 000 substances) in EMPODAT Suspect – labelled as a separate 'matrix'. The EMPODAT Suspect will be updated accordingly to accommodate passive sampling data.

NORMAN Artificial Intelligence Workshop, which tool took place in Leipzig in October 2024 recommended to establish a new NDS module on:

20. Models Training and Validation Datasets Database – ALL open access models used to feed NDS modules with source code reported on GitHub and their training and validation datasets to support transparency.

After a preliminary screening, the models used in TerraChem project (terrestrial environment and biodiversity) and ToxAI tool (toxicity thresholds for numerous endpoints) were selected as a starting point. **All members of the network are invited to contribute to the design of the database in 2026.** The long-term vision is to establish a transparent link to all source data/parameters used for model predictions (including those not yet stored in the NDS) as a part of the EU Common Data Platform and OSOA strategy. The activity is linked with JPA 2026 WG-2 proposal on 'Blueprint for linking ecotoxicity to different levels of biodiversity damage, building on and integrating various NORMAN data streams: Focus on workflow example implementation'.

The NDS has been proposed to be a part of the European infrastructure hosting data on microplastics (see also JPA 2026 proposal on 'Defining guidelines for including microplastics into NORMAN Database System'. This will require development of:

21. Microplastics Database

The structure of database will be based on to-be-developed harmonised Data Colection Templates (DCTs) covering major sampling and analysis techniques. Additionally, a strategy how to archive/refer to complex spectral datasets (such as particle based as well as hyperspectral data; FTIR, Raman, GC-MS) will be developed. This will allow for inclusion of microplastic data obtained from instruments using different measurement approaches and guarantee data comparability on a long term. An interlink with EMPODAT and EMPODAT Suspect data specifically addressing plastic additives and NIAS in the same samples as those submitted for microplastics analysis will be created (if feasible; see also JPA 2026 proposal on 'Plastic associated chemicals').

***Description of the proposed activity and expected outcomes for 2026 (and beyond):***

The proposed task for maintenance of the NDS and its continuous upgrade in 2026 include:

- Continuous upgrade, maintenance of all modules; curation and upload of newly provided datasets.
- Curation of interlinking of all NDS modules with new data inputs.
- NDS Chemical Occurrence Data (EMPODAT): maintenance, upgrading and feeding of new data into the database; sharing the data with IPCHEM.
- Upload of JDS5 data into NDS, link to prioritisation modules.
- Update of the 'Hazards and Properties' database module for CMR/ED properties and new datasets, interlinking with the prioritisation tools for target and suspect screening.
- Further development of API portal allowing for automated data sharing with external databases.
- Assigning DOIs to newly contributed datasets in the NDS.
- Further development of the Literature Chemical Exposure Database (task supported by TerraChem project); upload of already available datasets; establishment of a workflow for automated processing of AI-obtained data into the 'NORMAN format'.
- Further development/upgrade of visualisation tools in the NDS and prioritisation tools.
- Update of EMPODAT Suspect module to accommodate NTS passive sampling (suspect screening) data from JDS5 as a separate matrix.
- Upload of new data from JDS5 and Black Sea EU4EMBLAS projects into the ARBs/ARGs module.
- Upgrade of Substance Factsheets module in the NDS 2.0. (link to prioritisation modules).
- Development of the Models Training and Validation Datasets Database *(if approved, in collaboration with Data Science CWG)*.
- Preparation for possible development of the Microplastics Database (in collaboration with WG-4).

***Added value / Link with other NORMAN activities and / or other projects***

| | |
|---|---|
| | The proposed tasks will benefit all WGs and CWGAs in the NORMAN network. |
| **Participants** | EI, all interested members |
| **Proposed in-kind contribution** | All – contribution of existing data and design of new NDS modules<br>EI – overall coordination, development of beta versions of Literature Database and EMPODAT Suspect and population with test data |
| **Contribution needed from NORMAN Association**[1] | Maintenance and continuous update of the NDS:<br>- EI: 42,000 €<br>Rental of the server hosting the NDS, management and backup system:<br>- EI: 10,600 €<br>Development of the prototype of the Models Training and Validation Datasets Database<br>- EI: 4,500 € |

[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.)