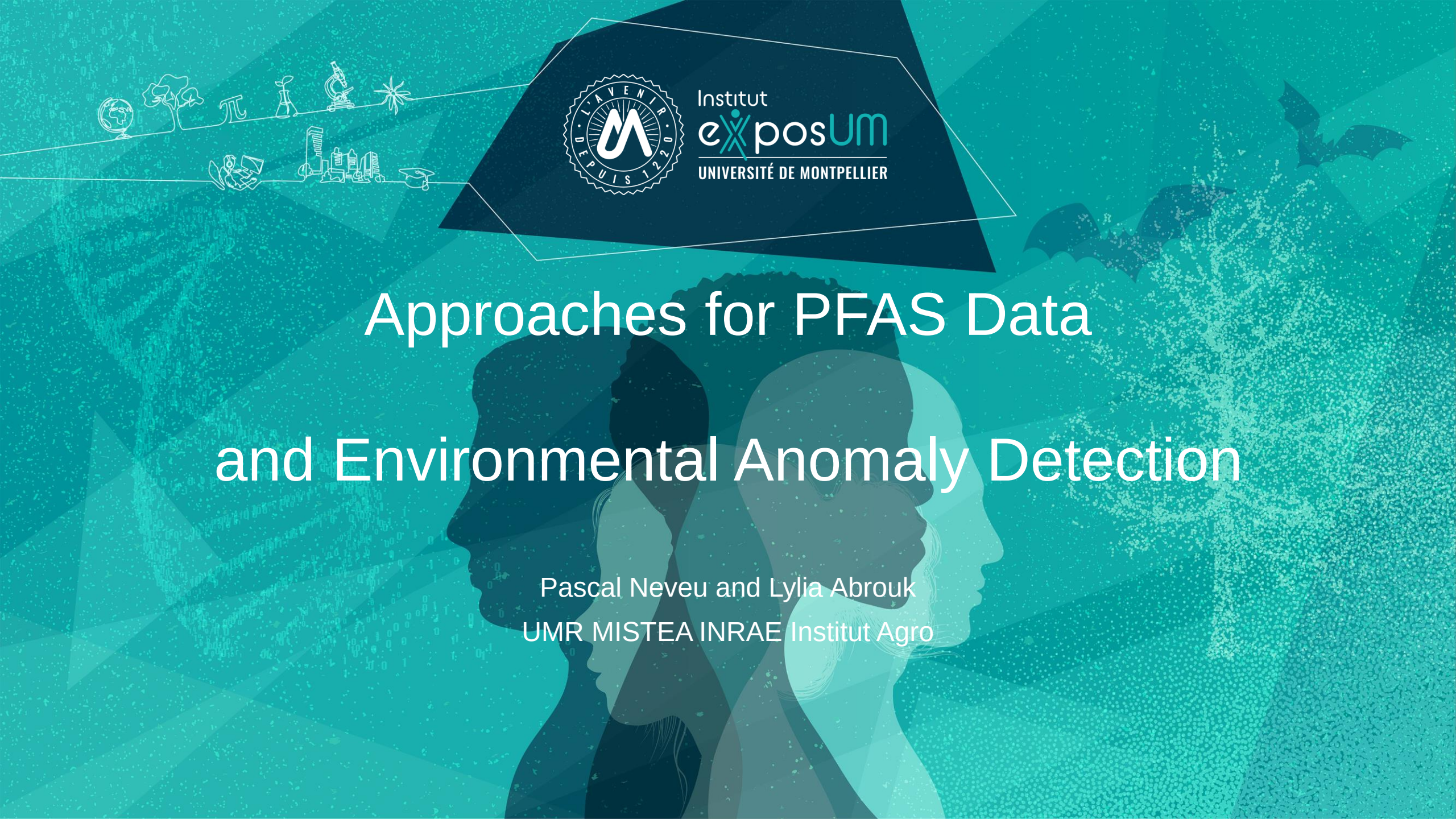


Approaches for PFAS Data and Environmental Anomaly Detection

Pascal Neveu and Lylia Abrouk

UMR MISTEA INRAE Institut Agro

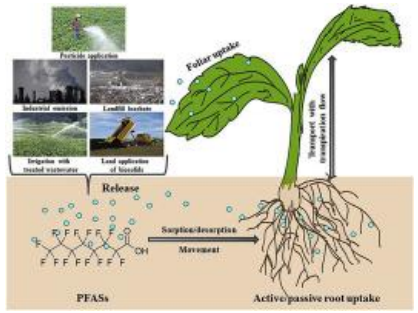
What are PFAS?

