

Proposals for NORMAN Joint Programme of Activities 2024

Title	NORMAN WG-1 Prioritisation
Type of activity	Working Group activities
Leader	INERIS, EI and UBA
Topic / activities	<p>Background / Justification for the proposed activity:</p> <p>Better chemical management in line with the goals of the EU Chemicals Strategy and Zero Pollution Action Plan strongly relies on identification and control of pressures through efficient monitoring and knowledge of the properties and use patterns of chemicals. Given the high number of chemicals present in our environment, prioritisation of chemical contaminants is highly demanded by regulators and decision-makers in order to identify and justify the most urgent actions necessary to achieve these goals.</p> <p>The tasks of WG-1 aim to:</p> <ul style="list-style-type: none"> - Ensure that all relevant information for assessment of chemical risks is maintained and regularly updated in a timely manner; - Integrate and exploit this information to identify lists of chemicals in need of priority actions (each priority list corresponds to an action); - Upgrade the prioritisation scheme in a constant, dynamic manner (integration of innovative techniques) for higher added value of the prioritisation work and a more reliable and meaningful interpretation of the prioritisation results. <p>Description of the proposed activity and expected outcomes for 2024 (and beyond):</p> <p>Task 1: Support the prioritisation work of the Commission services at European level and provide comment on relevant consultation documents (PARC, EWS, where relevant) (ALL)</p> <p>NORMAN participates as a stakeholder in the WG Chemicals of DG ENV and intends to further contribute to the activities related with the review of the list of WFD Priority Substances and the Watch List.</p> <p>In 2023 NORMAN WG-1 provided proposals for the 5th review of the Watch List. In 2024, NORMAN will be able to contribute to the PARC partnership on the following prioritisation-related topics:</p> <ul style="list-style-type: none"> ▪ Collaboration in the <i>development of the mechanism for priority setting in environmental and multi-source monitoring – MonitoringFrame project</i> (PARC T4.2.1). The NORMAN scheme was recognised as an efficient and pragmatic way to address regulatory questions while addressing existing data gaps. Next steps will be to extend the scope of the NORMAN approach: the prioritisation should cover multi-compartments and should address human and environmental health. Overall, the ambition is to define a cross-regulation and cross-compartment prioritisation mechanism, where the selection of chemicals/matrices/endpoints (effects) would be based on multiple lines of evidence. ▪ Contribution to the <i>selection / prioritisation of EDCs for the PARC pilot monitoring study</i>: Suspect screening of the list of candidate EDCs in the samples of the DSFP to have an overview of the estimated frequency of occurrence of these compounds and estimations of semi-quantified concentration levels. ▪ <i>Early Warning System for Europe</i> (EWS) NORMAN can contribute with NTS, but also EMPODAT, substance info (hazardous info) and prediction models for prediction of hazardous properties. ▪ Potential collaboration of NORMAN WG 1 and WG Soil with JRC to work on a priority list of substances for soil. <p>Task 2: Collection and prediction of compound-specific information in support of prioritisation (UBA / EI / NKUA / DERAC / INERIS)</p> <p>After finalisation of the harmonised ecotox DCTs, the still pending extraction and compilation of additional experimental ecotoxicity data from other existing ecotox databases, i.e. the REACH portal and the UBA ETOX database, will be started in 2024. Regulatory quality targets for various matrices, including re-use, soil and marine waters, will also be collected and compiled (see also task 3).</p> <p>Moreover, data on physico-chemical properties (e.g. Kow, Koc, BCF) as well as hazardous properties (i.e. related to ED, CMR, PBT, PMT) will be retrieved from REACH registration dossiers and future CLP dossiers and integrated in the Substance Factsheets to support the hazard score in the prioritisation module. For this purpose, a respective DCT will be developed. In case of the lack of experimental data on these hazards, prediction models using artificial intelligence will be developed and used to fill the gaps.</p> <p>The P, B, T, CMR and ED scores had been assigned to 76,269 compounds in the SusDat list by UBA and EI using the JANUS model. However, difficulties to run the model for the remaining ca. 30,500 chemicals (before update of SusDat by additional ca. 14,000 substances in the end of 2023) persist. Hence, it is proposed to develop a quantitative model for persistence, based on a set of 8000 reported</p>

degradation half-lives for the OECD 302 tests. The models will allow to **predict the DT50 in water for all compounds in SUSDAT** that are covered by the model domain.

In order to ensure that only reliable predictions are reported in the Substance Factsheets and used for the prioritisation of chemicals which lack experimental data, it is proposed to develop a **new General Application Domain (GAD)** that can be adapted to all models that are used in NORMAN. The GAD should ensure that the prediction models are not overfitted, while structures that are clearly not covered by the model are detected with high confidence.

