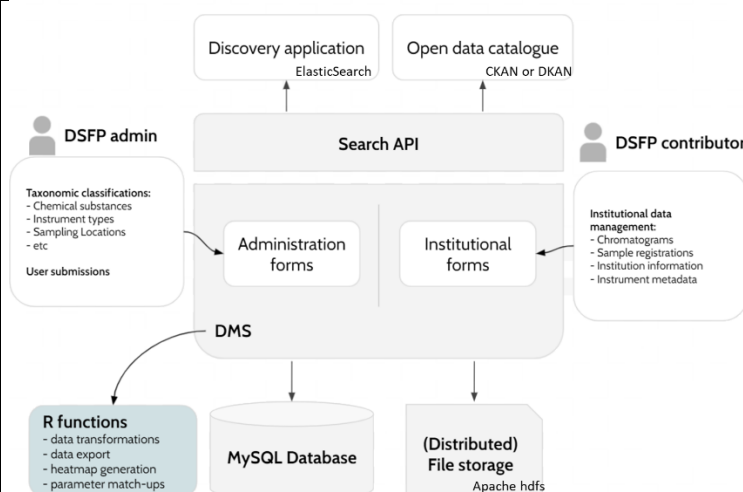




## Proposals for NORMAN Joint Programme of Activities 2022

Title	Development and upscale of DSFP - Phase II
Type of activity	Database development and upscale
Leader	Nikiforos Alygizakis, Jaroslav Slobodnik (EI-Environmental Institute), Nikolaos S. Thomaidis (NKUA-National and Kapodistrian University of Athens)
Topic / activities	<p><b>Background / Justification for the proposed activity:</b></p> <p>A platform for archiving, processing, analysing, data mining and retrieving information for thousands of contaminants of emerging concern (CECs) contained in high resolution mass spectral (HRMS) data has been developed. The database is a 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 2021, the platform has been populated with HRMS data of more than 2,500 unique environmental samples around Europe and beyond. The samples covered water (50.02%), biota (22.78%), wastewater (18.81%), sediment (3.98%) and soil (3.11%) matrices. DSFP has been used in a series of sampling campaigns and collaborative trials of NORMAN Association, 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.</p> <p>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 <b>valuable asset</b> for the future activities of NORMAN Association. It is also expected to play an important role in PARC, considering that its functionalities will be expanded to human biomonitoring as well. 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. For this reason, JPA proposal to upscale and improve DSFP has been launched.</p> <p>The upscale involves creation of a Dataset Management System (DMS), which is 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 instrumental setups, upload and screen its own data under dedicated user panels. Data accessibility is organizational, but sharing of assets across organizations is possible. The DMS is built on top of popular open-source CMS software (Drupal), which handles user membership and grouping, and access control on top of the different data types. DMS administrators are responsible for handling user accounts, overlooking the data contribution process, manage the meta-data vocabulary and tagging data assets contributed to the platform. All metadata records and user information are stored in a relational database. For large file management, each organization may setup its own repository and DSFP may access the mzML files remotely.</p> <p>The DMS installation is accompanied by a search index (API), that keeps track of the latest data assets contributed to the platform. This Search API allows the multi-faceted retrieval of information, according to predefined metadata parameters. The underlying index will be fed automatically by the DMS every time a new data asset is made public. The Search API will be based on popular data indexing mechanisms, such as Elasticsearch or Apache SOLR.</p> <p>Two major front-end components are envisioned: a Discovery application and an Open Data Catalogue. The Open Data Catalogue will serve as the FAIR data showcase of the repository, listing meaningful collections of data that administrators have cherry-picked along with data contributors. The Data Catalogue will feature comprehensive metadata records per data collection/dataset, including geotemporal descriptors, responsible organization information, keywords and tags. It will also include data in downloadable form, but also retrievable through open APIs that will allow the interconnection of the repository with external systems. The Open Data Catalogue will be based on popular data cataloguing software (e.g. CKAN).</p> <p>The Discovery application will be refactored using state-of-the-art web frameworks and UI libraries, <i>i.e.</i> ReactJS and Material UI. It will read in data from the DMS Search API and will communicate with R functions for additional data analysis and complex export functionality. The described building blocks of the upscaled DSFP are presented in Figure 2. DSFP will facilitate fast publishing, effective management and open access of research assets.</p>



**Figure 2.** Technical solution for DSFP

The technical solution has been designed in 2021; the DMS and relational database were set up and the administration and institutional forms have been created (<https://dev.norman-data.eu/>). The Search API is under construction and it is planned to be finalized together with the open data catalogue and discovery application until June 2022. The project will be tested by beta-testers (UFZ and NILU already expressed their interest); afterwards two revision rounds will follow. After the finalization, an **online webinar** will be organized to disseminate the progress and support laboratories in sharing their data.

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

- NORMAN NDS
- Cross-Working Group Activity on Non-target Screening (NTS)
- WG 1 Prioritisation
- PARC initiative
- National and EU-funded projects

<b>Participants</b>	EI, UoA, UFZ, NILU, <b>all NTS members</b> of NORMAN and any other interested members.
<b>Proposed in-kind contribution</b>	Working hours for implementation the project (EI/UoA). Organization of webinars.
<b>Contribution needed from NORMAN Association<sup>1</sup></b>	Total funds required for finalizing DSFP upscale - Phase II: 14,000 € that corresponds to working hours for programming and implementation of the Technical solution (cf. above).  Remaining budget of DSFP upscale (– Phase I) 9,000 € will be spent in 2022.

<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.)