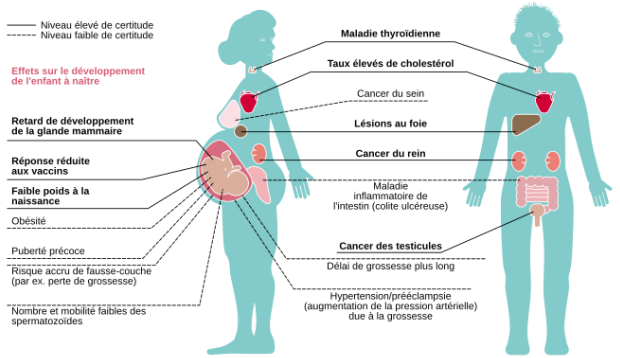
PFAS (per- and polyfluoroalkyl substances) properties: flame-retardant, stain-resistant, non-stick, emulsifying and waterproofing.

Used in many everyday products: kitchen utensils, textiles, food packaging, non-stick coatings, cosmetics, plant protection products, etc.

Stability of PFAS makes them highly resistant to degradation and persistent in the environment. Often called Forever Chemical

Some PFAS have proven toxicity and ecotoxicity





Institut exposUM
UNIVERSITÉ DE MONTPELLIER

Toxicity

Documented effects (known or suspected):

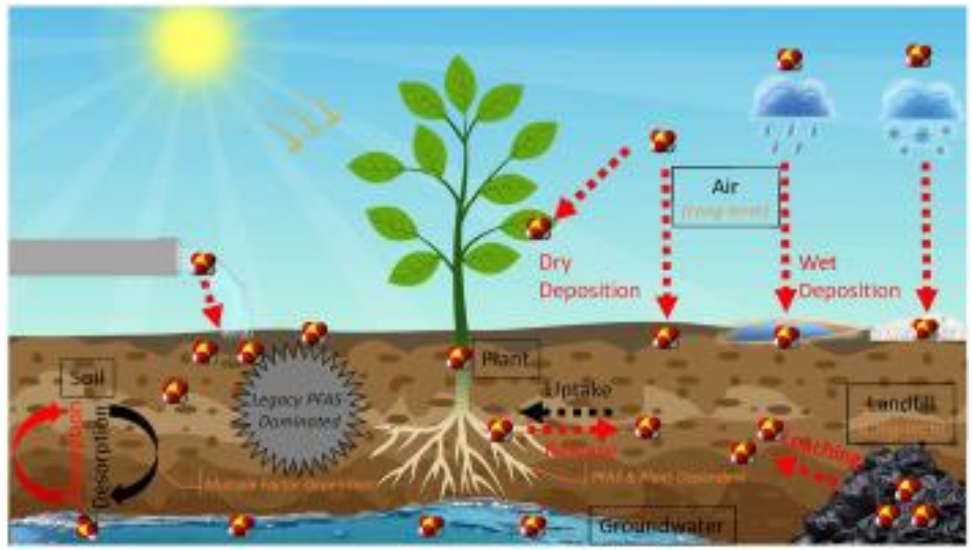
- Probable carcinogenicity (kidney, testicles)
- Impact on birth weight
- Alteration of the thyroid system (development)
- Alteration of cholesterol metabolism
- Alteration of steroidogenesis (gonadal development and reproduction)
- Alteration of the immune system (vaccine response)
- Impact on biodiversity
- ...

Data for a limited number of PFAS: PFOA, PFNA, PFHxS and PFOS

Impossibility of providing individual toxicity data for all PFAS (groups)

Where are PFAS?

Presence of PFAS pollutants:
air, soil, water, tap water, animals, plants… and humans.