Dedicated meetings will be organised with an appointed **Group of Experts in modelling and machine learning to reach consensus on the use of the predictive models to be used and the application of the GAD.**

Task 3: Collection of existing PNECs and prediction of PNECs (UBA / EcotoxCentre / DERAC / EI)

After the successful extension of prediction models for 3 additional fish species, 2 crustaceans and one insect species, it is proposed to **extend the set of prediction models, to also cover acute effects, as well as chronic effects in fish, daphnia and algae, for both aquatic species with sufficiently large experimental datasets.** The final aim is to enable a more robust statistical approach for the derivation of freshwater P-PNECs and their subsequent conversion for application to other matrices, like sediments and biota, e.g. by using Koc and BCF (see task 2). Thereby, the use of chronic-based endpoints will allow to lower the AF for deriving the P-PNEC, which will be appreciated by many regulatory bodies that are using the prioritisation results of NORMAN.

For 2024, it is therefore planned to **continue with the predictions of toxicity values for both, existing compounds and those that have recently been added to SusDat**, in order to be up to date with the list of Lowest PNECs. The upload of existing PNEC and the derivation of new PNECs will also include **MAC-EQS values.** This will allow the evaluation of peak exposure concentrations, such as those resulting from the German small water body monitoring project.

It is also planned to derive more robust **PNEC marine values** to support prioritisation of compounds in the **marine environment** (delayed from JPA 2023).

Moreover, new compartments, such as soil or sediments will be considered in 2024 to extend the scope of the current prioritisation scheme, thereby allowing consideration of multiple lines of evidence from different compartments. Regulatory quality targets in soil will be converted into quality targets for sewage sludge re-used in agricultural fertilisation (see task 5).

WG-1 will continue to **promote and coordinate the participation of Ecotoxicity Experts to derive and approve (i.e. vote for) new or revised Lowest PNEC values** for substances of the SusDat list, with a specific focus on substances that were highly prioritised in EU projects using the NORMAN Prioritisation Framework. The aim is to **progressively replace predicted PNEC values for substances prioritised in Cat 3 and 5, by experimentally-based PNEC values**, with a special focus on comprehensive datasets, including new data that had not been considered yet.

Task 4: Compilation of data / information regarding 'Use categories' and 'Chemical Functional Use' for all SusDat compounds (LCSB / UBA / INERIS / EI / OVAM (tbc))

The following has already been proposed at the WG-1 meeting in June 2022 and has subsequently been discussed among some experts. Some effort has already been made by EI to derive 18 use categories using the labels of the SLE lists, which are already implemented in the SEARCH function of SusDat. The full operational implementation of this task is postponed to 2024. With the ever-growing number of NORMAN SusDat substances, it becomes inevitable to group them by 'Use category' and 'Chemical groups category'. The NORMAN Substance Factsheets already contain information retrieved from the US EPA Dashboard, but at present this is not sufficient to allow searching in the NDS by e.g., all pesticides, all herbicides; all pharmaceuticals, all antibiotics; or by, e.g. all PFAS, pyrethroids, PAHs, ionic surfactants. The classification of the compounds is not trivial, since each substance has often multiple uses and associated sub-classes. The different approaches applied for categorisation of the uses (e.g. by REACH, US EPA Dashboard, PubChem) are not fully harmonised. In the same context, categorisation of transformation products and their linkage to single / multiple parents is another challenge that NORMAN wants to address.

The collection of this information is crucial to the prioritisation work (e.g. identification of contamination sources for which actions are required, exploitation of lists of substances).

To achieve this objective, we propose to organise a strategy within WG-1 in line with the on-going collaboration between NORMAN and PubChem addressing the following sub-tasks:

- Finalise and approve the Use Category DCT template (as agreed on in the last WG1 meeting). This will require controlled vocabulary and a common understanding about "use categories" and "functional uses". We should aim at a harmonised classification for NORMAN as well as for PARC.
- The DCT should be uploaded into a newly developed sub-module and should be used as an interface for experts to be able to interact online (to upload the information and amendments where needed). All changes made by the experts should be traceable (as it is done in the Ecotox module). All the

