



Matéo IORI
IG5 Promotion 2024-2025
Département Informatique et gestion

LIRMM
Montpellier

Stage de fin d'études
Département Informatique et gestion
Effectué du 1 avril 2024 au 29 août 2025

Stage de recherche

Professeur Polytech

Christophe FIORIO

Tuteur entreprise

Arnaud SALLABERRY

Professeur

Table des matières

1. Glossaire.....	4
2. Remerciements.....	5
3. Introduction.....	6
4. Analyse du contexte et des enjeux	7
4.1 Présentation du LIRMM.....	7
4.1.1 Histoire.....	7
4.1.2 Présentation.....	7
4.1.3 Rayonnement.....	8
4.1.4 Les principales activités.....	8
4.2 Présentation de l'équipe ADVANSE.....	9
4.3 Le projet DAE-PFAS.....	10
4.3.1 Historique du projet.....	10
4.3.2 Contexte du projet	10
4.3.3 Naissance du projet DAE-PFAS.....	10
4.3.4 Mon implication dans la vie de ce projet	11
4.4 Ma mission dans le cadre de ce projet	11
4.5 Conditions de travail.....	12
4.5.1 Environnement de travail.....	12
4.5.2 Encadrement.....	12
4.5.3 Outils Utilisés	12
5. Déroulement du stage.....	13
5.1 Développement et refonte du site du PFAS Data Hub (PDH)	13
5.1.1 Analyse de l'existant	13
5.1.2 Reproduction des fonctionnalités.....	16
5.1.3 Améliorations fonctionnelles	18
5.1.4 Améliorations visuelles	21
5.1.5 Enjeux techniques et solutions	22
5.1.6 Résultats obtenus et validation.....	23
5.1.7 Perspectives	24
5.1.8 Analyse critique – Outil de filtrage.....	24
5.2 Phase exploratoire.....	25
5.2.1 Objectifs	25
5.2.2 Méthodologie.....	25
5.2.3 Résultats.....	25
6. Analyse du travail réalisé	26
6.1 Mon apport au laboratoire.....	26
6.2 Compétences développées	26
6.3 Analyse critique	27
7. Conclusions et perspectives.....	28
7.1 Devenir du projet	28
7.2 Devenir de ma contribution	28

7.3	Bilan personnel.....	28
7.4	Évolution de mon projet professionnel.....	29
8.	<i>Webographie</i>	30
9.	<i>Annexes</i>	31

1. Glossaire

UMR

2. Remerciements

Je tiens à exprimer mes sincères remerciements à toutes les personnes qui contribuent à la réussite de mon stage au sein de l'équipe ADVANSE du LIRMM. Leur soutien, leur expertise, leurs encouragements sont précieux et enrichissent grandement mon expérience professionnelle.

Je souhaite tout d'abord remercier mon tuteur sur le projet, M. Alexis GUYOT, pour sa supervision et ses conseils tout au long de mon expérience. Son envie de partager, son professionnalisme et sa disponibilité, malgré son emploi du temps chargé, sont d'une grande aide dans la réalisation de ce stage. Son implication et son sérieux seront une inspiration pour mon projet professionnel.

Je tiens également à exprimer ma reconnaissance envers mon tuteur dans l'équipe ADVANSE M. Arnaud SALLABERRY, permanent de l'équipe. Sa confiance et ses propositions sont des moteurs essentiels dans la réussite des projets.

Mes remerciements vont également à toute l'équipe du département pour leur accueil chaleureux. M. Vincent RAVENEAU et M. Pascal PONCELET contribuent au bon déroulement de ce stage par leur aide et leur gentillesse, en me donnant des conseils et en me proposant des sujets de thèses susceptibles de m'intéresser. Leur esprit d'équipe et leur bienveillance créent un environnement de travail convivial dans lequel je me sent à l'aise.

Je souhaite également exprimer ma gratitude envers Mme. Nadine JACQUET et Mme. Françoise FOURCADIER du département administratif pour leur aide dans les démarches complexes et la planification des déplacements.

Je remercie l'ensemble du personnel de l'entreprise pour leur amabilité et leur coopération. Leur accueil chaleureux et leur ouverture d'esprit rendent mon stage agréable et enrichissant.

Enfin, je tiens à remercier mon professeur Polytech M. Christophe FIORIO ainsi que ma famille pour la relecture de ce rapport.

3. Introduction

Mon stage de cinquième année se déroule au Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM) situé au nord de la ville. Ce laboratoire est reconnu pour la diversité de ses activités de recherches. Il prône des valeurs d'éthique, de durabilité et d'inclusivité à travers différentes initiatives telles que la [science ouverte](#), la transition écologique ([Green LIRMM](#)) et la promotion de [l'égalité et de la parité](#).

Mon objectif était de découvrir le monde de la recherche, et le LIRMM, par sa proximité, ses valeurs et le fait que de nombreux professeurs de Polytech Montpellier y sont chercheurs, s'imposait comme un choix naturel. J'ai été accueilli dans un environnement stimulant et collaboratif, en accord avec mes intérêts pour la science des données.

J'ai eu la chance d'intégrer l'équipe qui m'intéressait après une candidature spontanée et un entretien en présentiel. Cette expérience se révèle extrêmement formatrice et influencera pour sur mes choix professionnels futurs.

Cette équipe, c'est l'équipe ADVANSE (ADVanced Analytics for data SciencE), elle est spécialisée dans l'analyse de données complexes. Placé sous la supervision d'Alexis Guyot (post-doctorant) et d'Arnaud Sallaberry (enseignant-chercheur permanent), j'ai pu contribuer à un projet qui vise à développer des outils d'analyse et de visualisation pour mieux comprendre la pollution environnementale liée aux substances per- et polyfluoroalkylées (PFAS).

Mes travaux ont débuté par la refonte et l'amélioration d'un outil existant permettant le filtrage et la visualisation de données liées aux PFAS. Le projet en cours consiste à utiliser le Machine Learning pour mieux comprendre les contaminations à ces polluants en Europe.

Ce rapport a pour objectif de fournir une vision globale des missions que j'ai menées, de réfléchir sur les résultats obtenus, les choix méthodologiques effectués, ainsi que d'évaluer les compétences sollicitées et développées. Il s'agira également de mettre en lumière les aspects perfectibles et les axes d'amélioration identifiés au cours du stage.

Il est structuré en plusieurs sections : une analyse du contexte avec une présentation du laboratoire, de l'équipe et des enjeux du stage ; un développement détaillé des missions effectuées, suivi d'une analyse critique. En conclusion, je dresserai un bilan personnel de cette expérience, avec des perspectives d'évolution professionnelle. Cette structure permettra de couvrir l'ensemble des aspects significatifs de mon expérience de stage, tout en répondant aux exigences académiques du rapport.

4. Analyse du contexte et des enjeux

Dans cette section, nous allons explorer le contexte ainsi que les enjeux de mon expérience au sein du laboratoire. Nous commencerons par une présentation détaillée du LIRMM, en mettant en lumière son histoire son organisation et son rayonnement. Ensuite, nous nous concentrerons sur l'équipe ADVANSE au sein de laquelle j'ai évolué. Puis nous présenterons le projet auquel j'ai contribué et nous verrons mon rôle dans celui-ci. Enfin, nous analyserons mes conditions de travail, en abordant l'environnement dans lequel j'ai évolué, l'encadrement mis en place et les outils utilisés.

4.1 Présentation du LIRMM

4.1.1 Histoire

Le laboratoire a été créé en 1992, résultant de la fusion de deux entités du CNRS présentes sur le site de Montpellier : le CRIM (Centre de Recherche en Informatique de Montpellier) et le LAMM (Laboratoire d'Automatique et de Microélectronique de Montpellier). Cette initiative visait à regrouper et stimuler les synergies entre les chercheurs en informatique, microélectronique et robotique, sous une seule structure pluridisciplinaire au service de la recherche et de l'innovation. Dès ses débuts, le laboratoire réunit environ 200 membres, chiffre qui doublera en 30 ans pour atteindre aujourd'hui plus de 400 personnes, dont plus de 160 chercheurs et enseignants-chercheurs, une quarantaine de personnel administratif et technique et plus de 130 doctorants et post doctorants.