PFAS

We still don’t know much about them

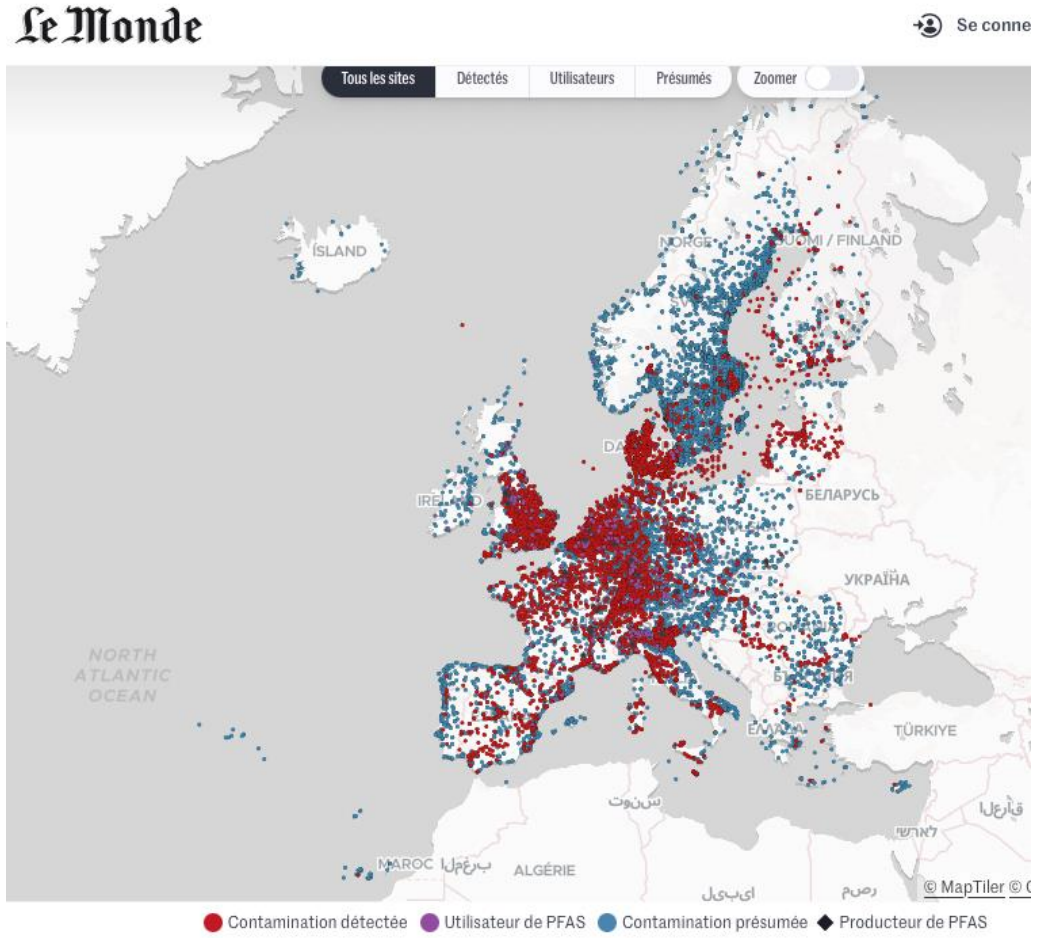
Data for a limited number of PFAS: PFOA, PFNA, PFHxS and PFOS

- Their inventory, characterization, dynamics, signatures
- Their toxicity, effects and interactions
- Behavior in complex ecosystems and pollution chains
- Identification of impacts and anomalies detection

Institut exposUM UNIVERSITÉ DE MONTPELLIER

Need data to better understand PFAS

In this regard, *Le Monde* has launched an initiative of data publication



Project Objectives

- Address PFAS impacts through a **multidisciplinary** approach
- Develop collaboration between data science, pollutant chemistry, environmental impact studies, and environmental health
- Raise awareness among scientific and industrial communities on PFAS-related issues
- Intensify research and innovation on PFAS issues.