- information (use and functional categories) will be directly linked and searchable from the Substance Factsheets / SusDat and the prioritisation system.
- c) Discuss in consultation with the WG what the best platform is to create a group repository for NORMAN to exchange expert views and establish harmonised "use" terms (e.g. GitLab, GitHub page for the NORMAN network similar to MassBank consortium). Should the platform be open to NORMAN members only? LCSB proposes that it should also be open to external experts like the PARC consortium?
 - d) The DCT is meant to be a living tool. The experts in WG-1 are invited to regularly review and upgrade the categories (e.g. add the mode of action) and harmonise the drop-down entries. Volunteers in WG1 will be appointed to be in charge of complex groups of compounds (e.g. all REACH or PFAS compounds) to suggest their categorisation.
 - e) Extraction of "Use" information from the various SLE lists (already available) and export of this information into PubChem (LSCB & PubChem Team, part of SLE activities) and into the DCT (citing SLE as the reference source). WG-1 members will help to prioritise the most relevant lists.
 - f) Pilot efforts to retrieve "Use" information from PubChem and other sources (e.g. Wikipedia, Wikidata, US EPA Dashboard, ChemSpider etc.) and integrate them into the Use category module. Translation tables will be created for each source to ensure harmonised entries compatible with the NORMAN Use category module (EI, LCSB, UBA, to be discussed).
 - g) Programming the workflow resulting from the pilot efforts (cf. e) for the transfer of the information retrieved from PubChem and other sources into the Use category module (EI).

Task 5: Prioritisation framework: follow-up activities to apply the new workflow beyond the freshwater compartment and across compartments and for specific chemical groups (ALL)

The work in 2023 has focused on the finalisation of the paper, the development of the infrastructure (back-end) to implement the online prioritisation tool linking the compounds' data in SusDat, target monitoring data (EMPODAT), suspect screening data (EMPODAT-SUSPECT under construction) and PNEC values (ECOTOXICOLOGY database). This operation has represented an important computational challenge (big data) which has required the preparation of structured data and indexing methods, the testing and further optimisation of the IT technology to allow fast data processing. In 2024 the infrastructure will be ready, and it will be possible to make operational the link between target and suspect data using the workflow applied in the WW pilot study (publication under submission). The focus of 2024 will be on:

- a) **Development of the IT infrastructure (EI and ALL):** the remaining tasks in 2024 will be: i) programming the workflow; ii) screening all the data collections / projects in the DSFP + EMPODAT; iii) storing the outputs in EMPODAT-SUSPECT database as standardised structures (to allow for computational processing); iv) designing the layout of the user interface (front-end), i.e. defining the search fields and the parameters that the user will be able to choose, visualisation of the outputs, etc. These features will be defined in consultation with the experts in WG-1 (meetings to be organised in 2024) and PARC (Collaboration in Task 4.2.1) before the programming phase (low fidelity >> high fidelity prototype).
- b) Application of the **prioritisation workflow** to the **list of candidate EDCs** provided by **PARC (collaboration WG-1 and PARC)**. A comprehensive compilation of information on regulated and non-regulated potential EDCs has been performed in PARC T4.2 in 2023. The list has been submitted for curation and inclusion in the SLE. In 2024 it will be possible to start the screening of data collections available in the NDS and prioritise the compounds using the NORMAN scheme.
- c) Application of the **prioritisation scheme (target and SS) to different data collections** (see below) representing as far as possible different environmental compartments: it will be possible to characterise the profile of the compounds across multiple environmental compartments and identify the priority compounds supported by multiple lines of evidence:
 - Danube River Basin Specific Pollutants based on wide-scope target and suspect screening using JDS4 data;
 - Black Sea Specific Pollutants based on wide-scope target and suspect screening using EMBLAS project data;
 - Prioritisation of contaminants in top predators and their prey using LIFE APEX project data.
- d) **Prioritisation of contaminants in soil (WG-1 and WG-7):** the work on prioritisation of contaminants in soil will remain *in stand-by until a critical mass of soil datasets* is available in the NDS. The newly EU-funded project TerraChem is expected to feed new data points in the NDS in near future.
- e) Identification of list of ranked SusDat chemicals relevant for **acquisition of their chemical standards or MS(MS) information** allowing for target screening (**Eawag, LCSB**). This is an ongoing effort which started with a first test in 2022. A list of candidate compounds for acquisition of mass spectra in MassBank was prepared by EI and INERIS in 2022, using the results of the WW case study (NTS prioritisation scheme applied on the list of 65,690 SusDat compounds suspect screened in 84 WW effluents from 19 countries (LC-HRMS in ESI+/-). The process will be pursued in 2024. Thanks to the results of the application of the prioritisation workflow on multiple data collections it will be possible identify the overlaps in the priority groups and in this way obtain evidence for acquisition of standards for MassBank. The final selection of a list of the "most wanted" spectra will consider the information available in NORMAN SLE, PubChem and CompTox, and check which compounds are not in MassBank or other open spectral libraries already. Once purchased and measured, the data should become available in the NORMAN Database System (e.g. MassBank, NORMAN-SLE).