4.1.2 Présentation

Situé sur le campus Saint-Priest à Montpellier, Le LIRMM est dirigé depuis 2023 par Mme. Marianne HUCHARD, Professeur en informatique. C'est une **unité mixte de recherche** placée sous la double tutelle de l'Université de Montpellier (UM) et du Centre National de la Recherche Scientifique (CNRS). Il bénéficie également de partenariats avec l'Université Paul-Valéry Montpellier 3 (UPVM), l'Université de Perpignan Via Domitia (UPVD) et l'Institut National de Recherche en Informatique et en Automatique (INRIA). Le laboratoire aspire à être à la pointe de l'innovation dans ses trois domaines, comme le suggère Philippe POIGNET, ancien directeur :

“L’objectif du LIRMM est de créer et d’innover dans le domaine de l’informatique, de la robotique et de la microélectronique. De concevoir les nouveaux algorithmes en intelligence artificielle et en analyse de données. De développer de nouveaux logiciels, concevoir de nouveaux robots et algorithmes pour les piloter. Également concevoir de nouveaux circuits intégrés et systèmes embarqués de demain.” - Sur le site officiel, consulté en juin 2025.

Fort de ses 30 ans d’existence, le LIRMM s’est imposé comme un acteur majeur de la recherche scientifique en France et en Europe, notamment grâce à sa capacité à mener des projets interdisciplinaires et à transférer ses innovations vers le monde socio-économique.

4.1.3 Rayonnement

Le laboratoire se distingue aujourd’hui par la création de start-ups issues de ses travaux, telles que [AcuSurgical](#), spécialisée dans la robotique chirurgicale pour les opérations de la rétine, ou [Algodore](#) – développée par le directeur de Polytech, M. Lionel TORRES – qui propose des solutions innovantes de traçabilité pour sécuriser les circuits électroniques. Le laboratoire rayonne aussi à l’internationale par ses nombreux projets de recherche financés à l’échelle [européenne](#), tels que “[Trustworthy AI for CCAM](#)” (intelligence artificielle de confiance pour les véhicules sans conducteur) et “[GUARDEN](#)” (Préserver la biodiversité et les services écosystémiques essentiels dans tous les secteurs et à toutes les échelles), tous deux soutenus dans le cadre du programme [Horizon Europe](#). Ces sont des exemples d’excellence scientifique reconnue, avec des publications dans des conférences et revues internationales, attestant de la portée technologique incarnée au LIRMM.

4.1.4 Les principales activités

Le laboratoire est logiquement structuré en trois départements scientifiques : Informatique, Robotique et Microélectronique, couvrant un large spectre de thématiques allant des fondements théoriques aux applications industrielles.

- Les recherches du [département Robotique](#) se concentrent sur la robotique médicale et de manipulation, la robotique sous-marine, la perception, l’interaction physique homme-robot ou encore la robotique humanoïde. Ces travaux trouvent des applications directes dans le monde de la médecine, de l’environnement, de l’industrie et aussi du quotidien humain.
- Le [département Microélectronique](#) vise à concevoir des systèmes électroniques intégrés toujours plus intelligents en mettant l’accent sur l’efficacité énergétique, la sécurité et la fiabilité. Ses recherches s’inscrivent dans le champ des objets communicants appliqués à l’environnement, y compris les environnements complexes (spatial, radiatif, haute température, vivant).

- Le département informatique couvre un large éventail de recherches, fondamentales et appliquées. Ses travaux portent sur les fondements du calcul, la science des données et l'intelligence artificielle, la bioinformatique, le génie logiciel et l'analyse d'image. Ces recherches entrent dans des domaines extrêmement variés et les domaines d'applications sont infinis. Le département entretient des partenariats avec des hôpitaux, start-ups, entreprises régionales ou grands groupes. Mon responsable « entreprise », M. Arnaud SALLABERRY est le responsable adjoint de ce département.

Le département d'Informatique est composé de 15 équipes, celle qui m'a particulièrement intéressé et où j'effectue mon stage est l'équipe ADVANSE. Nous la présentons dans la partie suivante.

4.2 Présentation de l'équipe ADVANSE

L'équipe ADvanced Analytics for data SciencE (ADVANSE) est l'une des équipes de recherche du département d'Informatique du LIRMM. Ses travaux s'inscrivent dans le domaine de l'analyse des grandes bases de données, avec pour objectif l'extraction de nouvelles connaissances à partir de données complexes et hétérogènes. Les activités de recherche de l'équipe ADVANSE se déclinent en trois axes principaux :

- Exploration de données (Data Mining) : développement de méthodes pour l'extraction de savoir, de connaissance, de motifs répétitifs à partir de grandes quantités de données.
- Visualisation analytique : conception d'interfaces visuelles facilitant le raisonnement analytique, permettant aux utilisateurs d'explorer et d'interpréter des données complexes facilement pour soutenir la prise de décision.
- Apprentissage automatique (Machine Learning) : utilisation et élaboration de modèles d'intelligence artificielle pour donner aux ordinateurs la capacité d'apprendre à partir des données. L'accent est mis sur l'explicabilité et l'architecture des modèles, qui sont des enjeux de recherche dans l'ère du temps.

L'équipe ADVANSE adopte une approche interdisciplinaire, combinant des travaux théoriques et expérimentaux, et collabore avec divers partenaires académiques et industriels. Elle est dirigée par le Professeur Pascal PONCELET, avec pour adjointe la Professeure Sandra BRINGAY.

Je vais à présent présenter le projet auquel j'ai activement contribué.

4.3 Le projet DAE-PFAS

Le projet DAE-PFAS, au cœur de mon stage, s'inscrit dans une dynamique de recherche interdisciplinaire mêlant science des données et enjeux environnementaux liés aux PFAS. Une annexe présentant ces composés chimiques et leurs impacts est fournie en fin de rapport [[lien](#)]. Sa lecture est recommandée avant de poursuivre, afin de mieux comprendre le contexte de la suite du rapport.

4.3.1 Historique du projet

Le *Forever Pollution Project (FPP)* est une initiative journalistique européenne menée par le journal *Le Monde* publié en 2023. Ce projet représente un travail colossal de fouille et d'homogénéisation des données. Il a permis de cartographier la contamination aux PFAS à travers l'Europe en regroupant des milliers de données gouvernementales, industrielles et sanitaires. À partir de ces travaux, le CNRS a initié le projet PFAS Data Hub (PDH), qui a pour but de continuer à faire vivre ces données en les rendant accessible et en les alimentant régulièrement. Ce projet regroupe des data journalistes, des géologues et des chimistes, mais pas de vrai développeur ou analyste de données.

Mes travaux sont entièrement basés sur cette base de données, nous y ferons par la suite référence par les termes « base de données du PDH » ou « données du PDH ».

4.3.2 Contexte du projet

MaDICS (Masses de Données, Informations et Connaissances en Science) est un Groupement de Recherche (GDR) du CNRS créé en 2015, visant à promouvoir les recherches interdisciplinaires en Sciences des Données. Il sert de forum d'échanges et de soutien pour les chercheurs et les acteurs externes (industriels, médias, culturels) confrontés aux enjeux du Big Data. Les activités de MaDICS se structurent autour d'Actions, qui durent entre deux et quatre ans.

Mon stage s'inscrit dans le cadre de l'action MaDICS nommée « Détections d'Anomalies Environnementales » (DAE) portée par Mme. Lylia ABROUK. C'est au sein de cette action qu'a vu le jour le projet DAE-PFAS visant la détection d'anomalies environnementales liées aux PFAS.

4.3.3 Naissance du projet DAE-PFAS

C'est en découvrant le projet PDH du CNRS par l'intermédiaire du chimiste M. Pierre LABADIE, impliqué dans celui-ci, que Mme. Lylia ABROUK a mis en place un projet de recherche interdisciplinaire à fort potentiel : DAE-PFAS. Ce projet est une extension directe

du PDH et lui ajoute une dimension, en le mettant en relation avec des experts en analyses de données. Les ambitions sont maintenant plus grandes, DAE-PFAS vise à exploiter les données du PDH pour faire des analyses sur la contamination liée aux PFAS, en croisant des approches en data science, visualisation et intelligence artificielle. Des outils facilitant l'exploration, l'analyse et la visualisation de ces informations sont donc en cours de développement.

Porté par le CNRS (via le MITI) et par l'Université de Montpellier dans le cadre du projet ExposUM, DAE-PFAS devrait évoluer dans les prochaines années en un ensemble de sous-projets ciblant différents volets (environnement, santé...).

