

PET – An Open-Source Shiny Application for Scalable Quality Assurance and Exploration of Transnational River Basin Monitoring Data

Severin Osten, Steffen Kittlaus, Ottavia Zoboli

Institute of Water Quality and Resource Management (IWAG), TU Wien, Vienna, Austria

River basin management is an inherently interdisciplinary and transnational task, increasingly challenged by the widespread occurrence of pollutants originating from diverse industrial and urban sources. Many of these substances remain insufficiently understood regarding their environmental fate and ecological impacts. Within the European Union, these challenges are addressed through the Water Framework Directive (2000/60/EC) and associated guidance documents, which explicitly require the establishment of integrated, multi-compartment inventories of emissions, discharges, and losses. These data infrastructures form the foundation for evidence-based assessment, modelling, and decision-making in river basin management.

In this context, the Institute of Water Quality at TU Wien has developed a centralized, multi-compartment concentration database within the framework of the Tethys project. The database is designed as a robust and scalable infrastructure to support national and transnational monitoring activities and is already operational for data contributions from multiple partners across the Danube River Basin. However, the practical use of such complex database systems remains a major challenge. Data import, validation, curation, and exploration typically require substantial technical expertise in database technologies and programming, posing a significant barrier for many domain experts.

To address this gap, we present the Pollution Evaluation Tool (PET), an open-source R package implemented as a Shiny application using the {golem} framework. PET provides a user-friendly, scalable interface for interacting with databases structured according to the Tethys data model. It fully integrates with existing authentication mechanisms and complies with local database constraints, enabling secure and standardized access across institutions.

The currently operational module of PET focuses on structured data quality assurance. It implements standardized, best-practice QA/QC workflows that integrate quantitative and qualitative cross-table consistency checks across core database entities, including sampling sites, samples, measurements, determinants, laboratories, analytical methods, and associated classifications and standards. These workflows support systematic detection of inconsistencies, missing information, and structural errors, thereby improving data reliability for downstream applications such as emissions modelling and large-scale assessments.

Beyond quality assurance, PET provides advanced filtering, constrained selection based on valid metadata, and functionality for updating and creating database entries through guided interfaces, substantially reducing the need for direct interaction with the database backend. A second development phase, currently conceptualized, will extend the system towards comprehensive cross-database exploration, including flexible visualization, environmental statistical analyses, metadata management, complex error detection, and tailored dataset generation.

This work represents a methodological and infrastructural contribution by operationalizing standardized, scalable, and open workflows for quality assurance and exploration of large, transnational environmental monitoring databases. PET lowers technical barriers for domain experts and supports the practical implementation of FAIR data principles, thereby enabling more effective use of emerging large-scale river basin data infrastructures.