Task 6: Prioritisation of contaminants in reused matrices (WG-1 and WG-5): In 2023, WG-5 revised the design of six EMPODAT DCTs to allow their application to reused matrices. New occurrence data

based on literature review will be compiled in EMPODAT for reused practices, i.e. stormwater, reclaimed water and sewage sludge in 2024. Moreover, new quality targets for reused matrices in agricultural practices were collected and uploaded in the ECOTOX database and further quality targets in sewage sludge for agricultural fertilisation will be derived (Task 3 in WG-5).

Thanks to this new input in 2024 it will be possible to start testing the NORMAN prioritisation workflow on the specificities of the reused matrices. In this case, the user interface will allow the user to select matrix and the reuse route, and the system will automatically select the proper PNEC, depending on the matrix and the prioritisation objective. For example, for prioritisation of contaminants of concern in reclaimed water for reuse in irrigation, it will be possible to select the wastewater effluent datasets labelled as "reclaimed water / reuse irrigation" and match the concentration data with the corresponding PNEC_crop, specific for protection of human health.

Task 7: Improvement of the features of the prioritisation tool (expert consultation and programming activities) (ALL) (task planned in previous JPAs, postponed to 2024) (EI, INERIS, UBA, UFZ, NKUA, OVAM, DERAC, LCSB, KWR)

Further additional options for improvement of the prioritisation tool have been discussed in the Database workshop in 2023 and within the Prioritisation WG-1. At the moment, the output of the prioritisation tool is a downloadable Excel file with the list of substances sorted by category and associated scores.

WG-1 proposes to explore the integration of unsupervised clustering as a new feature besides the conventional output (table):

- Principal component analysis (PCA)
- Hierarchical clustering
- Self-organising maps
-

As regards the visualisation options, it will be possible to consider: Network analysis, Trends analysis, Heatmaps and Venn diagrams. All these tools will allow to draw more conclusions about the prioritisation results (beyond the action categories).

Task 8: New development of the prioritisation tool in collaboration with PARC (EI, INERIS, UBA, NKUA, UFZ, OVAM, LCSB)

The development of the **Prioritisation project in PARC** involved the start of a collaboration activity with WG-1. In the PARC project it was decided to create a system flexible for new prioritisation queries (regulatory and research needs). This would include the extension of the list of indicators and the possibility for the user to customise them, including the weight of each indicator for the scoring system. It was proposed that the NORMAN Database System could provide the infrastructure for the programming of the prioritisation tool proposed in PARC. If this option is confirmed, a new prioritisation tool will be developed and in the beginning it will co-exist with the current one. This task will be organised in two steps:

- 1) Meetings (virtual, if possible 1 physical meeting) for consultation of the WG-1 experts and PARC: proposals / desk work to revise the features of the prioritisation tool.
- 2) Programming work: expected to start in the second part of 2024 after agreement with the WG-1 experts, based on dedicated instructions derived by a smaller team of experts.

Task 9: State of the environment in Europe (task planned in previous JPAs, will be postponed to 2024) (EI, UBA, INERIS and ALL)

We would like to integrate the results of single substances risks at certain sites into a total effect, by mapping the concurrent occurrence of chemical contaminants in Europe, across compartments: e.g. "how many compounds exceed the lowest PNEC at each site"?, or "what is the total mixture risk at each site"? These might become new chemical indicators to measure progress in the quality of the environment e.g. as a result of remediation measures and inform policy decisions (cf. pesticides indicator of the EEA) about the trends. This would allow NORMAN to, e.g., support the Zero pollution monitoring report of the EEA. Moreover, analysis of datasets regarding the minimum (most toxic compounds) compared to the total effect (mixture risk) might allow to derive an indication of a retrospective mixture allocation factor (MAF) in support of the COM EGD initiative.