4.3.4 Mon implication dans la vie de ce projet

J'ai eu la chance et l'honneur de présenter mes travaux lors d'une session du GDR MaDICS à Toulouse fin mai. Mon travail a également été présenté par mon tuteur à la conférence INFORSID à Pau en juin, où il a été très bien accueilli, et le sera à nouveau en juillet à PFIA à Dijon.

4.4 Ma mission dans le cadre de ce projet

Ma mission dans ce projet s'inscrit au croisement entre développement logiciel et recherche. Dans un premier temps, j'ai été chargé de concevoir une nouvelle interface web en m'inspirant du site du [PDH](#) avec l'objectif de proposer un outil plus parlant visuellement et plus adapté aux besoins des chercheurs. Ce travail a abouti à une version permettant des filtrages plus poussés et une visualisation plus claire des données. Ce travail est révolu et sera détaillé dans la partie X.X

Dans un second temps, (en cours) ma mission s'oriente vers un axe plus exploratoire du projet : l'analyse des données PFAS via des méthodes de clustering. L'objectif est de mieux comprendre les sites contaminés, en comprenant leurs origines, leurs causes et en dressant des profils de contaminations. Cela implique également la conception de visualisations claires permettant aux experts d'interpréter les résultats générés par nos algorithmes facilement, sans avoir besoin de comprendre la logique intrinsèque de ces mêmes algorithmes. Ce travail est en cours et se développera sur le restant de la durée de mon stage.

Mon travail est resté aligné sur les axes prévus au début du projet : d'une part l'amélioration des outils de sélection des données, et d'autre part l'intégration d'outils d'analyse exploratoire et explicative, dans le but de produire à terme une contribution scientifique valorisable.

4.5 Conditions de travail

Nous allons maintenant passer au détail de mes conditions de travail. Je vais commencer par parler de l'environnement de travail, avant de décrire l'encadrement mis en place puis de finir sur les outils que j'ai beaucoup utilisé lors de ce stage.

4.5.1 Environnement de travail

Mon stage s'est déroulé en présentiel dans le bâtiment 5 du campus Saint-Priest.

Ce campus est un pôle technologique qui a été conçu pour favoriser les synergies entre recherche académique, innovation industrielle et formation. Il regroupe aujourd'hui une grande partie des acteurs montpelliérains de la recherche en sciences et technologies du numérique.

Le bâtiment 5 accueille la quasi-totalité de l'IES (environ 200 personnes), une partie du LIRMM (environ 100 personnes), une antenne du centre INRIA Sophia Antipolis (environ 100 personnes), ainsi que des espaces réservés aux start-ups et à la recherche partenariale avec les entreprises locales.

C'est un environnement propice à la rencontre entre disciplines scientifiques, ce qui favorise les travaux interdisciplinaires — un aspect essentiel pour faire progresser la recherche.

4.5.2 Encadrement

L'encadrement durant mon stage s'est inscrit dans une dynamique d'autonomie. Les membres permanents de l'équipe alternant entre enseignement et recherche, leur présence au laboratoire est variable. Mon référent, très sollicité en début de stage, a été régulièrement en déplacement. J'ai dû adopter une démarche proactive : aller chercher l'information auprès des membres de l'équipe plutôt que d'attendre qu'elle me soit apportée. Un suivi hebdomadaire a été mis en place chaque vendredi pour faire le point sur l'avancement des travaux en cours et la prévision des prochains.

4.5.3 Outils Utilisés

J'ai utilisé GitLab pour le versionnement du code, Visual Studio Code comme éditeur, et Discord pour les échanges informels. Pour l'aspect recherche, Google Scholar m'a permis d'identifier les travaux académiques pertinents, et Zotero m'a servi à structurer et gérer la bibliographie de manière rigoureuse.

Il est maintenant temps de rentrer dans le cœur de mon travail, la prochaine partie va détailler les tâches que j'ai réalisé lors de ce stage.

5. Déroulement du stage

Pour structurer l'analyse de mes travaux au cours du stage, je vais commencer par décrire le développement du premier outil. Ensuite, je parlerai de la seconde phase, qui est bien plus exploratoire. Les éléments présentés lors de cette seconde partie seront plus abstraits que ceux présentés dans la première.

5.1 Développement et refonte du site du PFAS Data Hub (PDH)

Durant les 2 premiers mois de mon stage, mon travail a consisté à améliorer un outil existant. Je vais le présenter, identifier ses limites et détailler les améliorations implémentées. Je parlerai des problèmes rencontrés et des solutions mises en place avant de présenter les résultats et analyser mon travail avec un regard critique.

5.1.1 Analyse de l'existant

Le site du PDH constitue aujourd'hui la seule plateforme publique donnant accès à des données relatives à la contamination liée aux PFAS en Europe. Mis en place et maintenu par Luc Martinon, data journaliste. Son but est d'offrir aux chercheurs (notamment chimistes et environnementalistes) un outil leur permettant de visualiser et d'accéder aux données du PDH. Voici un scénario d'utilisation basique pour un chercheur :

1. L'utilisateur filtre les données qu'ils l'intéressent.
2. Pour visualiser l'information l'utilisateur à sa disposition une carte qui se met à jour en direct et une fonctionnalité de clic pour analyser les données d'un point.
3. L'utilisateur exporte les données filtrées pour effectuer des analyses complémentaires.

Ci-dessous une capture d'écran de l'interface de l'outil.

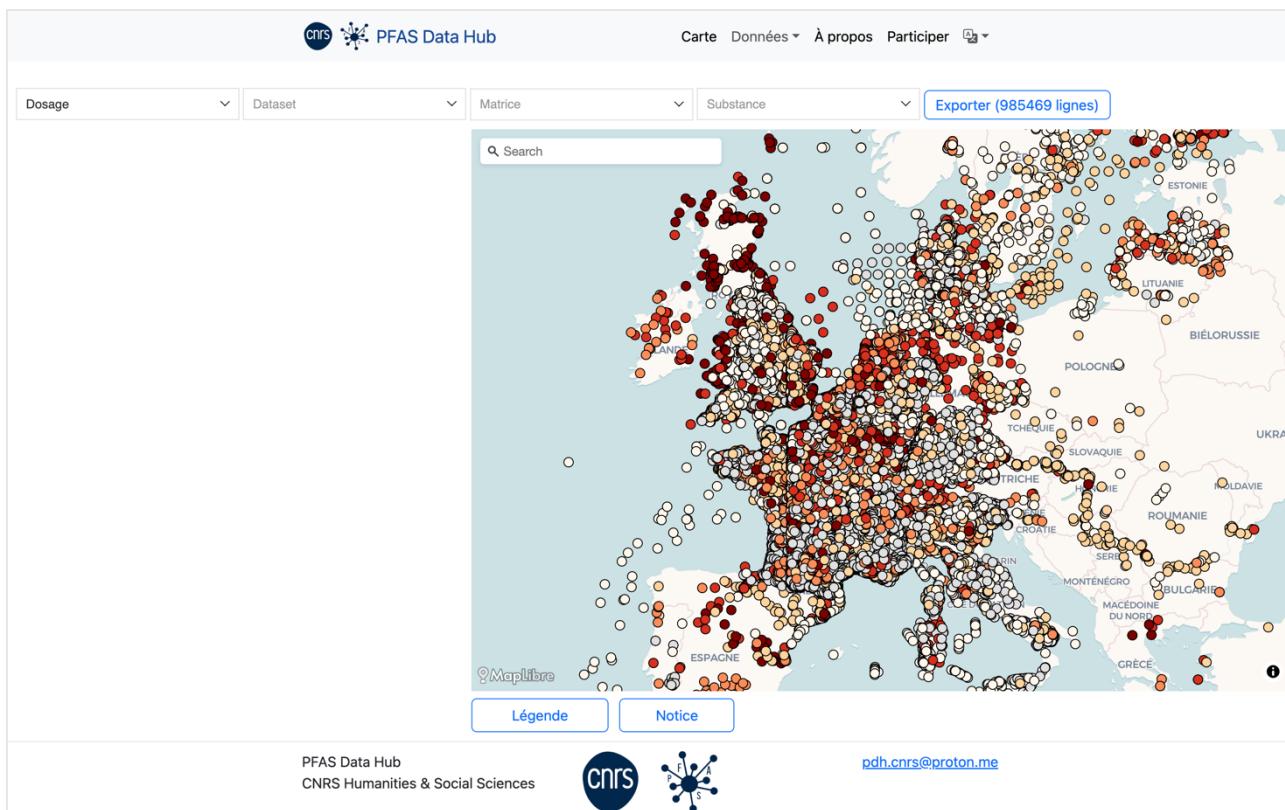


Figure X : Capture d'écran de l'interface du PDH

Les fonctionnalités principales de cet outil, son architecture technique et ses limites sont détaillés dans les prochaines sous-parties.

