

# Internship Proposal

## Monitoring and Visualization of Water Resource and Environmental Data Contributing to Rivers:

### Data Integration, Automated Cleaning, and Visualization Dashboard

*Suivi et Visualisation des Données Météorologiques et de Ressources Hydriques Contribuants aux Rivières : Intégration de Données, Nettoyage Automatique, et Dashboard de Visualisation*

#### Supervisors (*co-encadrants*)

Walter Rudametkin (Professor, DiverSE IRISA) <[Walter.Rudametkin@irisa.fr](mailto:Walter.Rudametkin@irisa.fr)>

Johann Bourcier (Associate Professor, HDR, DiverSE IRISA) <[Johann.Bourcier@irisa.fr](mailto:Johann.Bourcier@irisa.fr)>

Laurie Boithias (Associate Professor, M-TROPICS GET) <[Laurie.Boithias@get.omp.eu](mailto:Laurie.Boithias@get.omp.eu)>

#### Research team (*affiliation du promoteur*)

We are looking for one or two interns to collaborate on this subject. The interns will join the DiverSE project-team at the IRISA laboratory in Rennes, France, and will collaborate with the GET laboratory in Toulouse.

#### DiverSE Project-team

<https://www.diverse-team.fr/>

IRISA Rennes <a href="https://www.irisa.fr/">https://www.irisa.fr/</a> Campus de Beaulieu 263 av. du Général Leclerc 35 042 RENNES cedex	Géosciences Environnement Toulouse (GET) <a href="https://www.get.omp.eu/">https://www.get.omp.eu/</a> Observatoire Midi-Pyrénées 14, avenue Édouard Belin 31400 TOULOUSE
--	---

#### Keywords (*mots clés*)

Software Science, Datascience, DevOps, Web programming

## Project description (*description du projet*)

The proposed internship is situated within the context of monitoring bacteriological contaminants in rivers, aiming to establish links between human activities, weather conditions, and river contamination. The primary objective of this internship is to design a comprehensive data management system, including automated cleaning of field-collected data, integration of data from multiple sources, and the creation of a visualization dashboard for in-depth analysis over an extended period.

*(Version Française)*

*Le stage proposé s'inscrit dans le contexte de la surveillance des contaminants bactériologiques dans les rivières, visant à établir des liens entre les activités humaines, les conditions météorologiques et la contamination des cours d'eau. L'objectif principal de ce stage est de concevoir un système complet de gestion des données, incluant le nettoyage automatique des jeux de données collectées sur le terrain, l'intégration de données provenant de multiples sources, et la création d'un tableau de bord de visualisation pour une analyse approfondie sur une longue période.*

## Objectives (*objectifs*)

- Systematic and Automated Data Cleaning: Develop automated methods for cleaning raw data from field collection, removing errors, outliers, and duplicates, thus ensuring data quality.
- Design of a Visualization Dashboard: Design and develop an interactive dashboard for visualizing the collected data. This dashboard will enable users to monitor real-time levels of rain, wind, solar radiation, temperature, humidity and other data from the monitored rivers.
- Integration of Multiple Data Sources: Implement a data integration system that combines data from various sources such as meteorological and hydrological field sensors, for a more comprehensive analysis of factors that contribute to, for example, the spread of bacterial contaminants and extreme weather events in watersheds and river systems.
- Develop a data pipeline inspired by DevOps CI/CD approaches for technicians to easily feed the data into the system to obtain a close to real-time monitoring system.

*(Version Français)*

- Nettoyage Systématique et Automatique des Données : Développer des méthodes automatisées de nettoyage des données brutes issues de la collecte sur le terrain, en éliminant les erreurs, les valeurs aberrantes et les doublons, garantissant ainsi la qualité des données.
- Conception d'un Dashboard de Visualisation : Concevoir et développer un tableau de bord interactif pour visualiser les données collectées. Ce tableau de bord permettra aux

*utilisateurs de suivre en temps réel les niveaux de contamination bactériologique dans les rivières surveillées.*

- *Intégration de Données Multiples : Mettre en place un système d'intégration de données provenant de diverses sources telles que capteurs sur le terrain, données météorologiques, et données démographiques, pour une analyse plus complète des facteurs de contamination.*
- *Développer un pipeline type DevOps CI/CD permettant aux techniciens d'introduire facilement les données dans le système afin d'obtenir un système de surveillance en temps quasi réel.*

## **Methodology (méthodologie)**

The internship will involve a research, development, and implementation phase of the technical solutions necessary to achieve the stated objectives. You will work in collaboration with an experienced team in the field of environmental monitoring and use data processing and visualization tools to successfully complete your project.

*(Version Française)*

*Le stage impliquera une étape de recherche, de développement, et d'implémentation des solutions techniques nécessaires à l'atteinte des objectifs énoncés. Vous travaillerez en collaboration avec une équipe expérimentée dans le domaine de la surveillance environnementale et utiliserez des outils de traitement de données et de visualisation pour mener à bien votre projet.*

## **Expected results (Résultats attendus)**

- Source code for automated data cleaning scripts.
- Interactive data visualization dashboard.
- Detailed report explaining the methodology, obtained results, and any recommendations for improving the monitoring of watersheds and rivers.