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

- Support the preparation of the programme of activities of the PARC partnership, with particular focus on inventory of existing data and common knowledge on chemicals in order to identify gaps in data and anticipate future trends.
- Support to development of an Early Warning System for Europe (PARC 8.2)
- Support to DG ENV / JRC for the review of the list of WFD Priority Substances, the Watch List for surface water and the Watch List for groundwater.
- Support other Member States or River Basin authorities if requested.
- Links with WG-8; Support prioritisation of CECs as Sea Specific Contaminants in the marine environment (OSPAR, HELCOM, MED POL, Black Sea Commission).
- Links with WG-5 for identification of priority contaminants in treated wastewater intended for reuse
- Links with WG on Soil and the Terrestrial environment for prioritisation activities.
- Link with NTS CWG, in particular as regards DSFP, SLE and the definition of a mechanism for updating the NORMAN List of emerging substances as top priority substances of SusDat.
- NORMAN MassBank: prioritisation of relevant compounds whose mass spectra are not yet included

		<p>in this online repository.</p> <ul style="list-style-type: none"> - Interlinking with the entire NORMAN Database System.
Participants		<p>INERIS, EI, UBA, Ecotoxcentre, NKUA, LCSB, OVAM, UFZ, KWR, DERAC (All WG-1 are welcome)</p>
Proposed contribution	in-kind	<p>In-kind contribution is foreseen in all proposed activities as a complement to NORMAN funding.</p> <p>INERIS: coordination activities; contribution in dataset validation for modelling, conceptualisation of the modules, interface / visualisation; use categories for substances;</p> <p>UBA: coordination activities, with a focus on Ecotox and machine learning Expert Group; compilation of datasets and their validation for modelling, conceptualisation of the modules, interface / visualisation; DCT development, use categories for substances;</p> <p>EI: programming of the database modules, derivation of prediction models; running the predictions for SusDat compounds; running of suspect screening workflow (DSFP) for prioritisation case studies;</p> <p>NKUA: programming;</p> <p>LCSB: use categories for substances: preparation of meetings, communication with other data sources</p> <p>Ecotoxcentre: contribution in compilation of existing PNECs and validation of datasets for the ECOTOX module;</p> <p>DERAC: compilation of existing PNECs and connection with WG-5 for prioritisation of substances in matrices of interest for "reuse";</p> <p>UFZ: conceptualisation of the modules, interface / visualisation; use categories for substances;</p> <p>OVAM: expert advice for the development of the prioritisation tool in collaboration with PARC.</p>
Contribution from Association¹	needed NORMAN	<p>Total net budget under JPA 2024: 23,000 € (+ 17,600 € postponed from 2023)</p> <p>Task 1: Support the prioritisation work of the Commission services at European level and comment on relevant documents and queries (PARC, EWS, where relevant). (in kind)</p> <p>Task 2: Collection and prediction of compound-specific information in support of prioritisation (and general application domain) (EI/ UBA/ NKUA):</p> <ul style="list-style-type: none"> - Retrieval and harmonisation of ecotox data from external databases (2,000€ not spent in 2023) - Compilation and model prediction of hazardous properties (2,000€) - General Application Domain (2,000€) <p>Task 3: PNEC derivation module (EI/ UBA/ NKUA):</p> <ul style="list-style-type: none"> - Model development for predictions of acute and chronic toxicity values for additional 10 species (5,000 €) - Application of the models for existing and new compounds in SusDat (2,000 €) - Compilation and uploading of regulatory PNECs (1,000 €) <p>Task 4: Compilation of data / information regarding 'Use categories' and 'Chemical Functional Use' for all SusDat compounds (LCSB / UBA / INERIS / EI / OVAM)</p> <ul style="list-style-type: none"> - Organisation and harmonisation / Collection and compilation of data / information available on 'Use categories' and 'Chemical Functional Use'. (LCSB: 3,000€ not spent in 2023) - Programming the workflow resulting from the pilot efforts for the transfer of the information retrieved from PubChem and other sources into the Use Category module. (EI: 3,600€ not spent in 2023) <p>Task 5: Prioritisation framework: follow-up activities to apply the new workflow beyond the freshwater compartment and across compartments and for specific chemical groups (ALL)</p> <ul style="list-style-type: none"> - Programming work of the back-end and interface of the extended prioritisation module bringing together suspect and target screening data (EI: 4,000€) <p>Task 6: Prioritisation of contaminants in reused matrices (WG-1 and WG-5) (budget in WG-5/NDS)</p> <p>Task 7: Improvement of the features of the prioritisation tool (consultation and programming activities) (ALL)</p> <ul style="list-style-type: none"> - Design and update of the interface for spatial visualisation of the prioritisation results (to be agreed in consultation within WG-1 further to the outcomes of the NORMAN Database Workshop (EI: 5,000€ not spent in 2023) <p>Task 8: New development for the prioritisation tool in collaboration with PARC</p> <ul style="list-style-type: none"> - Programming work of the back-end and interface of the prototype of the new prioritisation module in collaboration with PARC (EI: 7,000€) <p>Task 9: State of the environment in Europe</p> <ul style="list-style-type: none"> - Programming of the module and visualisation of the results (EI / UBA / INERIS: 4,000 not spent in 2023)

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.)