5.1.1.1 Fonctionnalités principales

Le filtrage constitue une fonctionnalité essentielle pour tout outil visant à afficher un volume important de données. L'interface du PDH, présentée ci-dessus, offre à l'utilisateur la possibilité de filtrer les données selon plusieurs critères :

- Catégorie : Les données peuvent représenter des sites de production de PFAS, des sites qui utilisent des PFAS, des sites de contamination présumée, ou des mesures effectuées (avec des mesures de concentrations de polluant).
- Dataset : Les données ont été regroupées et homogénéisées dans un outil unique par les membres du projet mais viennent de plusieurs sources, et donc plusieurs 'datasets'.
- Matrice : Nature du lieu de mesure (eau, sol, air...).
- Substance : La famille des PFAS est très large, le site du PDH propose de filtrer par substances (les substances sont des composés chimiques de la famille des PFAS).

Un autre élément important est d'accéder à la donnée. Lors d'un clic sur un point, l'interface du PDH propose un panneau latéral affichant les données complètes :

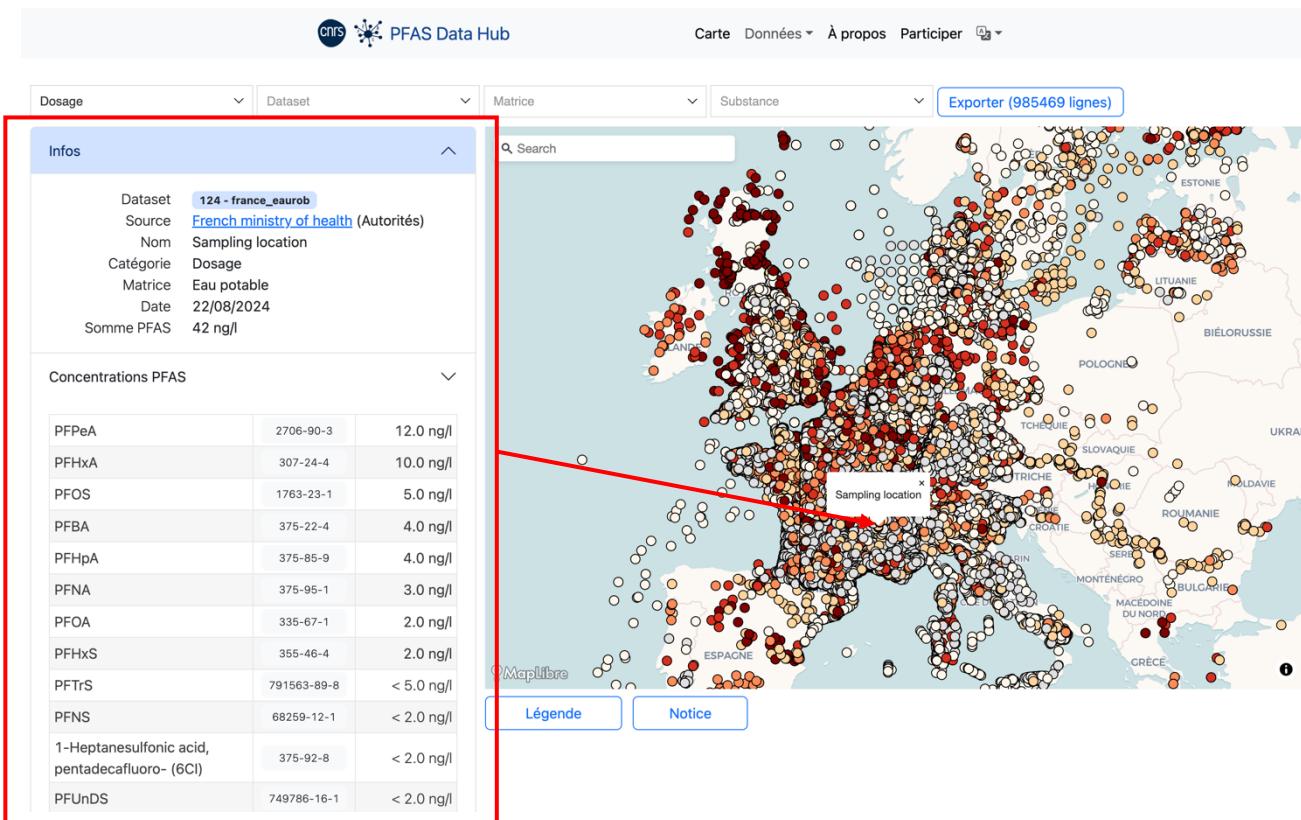


Figure X : Panneau latéral avec le détail des informations du point sélectionné

Il est possible sur la gauche de la figure X de voir les mesures de concentrations réalisées pour le point sélectionné ainsi que d'autres informations.

Les données peuvent être exportées au format CSV.

5.1.1.2 Architecture

Ce projet repose sur une architecture multi services orchestrée via Docker Compose avec les éléments suivants :

Frontend : C'est un site statique principalement composé de pages Markdown. La page carte interactive constitue l'élément le plus dynamique, développée en HTML / JavaScript pur. La carte en elle-même est développée à partir de MapLibre.

Backend/API : FastAPI et DuckDB sont utilisés pour accéder aux données efficacement.

5.1.1.3 Limites

Ce site est le premier à mettre à disposition ce jeu de données du PDH, à été développé et est maintenu par une seule personne : Luc Martinon – qui n'est même pas développeur –. Il est normal qu'il contienne certaines limites.

Premièrement, les fonctionnalités de filtrage sont incomplètes :

- Il est impossible de filtrer par zone géographique ainsi que par date de mesure.
- Les substances sont indépendantes de l'unité dans laquelle elles ont été mesurées. Par exemple, si je sélectionne la substance PFOA – un des PFAS les plus répandu –, mes résultats vont contenir des mesures en ng/L (valeurs entre 0 et 40) comme en ng/Kg (valeurs entre 10 et 5000). Cela fausse les échelles et peut engendrer des erreurs dans les analyses

De plus, avec plus de 1 million de points à afficher, la visualisation du PDH n'est pas claire. Les points se superposent, il est difficile d'identifier des tendances / clusters. La légende de couleur est logarithmique et manque de précision / nuance.

Il est important de noter que pour ces causes, M. Martinon reçoit des demandes de filtrage personnelles par certains chercheurs – qui veulent par exemple isoler une région de l'Europe dans leurs analyses –. Il traite ces demandes au cas par cas et pourrait gagner du temps si l'outil du PDH était plus performant.

Dans les prochaines parties je vais détailler le travail réalisé pour reproduire les fonctionnalités dans un nouvel outil en repartant de zéro, puis présenter les améliorations mises en place.

5.1.2 Reproduction des fonctionnalités

Dans le but d'avoir une version propre au LIRMM et facilement maintenable par les membres de l'équipe, il m'a été demandé de développer l'outil en repartant de zéro en utilisant les technologies suivantes :

- Svelte comme framework pour le frontend.
- PostgreSQL pour la base de données.
- SvelteKit (intégré naturellement dans Svelte) pour la liaison du frontend à la base de données.
- Leaflet pour la cartographie car très simple à mettre en place pour des résultats rapides.

Le système de filtrage repose sur une rune svelte centrale qui contient l'état des filtres. Dans Svelte, une rune est une variable réactive, qui fonctionne un peu comme une "valeur intelligente" : dès qu'on la modifie, tous les éléments qui en dépendent se mettent automatiquement à jour. Elle est basée sur le design pattern [X] observable [X], et permet de propager facilement les changements dans l'interface. Toute modification de filtre dans l'application change l'état de cette rune et entraîne une nouvelle récupération de données. Cette approche utilise l'approche réactive native et très efficace de Svelte.

