# A Geospatial Knowledge Graph for Analyzing the Fate and Transport of Environmental Contaminants

**Katrina Schweikert, Torsten Hahmann, Shirly Stephen**

## ABSTRACT:

Per and Polyfluoroalkyl substances (PFAs) are a group of thousands of chemicals that do not naturally break down in the environment and can bioaccumulate. Their manufacture and usage over several decades in various applications have made them widespread throughout the environment and following numerous environmental transport pathways to cause contamination in food and drinking water supply. The human health impacts are still being studied, but they are known to increase the risk of kidney and testicular cancer and decrease immunity and the effectiveness of vaccines in children, among other documented negative health impacts. The sampling and investigation of PFAs environmental contamination in the U.S. is currently done across various agencies at the state and federal level that manage agriculture, drinking water, and environmental regulations. The heterogeneity of the data regarding PFAs sampling and information on known and suspected contamination sources make it challenging to build a holistic picture of contamination. Integrating the heterogeneous data together will help to answer ongoing questions about exposure, environmental fate and transport, and resource prioritization. Additionally, in some states sampling data is sparse due to limited funding and no definitive Federal Regulations for a maximum concentration level in food, water, and soil. This makes modeling, estimation and other hypothesis testing regarding spatial patterns of contamination critical to risk assessment and prioritization of limited resources.

To help gain a more holistic picture of PFAs contamination and guide testing and mitigation efforts, we construct a Knowledge Graph (KG) that combines sample data from state and federal testing programs, with information on the different chemicals, potential sources, and the properties of vulnerable sites. A KG is a special type of information system that links heterogeneous data together and aligns them to a common ontology or schema of concepts. We connect various sampling results with information on nearby potential sources, potentially impacted features of concern and environmental context such as hydrology, soil, and geology, all via the spatial relations of the various information. These connections will help better model the environmental fate and transport mechanisms of different PFAs. We also include regional administrative and organizational geometries to support aggregation of statistics and comparison with demographic measures of vulnerability and environmental justice.

We then demonstrate how KG queries and tools can be leveraged to investigate hypotheses about the complex contamination mechanisms and find spatial patterns for the environmental transport. As examples, we construct interactive maps that aggregate known contamination and potential risk to private well water supply and surface waters in the state of Maine. While the graph is currently still being expanded, these statewide summaries showcase the graph's analytic and visualization capabilities. Once completed, the KG will integrate data on PFAs contamination around the country to answer questions about where there is risk, where more testing is needed, and what patterns we can learn from the existing sampling data.

**KEYWORDS:** *geospatial knowledge graph, environmental contamination, environmental justice*

**Katrina Schweikert**, PhD Student, Spatial Information Science and Engineering, University of Maine, Orono, ME, USA

**Torsten Hahmann**, Associate Professor, Spatial Information Science and Engineering, University of Maine, Orono, ME, USA

**Shirly Stephen**, Postdoctoral Research Associate, Center for Spatial Studies, University of California, Santa Barbara, CA, USA