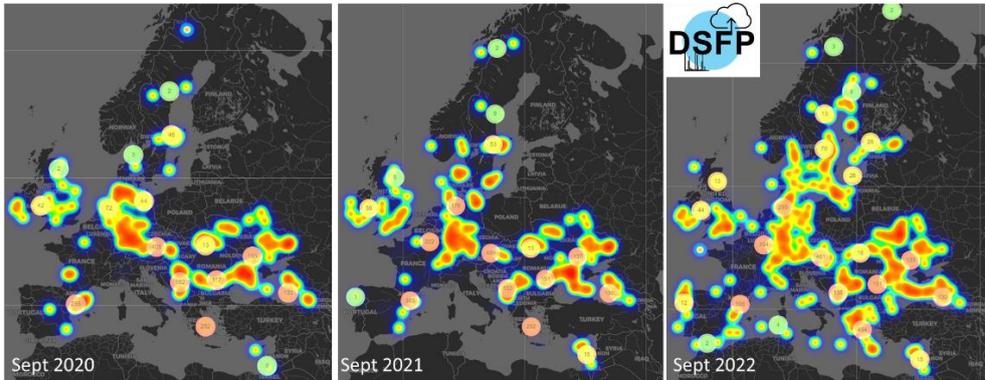


Proposals for NORMAN Joint Programme of Activities 2023

Title	Improvement and maintenance of DSFP
Type of activity	Databases
Leader	Nikiforos Alygizakis (EI-Environmental Institute & NKUA-National and Kapodistrian University of Athens), Jaroslav Slobodnik (EI), Nikolaos Thomaidis (NKUA)
Topic / activities	<p>Background / Justification for the proposed activity:</p> <p>The previous activities on upscale of DSFP were implemented successfully and established a significant infrastructure for NORMAN. The purpose of the upgrade was to bring the prototype to a production-ready facility with improved informatic characteristics (usability, maintainability, scalability, performance and extensibility).</p> <p>Briefly, the technical solution consisted of the following elements:</p> <ol style="list-style-type: none"> 1. The creation of a back-end Dataset Management System (DMS). DMS was designed as a single point of data contribution and administration. The users, who wish to contribute their chromatograms register through a registration wizard. Each user belongs to an organization and can register instruments, instrumental setups, upload and screen its own data under dedicated user panels. 2. The DMS installation is accompanied by a search index (API), that keeps track of the latest data assets contributed to the platform. The API assures interoperability and allows communication of DSFP with other external sources. Moreover, it provides programmatic access to the contributions and will allow batch contributions in the future. 3. A front-end interface was created. It has two major components: a compound discovery application and an open data catalogue. The open data catalogue serves as the FAIR data showcase of the repository, listing collections of data. The compound discovery application uses fast indexing technologies and NoSQL approach for searching of compounds in the chromatograms. The adopted representations are advantageous, because it allows extensibility and rapid search of compounds in the sample collections. 4. All contributions were transferred to the new facility, whereas previously used R and python functions are hosted as serverless functions in docker images, which are accessed via restful APIs. This approach allowed the execution of the NTS workflow at high scale in an isolated virtual environment, which contains all needed software and dependencies. <p>More information about the applied technical solution is available in the project proposals of previous years. DSFP is in position to facilitate fast publishing, effective management and open access of research assets. The platform enables archiving, processing, analysing, data mining and retrieving information for thousands of contaminants of emerging concern (CECs) contained in high resolution mass spectral (HRMS) data. The database is a truly unique effort to collect HRMS data from environmental samples that enables for risk assessment of CECs, which is then intended for use by the policy makers and regulators.</p> <p>As of September 2022, the platform has been populated with HRMS data of more than 3,196 unique environmental samples around Europe and beyond. Despite the upscale activity, a steady increase in numbers of samples was achieved in 2022. The number of samples is expected to increase exponentially during the next years due to the created facility and the increasing participation of NORMAN Association members in significant projects such as the PARC project. The samples covered water (36.87%), biota (30.49%), wastewater (20.75%), sediment (4.53%), groundwater (3.88%), soil (2.47%) and other matrices (1.01%).</p> <div style="text-align: center;">  <p>1850 specimens +30.8% 2421 specimens +32.0% 3196 specimens</p> </div> <p>Figure 1. Spatial coverage of samples contributed to DSFP as of September 2022</p> <p>DSFP has been used in a series of sampling campaigns and collaborative trials of NORMAN Association,</p>

	<p>providing satisfactory results in terms of identifying substances usually overlooked by target and non-target screening. DSFP is nowadays frequently used in various large-scale monitoring campaigns and acting as a safety-net for the detection of potentially hazardous substances. DSFP has proven to be useful for various activities of the NORMAN network, especially at the prioritisation of CECs and as an early-warning system for environmental chemical risks. DSFP, as part of NORMAN Database System (NDS), is a valuable asset for the future activities of NORMAN Association.</p> <p>The continuous development of functionalities of DSFP is of importance to expand its use and further enrich the database. Changes that would make DSFP more efficient and that would stimulate the interest of researchers to upload their data and thus increase the collection of the HRMS data remain still a very important objective.</p> <p>For this reason, the JPA proposal aims to maintain and improve the functionalities of DSFP. It is worth noting that a roadmap for DSFP has already been set.</p> <p>The following major elements of the roadmap can critically improve the infrastructure:</p> <ol style="list-style-type: none"> 1. Creation of a data exploration tool. The tools that have been previously developed within projects such as LIFE APEX (spatial distribution visualization, time-trend analysis and co-occurrence analysis) will be migrated 2. Incorporation of the IP scoring to DSFP for reproducible and transparent reporting of the results 3. Integration of DOI system to the repository 4. Possibility to import compounds that are currently not part of SusDat 5. Improvement of the API infrastructure and its documentation 6. Programmatic contribution of collections via a batch contribution mechanism 7. Creation of a ready product for scientific projects 8. Support for 4D-HRMS data 9. Establishment of the DSFP early-warning system (aligns with the goals of PARC project): Connection of laboratory instruments with DSFP web services, event broadcasting and trigger mechanisms for event handling 10. Expansion to multiple analytical platforms 11. Creation of a data marketplace 12. Better integration with internet search engines (Google, Bing etc.) that will increase the findability of the collections <p>Added value / Link with other NORMAN activities and / or other projects</p> <ul style="list-style-type: none"> ▪ NORMAN NDS ▪ Cross-Working Group Activity on Non-target Screening (NTS) ▪ WG 1 Prioritisation ▪ PARC initiative ▪ National and EU-funded projects
Participants	EI, NKUA, UFZ, NILU, all NTS members of NORMAN and any other interested members.
Proposed in-kind contribution	Working hours for implementation the project (EI/NKUA).
Contribution needed from NORMAN Association¹	Total funds required for improvement and maintenance: 14,000 €

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.)