

Design and Development of Digital Twin Prototype for Natural Environment

Amit Sharma^{1,3}, Felix Iglesias Vazquez², Frederic Ros³, Lynh HOANG-Vy-Thuy⁴

¹LE STUDIUM Institute for Advanced Studies, 45000 Orléans, France

²TU Wien, Institute of Telecommunications, Vienna-AU

³PRISME laboratory, University of Orléans, Orléans, France

⁴BRGM (Service géologique national | BRGM)

REPORT INFO

Fellow: **Dr. Amit Sharma**

From Chitkara University, Punjab, India

Host laboratory in region Centre-Val de Loire: PRISME Laboratory (EA 4229), Université d'Orléans, Orléans, France

Host scientist: **Prof. Frédéric Ros**

Period of residence in region Centre-Val de Loire: Oct 2024 – Oct 2025

Keywords:

Environmental Monitoring, Digital Twin, Time Series Prediction, Groundwater, Air Quality, Satellite Data, Hybrid Modeling, JUNON

ABSTRACT

The environmental monitoring and its efficient management highly dependent upon the integration of heterogeneous data sources, followed by advanced numerical approaches and artificial intelligence techniques. In alignment with the objectives of JUNON programme, this fellowship focused on design and development of operational Digital Twin (DT) for environmental monitoring dedicated to groundwater and air quality systems. The aim of this project is to propose a Digital Twin architecture which is capable of integrating diverse data from in-situ sensors, satellite observations, physics-based and data driven models and to map them in web platform. The proposed system allows historical visualization, prediction, forecasting, 2d spatial mapping, and model-oriented simulation through services and ensures extensibility and long-term maintainability. A major contribution of the fellowship was consolidation and extension of an initial Digital Twin architecture into fully functional web application, incorporates data manager, scheduler for periodic updates, modeling services and user-friendly web interface. The proposed architecture was validates using real-world data across Centre-Val de Loire region, demonstrating its capability to handle heterogeneous temporal and spatial resolutions, and data quality constraints. Along with the technical development, this fellowship also contributed towards methodological design for hybrid modeling policies, operational robustness, and Digital Twins role for natural resource management. The outcome of this work offers a solid foundation for future research publications, technology transfer, and continued collaboration within the JUNON programme.

1- Introduction

The sustainable management of natural environment requires progressively, refined, understanding or high resolution, characterization of environmental systems and their time dependent behavior to understand long-term dynamics. In recent years, the rapid expansion of environmental data combined with recent progress and methodological developments in computational modelling and artificial intelligence has enabled new

approaches for representing complex natural processes in a digital form. Among these techniques, the concept of Digital Twin (DT) has become more prominent for connecting observations, models, and bringing decision support tools all together within a unified system [1]. In the application of environmental sciences, Digital Twins are intended to deliver a dynamic Digital representation or replication of physical systems by iterative, integrating heterogeneous data streams, and the

computational models for data analytics [2]. Unlike conventional static approaches, these systems intended to update overtime, allowing continuous updating of system states and supports scenario-based evaluation. However, implementing this concept into operational system for assessment of groundwater levels/depth, their quality and air quality monitoring remains major technical and scientific challenge. The environmental systems like atmospheric compartments and aquifers are exhibit by (strong spatial heterogeneity), Complex interactions between natural processes and anthropogenic activities, and multi scale temporal dynamics [3]. The data describing these systems are collected through multiple sources, including in situ monitoring systems, meteorological observations, and satellite data each having distinct spatial granularities, uncertainties, and uneven temporal frequencies.

As a result, one of the main challenges concerns the consistent fusion of these heterogeneous datasets within a single or unified platform and visualization framework. Based on this context, the JUNON programme which is led by BRGM and supported by Centre-Val de Loire region, focuses on building, digital twins, dedicated to ground, water, and air systems. The primary objective of JUNON is not only to strengthen scientific knowledge, but also to establish digital services intended to support monitoring, prediction, and management of these natural resources. The work conducted during this fellowship forms a concrete contribution towards this scope by focusing on the conceptualization and implementation of an environmental DT prototype, dedicated to groundwater and air quality systems. The initial and original system architecture of DT were proposed and implemented in first functional version by initial developer Felix Iglesias Vazquez, establishing the fundamental principle of modularity and service-oriented design. Leveraging this foundation, the present work expanded the framework to incorporate additional environmental compartments, notably air quality, and to enhance the platform with new analytical and predictive approaches,