Une fonction réagit aux changements dans cette rune et lance la récupération des nouveaux points en accord avec les nouveaux filtres. Elle renvoie un objet "response" qui peut avoir 3 états interprétés différemment par l'application :

- Loading : dans ce cas, on affiche un écran de chargement.
- Erreur : on affiche un message d'erreur.
- Data : contient la donnée voulue, on affiche les résultats.

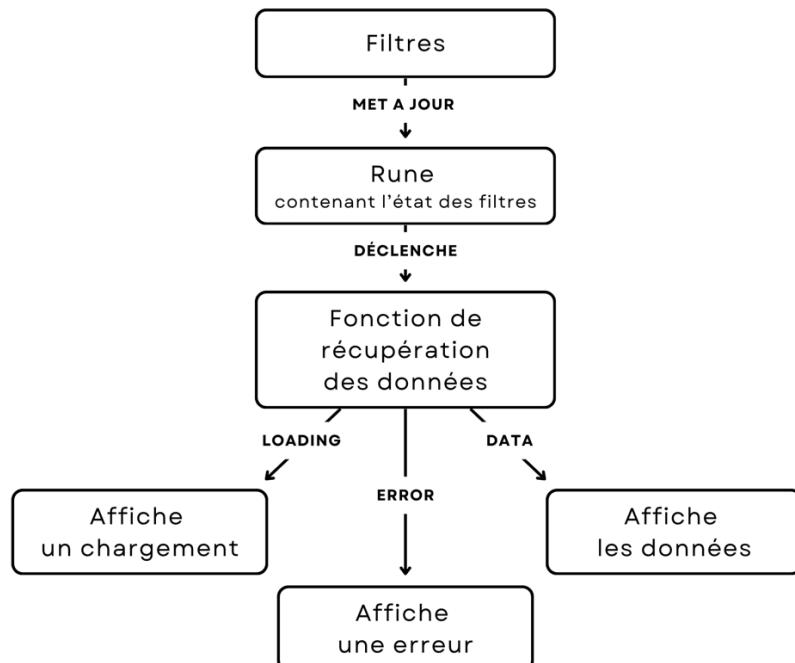


Figure X : Schéma détaillant le fonctionnement du filtrage

Passons maintenant à l'analyse des améliorations apportées, une première partie détaille les améliorations fonctionnelles, une seconde les améliorations visuelles.

5.1.3 Améliorations fonctionnelles

5.1.3.1 Filtrage avancé

- Filtrage par zone géographique

Nous avons mis en place un système de filtrage par zone géographique. L'utilisateur peut ainsi sélectionner un pays, une région ou un département d'un simple clic sur une carte interactive. Il est également possible de définir une zone personnalisée à l'aide d'un outil de sélection libre (lasso), permettant d'effectuer des analyses ciblées sur une ville ou sur une zone ne correspondant à aucun découpage administratif standard.

Area selection



Area selection

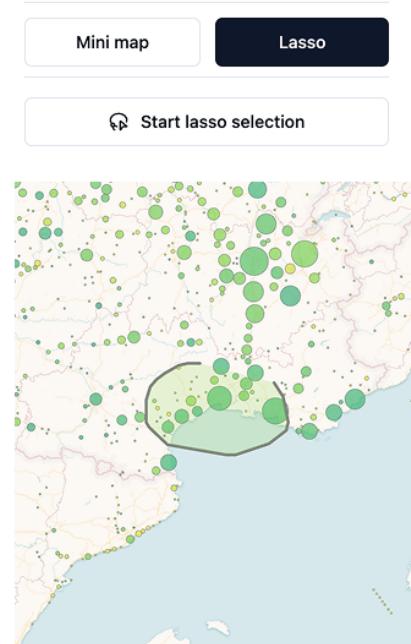


Figure X : Détail du composant de filtrage par zone géographique.

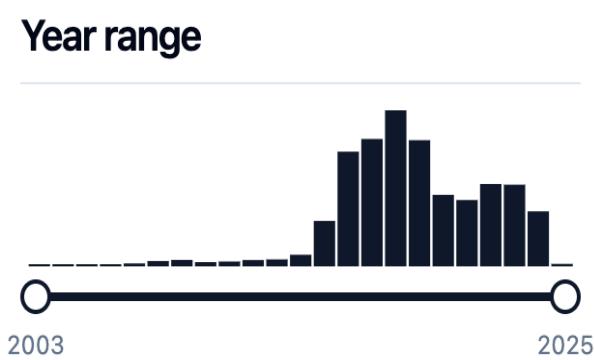
- Filtrage par substance

Un autre filtre essentiel mis en place concerne les substances. Pour chaque substance sélectionnée, l'utilisateur doit également choisir une unité, formant ainsi un couple cohérent substance/unité — ce qui permet d'éviter les incohérences évoquées en partie X. Un mode optionnel permet d'exclure les valeurs aberrantes (« outliers »), en appliquant la méthode des écarts interquartiles. Enfin, un curseur (slider) permet de filtrer les concentrations associées à chaque couple, offrant un contrôle précis sur la plage de valeurs à visualiser.



Figure X : Détail du composant de filtrage par substance.

- Filtrage par année de mesure



Il est aussi possible de filtrer par année de mesure, un détail qui n'est pas anodin puisque certaines données trop anciennes peuvent être considérées comme non fiables avec les évolutions des appareils de mesure.

Figure X (a gauche) : Détail du composant de filtrage par plage d'années.

Bien sûr, il est toujours possible de filtrer par matrices, dataset et catégorie comme dans la version initiale.

- Clic

Lorsqu'un utilisateur clique sur un point, nous avons repensé l'affichage des informations pour le rendre plus pertinent. Comme la plupart des points correspondent à des stations de mesure, plusieurs prélèvements peuvent avoir été effectués au même endroit. Nous avons donc choisi de regrouper les résultats par substance, sous forme de répartition (min, max, moyenne, quartiles...).

Cependant, cette approche présente une limite importante : elle ne restitue pas l'évolution temporelle des concentrations, pourtant essentielle dans le cadre du suivi environnemental. En effet, l'une des principales fonctions d'une station est de mesurer les variations dans le temps. Une amélioration envisagée — à intégrer dans les prochaines phases du projet — serait d'ajouter une visualisation temporelle. L'idée serait d'afficher une petite courbe pour chaque ligne permettant d'apercevoir rapidement la tendance de l'évolution de la concentration.

Details										X
9 points										
Substance	Unit	Detections	Mean	Min	Q1	Median	Q3	Max		
PFBA	ng/l	6/9	97.00	28.00	35.75	50.00	87.50	323.00		
PFBS	ng/l	8/9	21.13	6.00	8.75	12.00	22.75	65.00		
PFHxA	ng/l	8/9	11.25	6.00	7.50	11.50	13.25	20.00		
PFOA	ng/l	9/9	12.56	5.00	6.00	11.00	16.00	30.00		
PFOS	ng/l	8/9	2.06	0.88	1.24	1.90	2.39	4.34		
PFPeA	ng/l	5/9	8.80	5.00	6.00	6.00	9.00	18.00		

Figure X : Détail du panneau affichant les détails d'une station de mesure.

5.1.3.3 Autres améliorations

La logique d'export des données a également été repensée. Il est désormais possible de choisir les colonnes à inclure dans l'export CSV, ce qui permet de limiter la surcharge d'informations inutiles et de simplifier les analyses.

Par ailleurs, l'interface offre la possibilité d'enregistrer directement la carte affichée au format PNG, sans avoir recours à une capture d'écran — une fonctionnalité particulièrement utile pour illustrer des résultats ou réaliser des comparaisons visuelles. Enfin, d'autres améliorations ont été apportées à la navigation sur la carte : centrer la vue sur les données actuellement filtrées, ou zoomer automatiquement sur un cluster en cliquant dessus en sont des exemples.

Ces nouvelles fonctionnalités répondent directement aux limites identifiées et aux besoins exprimés par les utilisateurs du PDH – notamment les demandes récurrentes reçues par M. Luc Martinon.

5.1.4 Améliorations visuelles

Nous avons également ajouté une option de *clustering* des points, reposant sur l'algorithme F-SAC (*Fast Spatial Agglomerative Clustering*), développé au sein du LIRMM. Cette méthode permet d'agrérer les points proches pour alléger l'affichage, tout en conservant une information pertinente, rendant la carte lisible même en présence de plusieurs dizaines voire centaines de milliers de points.