Institut eXposUM UNIVERSITÉ DE MONTPELLIER

Data sciences for PFAS Challenges

- What information can be extracted from *Le Monde* data
  - Comparing sites (clustering, looking for anomalies)
- Enrich and reuse Le Monde data
  - Identifying relevant data sources: production sites, water treatment, blood donation, etc.
  - Propose a structuring and standardizing data for added value.
  - Link it to other data sources
  - Data consolidation
  - Data integration → Constitute reference data sets from multiple sources
  - Explore and analyze data sets and search for anomalies

Institut exposUM UNIVERSITÉ DE MONTPELLIER

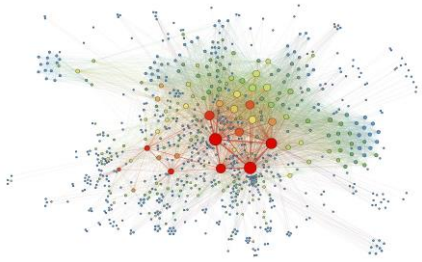
Partner Institutes and Laboratories

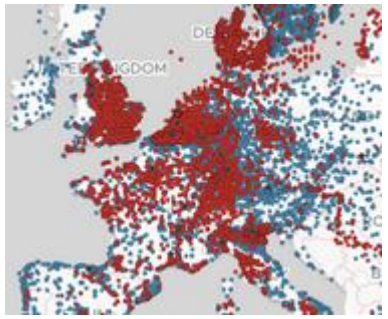
- **MISTEA** (INRAE and Institut Agro)
- **LIRMM** (CNRS UMR 5506, University of Montpellier)
- **EPOC** (Environnements et Paléoenvironnements Océaniques et Continentaux, UMR CNRS 5805)
- **PRODIG** (Pôle de Recherche pour l'Organisation et la Diffusion de l'Information Géographique, UMR 8586 CNRS)

Data harmonization

- Why use ontologies (knowledge models)?
  - Shared semantics for observed entities and their properties
  - Expressing semantics in a machine-readable form
- Use of ontologies is crucial in a cross-disciplinary context
  - enables advanced data management (sharing, linking, integration, etc.)

Our objective is to constitute and enrich an ontology network
→ essential for data and knowledge extraction



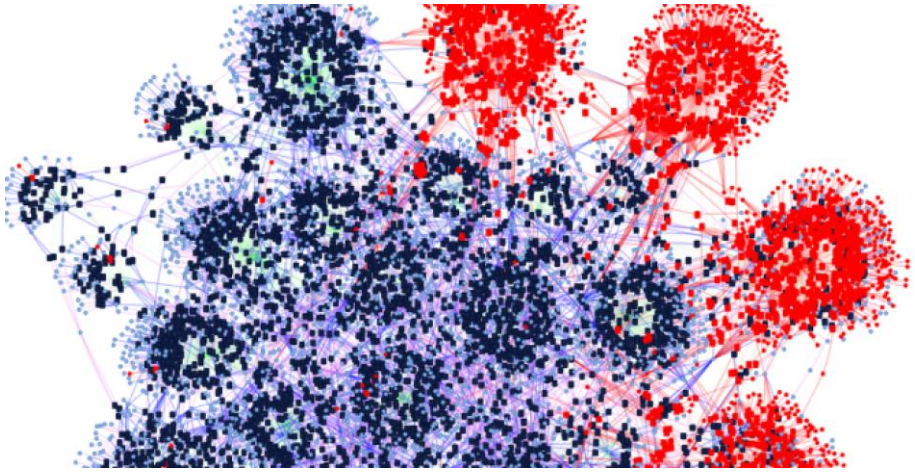


Data Collection and Consolidation

- Extract information from *Le Monde* data as a starting point
- Identify sources (sites, health, pollutants, etc.)
- Data harmonization for integration and consolidation over time and space
- Improve the quality and value of data
- Visualization to explore, compare and locate relevant data from different sites and sources

Data Analysis: ML Approaches

- Adaptation and use of supervised models for labeled data such as PFAS risks or PFAS signatures
- Adaptation and use of unsupervised models for knowledge extraction for clustering of sites, anomalies
- Development of semi-supervised learning to combine unsupervised and supervised models
- Visual analytics of clusters