where special attention was given on maintaining architectural coherence while integrating new services and further improving robustness, extensibility, and operational readiness. Therefore, the main objective of this work was to move from the experimental prototype towards a more functionally, complete and operational DT platform. This required the integration of new data sources, the deployment of multiple modelling services, and the development of a user-oriented web interface, enabling users to visualize historical data, perform predictions and forecast, and visualize spatial patterns. In addition to the implementation task, this work also aimed to address methodological aspects related to hybrid modeling, data quality handling, and the Digital Twins role as a tool for more informed decision-making for environmental efficient management. The project aims to design and develop DT prototype for natural resources such as measurement of groundwater levels/depths quantity and air quality. The design consists of several challenges in terms of data integration, real time monitoring, forecast. The proposed architecture is based on ENVRI (environmental research infrastructure) reference model and attempted to provide a design which is flexible, scalable, interoperable and reusable.

2- Experimental Details

The DT system developed in this work completely relies on open-source software stack where both front-end and back-end services are designed using Python. The proposed architecture is based on the FIWARE reference model for context-oriented applications. The FIWARE Orion Context Broker is implemented to manage the environmental entities and time-varying observations. The MongoDB is used to persist the historical time series data, while the FIWARE Cygnus is implemented to stream entity updates from Orion to the database. The FIWARE STH-Comet component is used to aggregate and query the time-series data, which allows efficient retrieval of historical data.

2.1 Overall System Architecture

Figure 1 depicts the proposed architecture of DT for environmental monitoring.

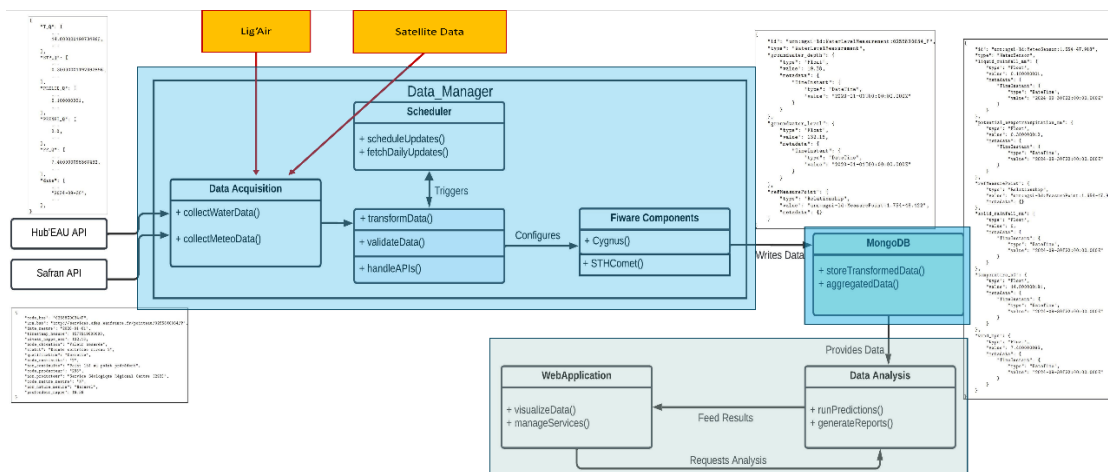


Figure 1: Proposed architecture of Digital Twins for natural environment

The proposed architecture is based on the ENVRI (environmental research infrastructure) reference model [4]. In this architecture, the data manager is the core module which is responsible for gathering the environmental data from different sources such as Hub'eau for ground water measurements [5], SAFRAN for meteorological data, lig'air for air quality data, and Copernicus for Satellite measurements. The data manager collects the data, transforms it to NGSI standard, and further injecting it to the MongoDB. The Cygnus component is used to update the database with the recent records and STH commit is used to retrieve the time series data from MongoDB. The second important part of the proposed architecture is the data analytical phase which is used for the visualization of historical records and forecasting future predictions.