Par ailleurs, la palette de couleurs a été entièrement repensée pour améliorer la perception visuelle des valeurs. Nous avons adopté la palette Viridis, conçue pour être perceptible par les personnes atteintes de daltonisme, tout en assurant une bonne lisibilité en niveaux de gris. Ce choix est soutenu par les recommandations de l'article de xxx (retrouver lien).

Ci-dessous une comparaison avant / après de la visualisation des mesures de PFOA (toutes unités disponibles).

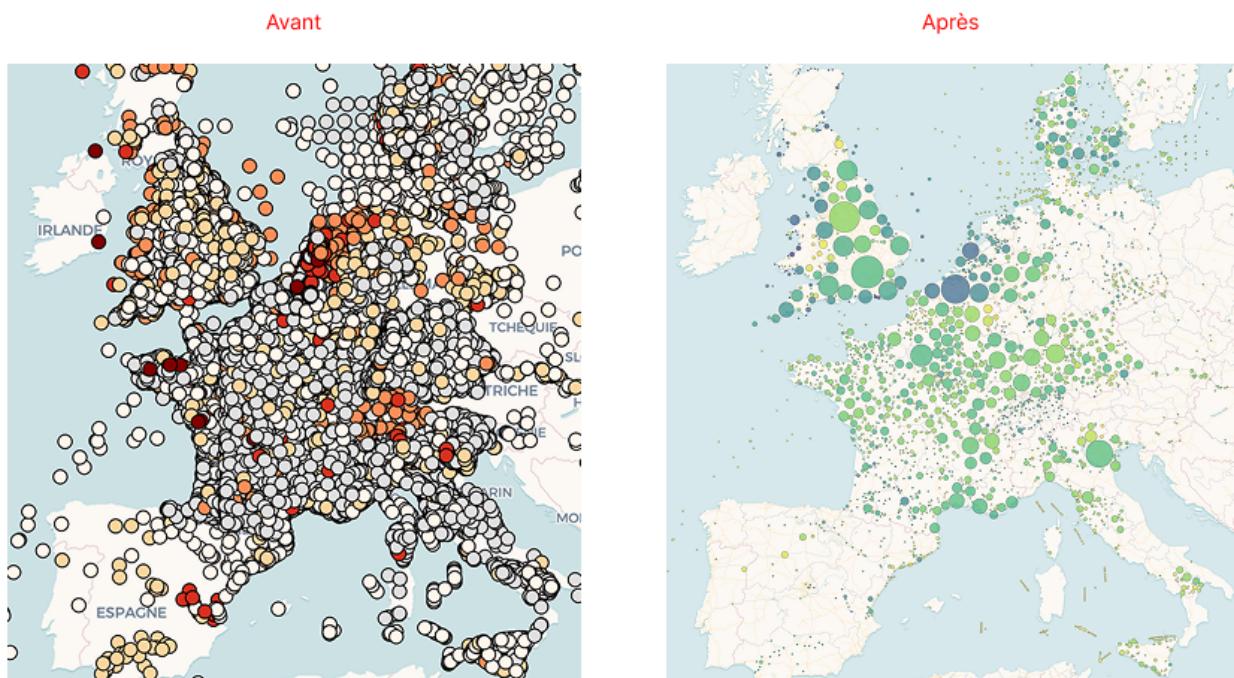


Figure X : Comparaison de la visualisation des mesures liés au PFOA (toutes unités)

Nous pouvons observer une surcharge frappante sur l'image de gauche, des points sont cachés. sur l'image de droite nous pouvons bien plus facilement observer la répartition des mesures sur le territoire.

Pour améliorer encore l'expérience utilisateur, plusieurs fonds de carte alternatifs ont été intégrés, tels qu'OpenStreetMap, des versions avec ou sans labels, ou encore une version topographique, afin de s'adapter aux besoins d'analyse spécifiques.

Enfin, une annexe dédiée présente l'évolution de l'interface utilisateur tout au long du projet.

5.1.5 Enjeux techniques et solutions

Plusieurs défis ont jalonné le développement :

- Taille des données : au-delà de 200 000 points, l'affichage en points classiques devient lourd à supporter dans le navigateur. Pour le mode cluster, c'est le temps d'exécution de l'algorithme F-SAC qui pose problème. Pour pallier cela, on effectue automatiquement une bascule vers un affichage en heatmap – bien moins lourde à afficher – lorsqu'il y a plus de 200 000 points à afficher sur l'écran. Pourquoi ne pas afficher tout le temps une heatmap ? Car on perdrait une dimension, la heatmap affiche uniquement la densité de mesure, et pas les valeurs de concentrations.

- Intégration de F-SAC : l'intégration de cet algorithme performant mais complexe a nécessité un travail approfondi. Il a fallu débattre de la version optimale. (Insérer annexe si assez de temps). [lien]
- Performances base de données : usage de materialized views et d'index spatio-temporels dans PostgreSQL pour accélérer les requêtes. (Insérer annexe si assez de temps). [lien]
- Mise en place d'un ETL (script) pour récupérer les données, les enrichir et les insérer dans la base de données. (Insérer annexe si assez de temps) [lien]

5.1.6 Résultats obtenus et validation

Ci dessous une présentation de l'interface finale du prototype mis en place. (Une annexe avec d'autres prises de vues l'explique plus en détail [pas encore]).

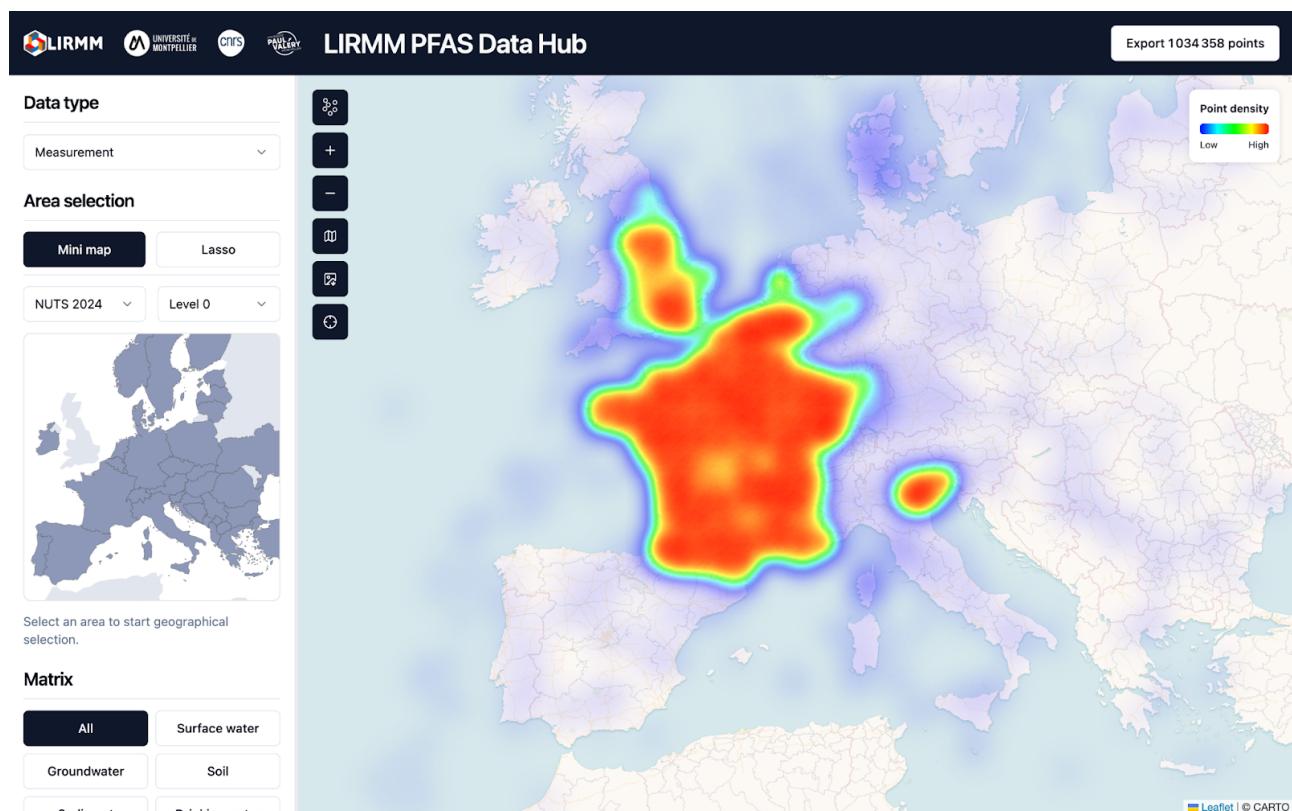


Figure X : Présentation de l'interface

L'interface est moderne, plus intuitive que l'ancienne version et offre aux chercheurs des fonctionnalités plus poussées.