Institut exposUM
UNIVERSITÉ DE MONTPELLIER

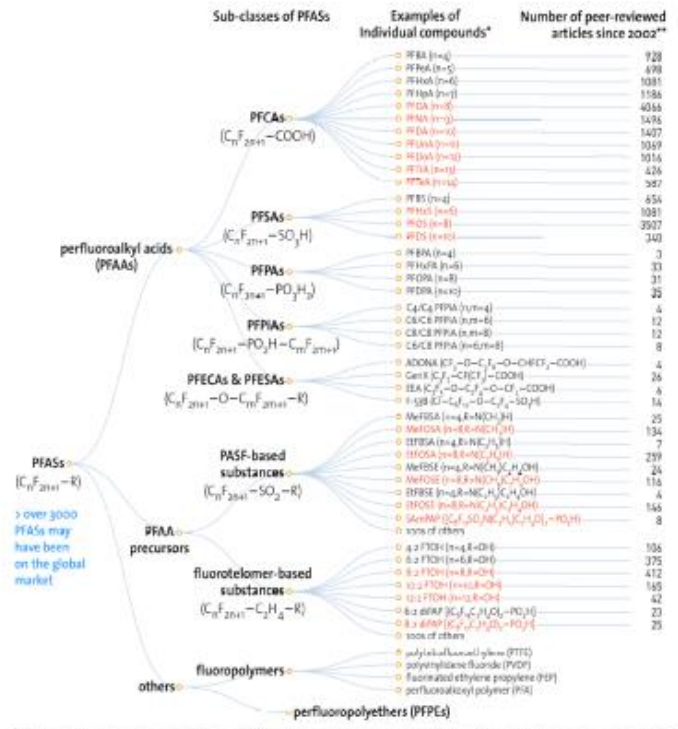
Data Analysis and Anomaly Detection

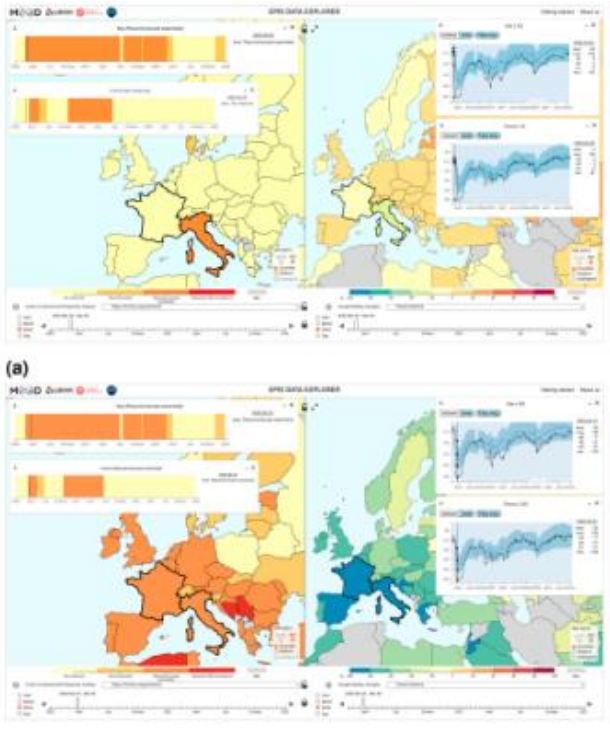
- Data analysis to detect patterns or anomalies in data sets.
- Hybrid approach: combining ontologies and machine learning algorithms.
- Spatio-temporal clustering to define normal behaviors before identifying anomalies.
- Expected results:
  - site not probable PFAS type (signature issue).
  - Unpredictable biodiversity impacts

Institut exposUM UNIVERSITÉ DE MONTPELLIER

Visual Analysis

- Development of visual interfaces to explore unstructured and structured data.
- Facilitating visual reasoning through interactive, coordinated views.
- Contribution to the explainability of machine learning models
- Model evaluations





Expected Outcomes

- Development of an ontology network for environmental PFAS for a better management of data and knowledge.
- Data analysis models and tools for identifying contamination clusters
- Extracted relevant information from various data sets
- Highlight anomalies based on constructed data sets
- Creation of a long term of network of partners.

Conclusion

The project is an opportunity to provide innovative solutions to contribute to a better targeting of guidelines to combat PFAS pollution.

- Identifying and integrating promising data sources
- Propose suitable methods for analyzing and visualizing data, enabling the discovery of anomalies
- Accelerating acquisition of knowledge with an integrated approach
- Building a multidisciplinary network
- Contributing to environmental and public health protection: makes unable predictive model of PFAS impact on public health and the environment.

Institut exposUM UNIVERSITÉ DE MONTPELLIER