*(Version Française)*

- *Code source des scripts de nettoyage automatisé des données.*
- *Tableau de bord interactif de visualisation des données.*
- *Rapport détaillé expliquant la méthodologie, les résultats obtenus et les recommandations éventuelles pour améliorer la surveillance des bassins versants et les rivières.*

## Desired profile (*Profil recherché*)

We are looking for a motivated intern with skills in data processing, programming (Python, R, etc.), and an aptitude for data visualization. An understanding of environmental issues would be an asset.

This internship topic offers a unique opportunity to contribute to the preservation of our water resources while developing valuable skills in data analysis and visualization. If you are interested in this topic, please don't hesitate to apply and join us in this crucial mission for the environment.

*(Version Française)*

*Nous recherchons deux stagiaires motivés avec des compétences en traitement de données, programmation (Javascript, Python, R, etc.), et une aptitude à la visualisation des données. Une compréhension des problématiques environnementales serait un atout.*

*Ce sujet de stage offre une opportunité unique de contribuer à la préservation de nos ressources en eau tout en développant des compétences précieuses en analyse de données et en visualisation. Si ce sujet vous intéresse, n'hésitez pas à postuler et à nous rejoindre dans cette mission cruciale pour l'environnement.*

## Computing resources (*Moyens informatiques*)

The research center will take care of all required tools and resources.

*À la charge de l'organisme d'accueil.*

## Skills Summary

The interns will develop their skills in Data science, Javascript, Web programming, DevOps, CI/CD, Linux, Docker, Git, and many other technologies.

As is a common practice in the DiverSE research team, all source code will be open sourced using either the GPL or the Apache License. It is expected that the students participate in related open source communities. This should also assist in the technological transfer from academic prototype to industrial-ready tools.

An advanced prototype to showcase the data, including a configurable and interactive dashboard, should be implemented and compatible with mobile applications.

## Bibliography (Bibliographie)

- [1] Boithias L., Auda Y., Audry S., Bricquet Jean-Pierre, Chanhphengxay A., Chaplot Vincent, De Rouw Anneke, Henry des Tureaux Thierry, Huon S., Janeau Jean-Louis, et. al. The Multiscale TROPICAL CatchmentS critical zone observatory M-TROPICS dataset II : land use, hydrology and sediment production monitoring in Houay Pano, northern Lao PDR. *Hydrological Processes*, 2021, 35 (5), e14126 [9 p.] <https://onlinelibrary.wiley.com/doi/10.1002/hyp.14126>
- [2] Ather Abbas, Sangsoo Baek, Minjeong Kim, Mayzonee Ligaray, Olivier Ribolzi, Norbert Silvera, Joong-Hyuk Min, Laurie Boithias, Kyung Hwa Cho. Surface and sub-surface flow estimation at high temporal resolution using deep neural networks. *Journal of Hydrology* Volume 590 2020. <https://www.sciencedirect.com/science/article/pii/S0022169420308301>
- [3] Abbas, A., Baek, S., Silvera, N., Souleuth, B., Pachepsky, Y., Ribolzi, O., Boithias, L., and Cho, K. H.: In-stream *Escherichia coli* modeling using high-temporal-resolution data with deep learning and process-based models, *Hydrol. Earth Syst. Sci.*, 2021. <https://hess.copernicus.org/articles/25/6185/2021/>
- [4] June Sallou. On reliability and flexibility of scientific software in environmental science : towards a systematic approach to support decision-making. *Software Engineering [cs.SE]*. Université Rennes 1, 2022. English. <https://inria.hal.science/tel-03854849v1>
- [5] Dorian Leroy, June Sallou, Johann Bourcier, Benoit Combemale. When Scientific Software Meets Software Engineering. *Computer*, 2021, 54 (12), pp.60 - 71. <https://inria.hal.science/hal-03318348v2>
- [6] P Laperdrix, N Mehanna, A Durey, W Rudametkin - Proceedings of the ACM Web Conference 2022. The Price to Play: a Privacy Analysis of Free and Paid Games in the Android Ecosystem. <https://dl.acm.org/doi/abs/10.1145/3485447.3512279>
- [7] Tricoire, Maxime, et al. "KevoreeJS: Enabling dynamic software reconfigurations in the Browser." 2016 19th International ACM SIGSOFT Symposium on Component-Based Software Engineering (CBSE). IEEE, 2016.
- [8] Mishra, V., Laperdrix, P., Rudametkin, W., & Rouvoy, R. (2021, July). Déjà vu: Abusing Browser Cache Headers to Identify and Track Online Users. In *PETS 2021-The 21th International Symposium on Privacy Enhancing Technologies*.