Le prototype a été présenté lors du Symposium (conférence) MaDICS à Toulouse en Mai. Les retours ont été très positifs, en particulier de la part de M. Luc Martinon, qui a exprimé un intérêt pour intégrer ces améliorations dans une future version publique du PDH.

Ce projet n'est pas hébergé, une documentation a été créée pour permettre à tout utilisateur d'installer le code, et de récupérer et formatter les données en lançant un script.

5.1.7 Perspectives

L'interface développée présente un prototype proposant une version améliorée de la version du PDH, à terme, cette version pourrait remplacer l'ancienne. Voici quelques pistes d'améliorations qui pourraient être analysées si tel est le cas :

- Passage à Maplibre (WebGL) pour un affichage plus rapide et capable de supporter davantage de points.
- Hébergement sur serveur hautes performances, pour accélérer les temps de réponse et permettre la consultation à grande échelle.
- Intégration future des résultats de recherche (SHAP, clustering, interpolation) directement dans la plateforme.
- Élargissement du jeu de données, automatisation de leur enrichissement et suivi en temps réel.

5.1.8 Analyse critique - Outil de filtrage

Analyse critique ?

5.2 Phase exploratoire

La seconde phase de mon stage est bien plus exploratoire

5.2.1 Objectifs

Objectifs de cette phase

5.2.2 Méthodologie

Méthodologie adoptée

5.2.2.1 Méthodologie de recherche

5.2.2.1 Application au projet

5.2.3 Résultats

6. Analyse du travail réalisé

6.1 Mon apport au laboratoire

Bien que le projet DAE-PFAS soit un projet collaboratif de long terme, appelé à évoluer avec ou sans ma participation, j'estime avoir bien contribué à son avancement durant ma période de stage. Grâce à mes bases solides en développement travaillées à Polytech, j'ai pu prendre en main les besoins techniques du projet et proposer des solutions efficaces dans un délai court. J'ai développé un outil fonctionnel et documenté. J'ai veillé à appliquer les bonnes pratiques de développement, tant sur le plan de l'organisation du code que de la clarté de l'interface, afin de garantir la maintenabilité de l'outil. Celui-ci pourra servir de socle pour les prochaines phases du projet, notamment lors d'analyses plus poussées... Enfin, les travaux exploratoires menés pendant le stage, constitueront une base de référence utile pour les futurs stagiaires ou chercheurs impliqués. Ils fourniront des premières pistes d'analyse, des prototypes et un cadre technique sur lequel s'appuyer.

6.2 Compétences développées

Au cours de ce stage, j'ai pu renforcer mon autonomie, mes compétences en développement ainsi qu'élargir mes connaissances en recherche scientifique, notamment dans un cadre de travail interdisciplinaire.

Sur le plan technique, j'ai tout d'abord découvert un nouveau framework : Svelte. Cela m'a amené à revisiter la documentation et à approfondir certains concepts fondamentaux du développement web. Cette remise à niveau m'a permis de revoir les bases. Avec l'expérience je commence à vraiment comprendre les enjeux liés aux frameworks web, en termes de performance, d'expérience utilisateur et de maintenabilité du code. J'ai également travaillé en Python sur plusieurs problématiques, j'ai créé des visualisations, des scripts pour modifier des données... Ce langage, que je continuerai à utiliser dans la suite du projet, m'a permis de consolider ma pratique en matière de traitement de données. Par ailleurs, j'ai développé des compétences en Machine Learning, notamment à l'occasion de la session MaDICS à Toulouse fin mai. Cette thématique, qui est centrale à la seconde phase de mon stage, m'a particulièrement intéressé par sa richesse et son potentiel infini.

En parallèle, mon immersion dans un environnement de recherche interdisciplinaire m'a permis d'acquérir des compétences spécifiques à la démarche scientifique. Les réunions hebdomadaires ont été l'occasion de confronter les idées, de discuter des orientations à suivre et de trouver collectivement des solutions à des problématiques complexes. J'ai assisté à des échanges rassemblant des chercheurs aux profils très variés (experts en visualisation, en base de données, ingénieurs ...), ceux-ci m'ont sensibilisé à l'importance de l'écoute, de la pédagogie et de la collaboration dans un projet de recherche. Cette capacité à travailler ensemble, malgré les différences de spécialités, m'a impressionné.

J'ai également appris la « méthodologie de recherche ». On peut retrouver des similitudes dans la manière de penser, de s'organiser quels que soient les problématiques scientifiques abordées. Avoir pratiqué cette méthodologie m'aidera à appréhender avec plus de recul les futures problématiques auxquelles je serai confronté. Faire de la recherche en m'appuyant sur des articles scientifiques m'a conduit à utiliser des outils que je n'avais jamais utilisé en dehors des cours, comme Google Scholar et Zotero.

Enfin, ce stage m'a permis de progresser en autonomie. L'encadrement, bien que toujours disponible en cas de besoin, était relativement souple. Il m'a donc fallu quelques jours pour prendre mes marques, organiser mon travail, poser mes questions au bon moment et gagner en assurance. Cette liberté à été déstabilisante au début mais je m'y suis habitué, j'ai pu développer à terme un rythme de travail efficace.

6.3 Analyse critique

6.3.1 Mentalité

L'intégration au sein de l'équipe n'a pas été immédiate. Dès le premier jour, le travail a commencé sans réel temps dédié à l'accueil ou à la présentation des membres de l'équipe. Cela m'a généré un léger sentiment d'isolement, il me manquait un cadre bien défini. Je me sentais un peu comme lâché dans le grand bain, sans trop savoir comment faire ce qu'on me demandait ni à qui demander de l'aide. Avec le recul, c'était totalement normal. J'ai juste été habitué à quelque chose de plus « doux » avec le stage en entreprise de l'année dernière, ou tout a pris plus de temps. Il y avait un nombre incalculable de protocole à respecter avant de pouvoir commencer à travailler. J'aurais pu faire davantage d'efforts dès le début pour aller vers les autres et m'intégrer plus activement. Ce point constitue une première piste d'amélioration pour mes futures expériences : prendre l'initiative dès le début dans le but d'être acteur pour établir des liens et faciliter la communication au sein de l'équipe.

6.3.1 Technique

J'ai ressenti des lacunes techniques lors de l'implémentation du système de clustering. Le code de base étant déjà écrit, je pensais initialement pouvoir l'adapter rapidement. En pratique, l'adaptation à mon cas d'usage a été complexe : il m'a fallu beaucoup de temps pour comprendre les subtilités de l'algorithme, ajuster les paramètres, et obtenir des résultats pertinents. J'ai eu tendance à me détourner de l'objectif initial, cherchant avant tout à obtenir un résultat "visuel" satisfaisant, au détriment de la finalité scientifique recherchée. Ce constat m'a permis de prendre conscience de l'importance de toujours garder en tête les objectifs du projet, même en cas de difficulté technique.

La seconde phase du stage, axée sur des outils de Machine Learning, je m'attends à devoir surmonter une nouvelle phase d'adaptation. Les bases acquises à Polytech sont correctes pour des projets encadrés, mais le niveau d'exigence scientifique attendu ici est plus élevé, notamment en matière de traitement de données environnementales à grande échelle, de choix de modèles, ou encore d'interprétabilité des résultats. Cela représente un défi stimulant, pour m'adapter je vais devoir m'organiser pour monter en compétence rapidement le moment venu.

La prochaine partie conclue le rapport, elle détaille le devenir de mes travaux, je vais y dresser un bilan et une réflexion sur mon avenir professionnel.

7. Conclusions et perspectives

7.1 Devenir du projet

Le projet DAE-PFAS dispose encore d'environ un an dans sa configuration actuelle. Dans la continuité d'un travail amorcé depuis un an déjà, au cours duquel les PFAS ont été explorés sous différents angles. Ce temps d'exploration a permis de dégager des axes de recherche plus ciblés, révélant la complexité et l'ampleur du sujet. En conséquence, il est prévu que DAE-PFAS évolue en un ensemble de sous-projets thématiques, chacun concentré sur un domaine spécifique, tels que la santé, ou l'environnement par exemple.

7.2 Devenir de ma contribution

Concernant mon travail, le prototype que j'ai développé pourrait à terme être intégré au site du PFAS Data Hub. Cette intégration nécessitera sûrement une phase de validation et d'ajustements, en collaboration avec un expert du domaine dans l'idéal. Le travail effectué dans la seconde phase de mon stage ayant été plus exploratoire, les travaux engagés resteront en suspens à la fin de mon passage. Ces éléments, bien documentés, serviront de point de départ pour un nouveau stagiaire prévu l'année prochaine.

7.3 Bilan personnel

Ce stage au LIRMM m'a offert une expérience enrichissante et formatrice, tant sur le plan technique que personnel. J'ai eu l'opportunité de découvrir le fonctionnement d'un laboratoire de recherche, avec une grande diversité de profils et une ambiance de travail

agréable. Cela m'a permis de m'introduire de la meilleure des manières au monde de la recherche scientifique, un univers stimulant et riche en idées.

Sur le plan des compétences, ce stage m'a permis de consolider des acquis essentiels pour la suite de mon parcours, qu'il s'agisse du développement logiciel, de la gestion autonome d'un projet ou encore de la collaboration dans un cadre interdisciplinaire. J'ai également découvert un domaine qui m'a particulièrement attiré : la visualisation de données. Réfléchir à la manière de représenter clairement des résultats complexes – cartes, graphiques, interfaces – a été une facette à la fois technique et créative du stage que j'ai beaucoup appréciée. Elle fait écho à mon intérêt pour l'UI développé lors de projets à Polytech, et ouvre peut-être une piste à explorer plus profondément à l'avenir.

7.4 Évolution de mon projet professionnel

Avant ce stage, je m'interrogeais sur la possibilité de poursuivre mes études par une thèse, ou de m'orienter directement vers le monde de l'entreprise. Ces quelques mois au LIRMM m'ont permis de clarifier ces points.

Aujourd'hui, je ne me projette pas dans une thèse à court terme. Ce choix est lié à la fois à des envies personnelles, à un besoin de stabilité, mais aussi au fait que les sujets que j'ai trouvé dans mes recherches et qu'on m'a proposés pendant mon stage ne m'ont pas pleinement convaincu. J'ai tout de même passé un entretien pour une thèse CIFRE en lien avec AcuSurgical — une entreprise cofondée par l'ancien directeur du LIRMM évoquée plus tôt dans le rapport — ce qui m'a permis d'explorer cette voie. Toutefois, je pense qu'il serait prématuré de m'y engager dès maintenant, notamment en raison du niveau académique élevé que ce type de parcours exige.

Pour autant, je n'exclus pas de revenir vers une thèse CIFRE à moyen terme. J'ai rencontré plusieurs doctorants ayant choisi de débuter une thèse CIFRE après quelques années de CDI, une trajectoire qui pourrait me correspondre davantage. Je compte d'ailleurs poser la question de l'éligibilité aux dispositifs CIFRE lors de mes prochains entretiens d'embauche, car cela pourrait influencer mon choix d'entreprise.

Dans l'immédiat, mon objectif est de trouver un poste en CDI dans une entreprise implantée dans la région de Montpellier. Je resterai attentif aux opportunités qui pourraient combiner activité en entreprise et recherche appliquée. Notamment dans une équipe de R&D.

8. Webographie

SOURCES :

[LIRMM - Robotics Place](#)

[LIRMM](#)

[Université de Montpellier](#)

[Université de Montpellier 2](#)

[LIRMM - Wikipédia](#)

[LIRMM - Équipes](#)

[SESAME](#)

[LIRMM - Équipe ADVANSE](#)

[MaDICS - DAE](#)

[DAE](#)

<https://www.lirmm.fr/european-projects-en/>

<https://arxiv.org/abs/2305.03139>

<https://cordis.europa.eu/project/id/101076911>

<https://www.umontpellier.fr/articles/30-ans-dinnovation-avec-le-lirmm>

<https://foreverpollution.eu/>

<https://www.operation-campus.fr/secteur-occitanie/pole-stics-st-priest-universite-de-montpellier>

Sources Annexe te&ds

ONU Environnement – [PFAS Global Problem](#)

Forever Pollution Project – <https://foreverpollution.eu>

France Info – “Des œufs contaminés aux PFAS dans l’Ain”, mars 2023.

ANSES – “Les effets sanitaires de certains PFAS”, 2022.

Santé publique France – “Exposition de la population française aux PFAS”, 2019.

ECHA – Proposal for PFAS restriction (2023)

Ministère de la Transition écologique – “Plan d’action PFAS”, 2023.

9. Annexes

Annexes.

Présentation des PFAS et des problématiques liées à leur présence dans l'environnement.

Le développement suivant fait office d'annexe TE&DS (Transition Écologique & Développement Soutenable).

Que sont les PFAS ?

Les PFAS (substances per- et polyfluoroalkylées) forment une vaste famille de plusieurs millions de composés chimiques synthétiques. Leur particularité réside dans la présence de liaisons carbone-fluor, extrêmement stables, ce qui leur confère des propriétés recherchées dans l'industrie : résistance à l'eau, aux graisses, aux hautes températures.

C'est pourquoi on les retrouve depuis les années 1950 dans de nombreux produits du quotidien : emballages alimentaires, textiles imperméables, revêtements antiadhésifs (comme le Téflon), mousses anti-incendie, cosmétiques, etc. Mais cette même stabilité chimique fait aussi des PFAS des polluants dits "éternels", car ils se dégradent très lentement dans l'environnement.

- En 2023, l'ONU a reconnu les PFAS comme un problème sanitaire mondial croissant [lien].

Pourquoi en retrouve-t-on dans l'environnement ?

Les PFAS sont relâchés dans l'environnement tout au long de leur cycle de vie : production industrielle, usage domestique, incinération ou enfouissement des déchets, rejets dans les eaux usées. Ces composés sont mobiles : ils voyagent dans l'air, l'eau, les sols, et peuvent parcourir de longues distances.

Des investigations comme le Forever Pollution Project² ont révélé des niveaux inquiétants de pollution à proximité d'usines, d'incinérateurs ou de sites militaires, avec parfois des teneurs dépassant les seuils sanitaires recommandés de plusieurs centaines de fois.

- Plus de 17 000 sites pollués ont été recensés en Europe [lien].

Quelles en sont les conséquences ?

Sur l'environnement

Les PFAS contaminent les eaux de surface, les nappes phréatiques et les sols. Ils s'accumulent dans les plantes, les poissons et les animaux, et perturbent les écosystèmes.

Sur les cultures et l'alimentation

En s'infiltrant dans les sols agricoles, les PFAS peuvent passer dans les cultures. Certaines zones en France ont vu leurs produits interdits à la consommation, comme des œufs ou des légumes [lien].

Sur la santé humaine

Les PFAS s'accumulent dans le corps humain et sont suspectés de provoquer divers effets : troubles hormonaux, cancers, baisse de la fertilité, affaiblissement du système immunitaire [lien]. Certains sont classés comme cancérogènes probables (PFOA, PFOS).

- En France, 100 % des échantillons sanguins testés contiennent des PFAS [lien].

Quelles mesures sont prises ?

En Europe

L'Union européenne a restreint certains PFAS via le règlement REACH. En 2023, cinq pays (dont la France) ont proposé une interdiction quasi totale de cette famille chimique [lien]. Le Parlement européen a aussi appelé à une interdiction des PFAS dans tous les emballages alimentaires d'ici 2030.

En France

Un plan d'action national a été lancé en 2023 avec plusieurs axes : renforcer la surveillance, identifier les sites pollués, réduire les émissions industrielles, informer le public [lien]. Plusieurs enquêtes journalistiques ont poussé à une prise de conscience publique.

- Depuis 2024, les industriels doivent déclarer leurs rejets de PFAS dans les cours d'eau.

Renvoi vers le texte