The experimental work performed during this fellowship is based on the design and development of a Scalable DT framework intended to represent environmental system, dedicated to groundwater resources and air quality. The proposed design follows a multi-tier architecture in which data acquisition, data management, modeling services, and web interface are clearly separated while remaining interoperable. The proposed design follows a scalable, service-oriented architecture, which is intended to combine heterogeneous environmental data, analytical models, and

interactive visualization applications within a unified execution platform. The architecture was intentionally adopted to accommodate both groundwater related processes and air quality Monitoring, while allowing further extensions possible without any major architectural changes. The proposed system is structured around four functional layers, such as Data acquisition and management, storage and synchronization, analytical and modeling services, and user interface. Each layer in the proposed design operates autonomously while exchanging information through explicit data, exchange mechanisms, ensuring system, reliability, and long-term maintenance of the overall framework.

2.2 Data Manger Implementation

The data acquisition process was implemented through a standalone data manager module which is developed in python. The data manager module is responsible for handling acquisitions, processing, and organization of environmental data originating from multiple heterogeneous data providers [6]. It is responsible for retrieving the groundwater, meteorological, air quality and satellite data, transforms them to NGSI-compliant entities and further injecting them to the FIWARE ecosystem.

For groundwater level/depth measurements, the data manager connects to public hydrological

Data services through Hub'eau API, While the meteorological variables are extracted through SAFRAN API. The air quality observations are gathered from regional air quality stations through lig'air API And the satellite derived atmospheric observations are obtained through Copernicus (Sentinel-2). The data manager module is responsible for collecting all data, preprocessing, and performs temporal synchronization, unit harmonization, and contextual metadata augmentation before exporting the processed data, either as structured JSON files or as standardized entities compatible with NGSI-compliant context frameworks.

2.3 Context Management and Database Layer

The persistent data storage and its coordination were Implemented through MongoDB database in combination with FIWARE Orion context broker. The environmental readings are modeled using standardized NGSI entities, which enables the consistent management of both static metadata as well as temporal varying measurements. The measurement points for water and meteorological data, air quality stations and satellite grid cells are represented as separate contextual entities, while the observed readings such as groundwater levels/depth pollutant concentration, atmospheric readings are stored as temporal attributes. The proposed design supports optimized retrieval of historical time series data while maintaining logical links between spatial locations and observed environmental readings.

2.4 Data Analytical, Prediction, and Forecasting Services

The data analytical services were carried out using independent Python based modules. These services incorporate visualization of historical time series data, prediction, forecasting (short term and long term), spatial interpolation, and physical modeling workflows [7]. The machine learning (ML) based prediction and forecasting services were implemented using various libraries such as XGBoost, fireTS, scikit-learn, and Darts time

series forecasting framework. The difference between the prediction and forecasting service is that the prediction assumes partial knowledge of future explanatory variables, whereas the forecasting service operates with condition that no explicit knowledge of future values, and future values must be estimated internally through historical patterns.

2.5 Spatial Analysis and 2D Mapping

In DT the spatial analysis service is based on the creation of 2D (two-dimensional) maps derived from non-uniformly spaced observation points. The geographical buffer and the polygonal regions are created across each air quality monitoring station to estimate the atmospheric pollutants from Copernicus of selected region and mainly to synch point based observations with spatial domains. The interpolations techniques are then implemented to create continuous spatial fields which represents groundwater levels/depths, or air quality indicators. This service provides intuitive visualization of environmental variations and also supports the comparative analysis across different locations for two different time periods. The satellite-based observations are processed through polygon-based averaging, where the gridded pixels intersect, predefined regions are aggregated to generate time-series data.

2.6 User Interaction Layer (Web based application)

The DT interface is realized through an interactive web application which is developed using Dash framework. The application provides an integrated platform for data visualization, execution of analytical services, model configurations, results and recommendations. Through the interface users can interactively select any data source, measurement points, environmental variables, date range, and modeling approaches with graphical controls. The results are presented in the form of dynamic graphs, tables, interactive maps which allows users to evaluate both historical and future scenarios. The developed application supports both groundwater and air

quality twins, which ensures methodological uniformity while addressing each domain specific requirements.

2.7 Execution Environment

The proposed design was deployed on a Linux-based server infrastructure and executed within segregated python virtual environments. For the simplification of deployment and reproducibility the selected services are containerized using Docker Compose. This deployment strategy allows components to be updated individually, while maintaining stable system behavior, which makes the system appropriate for both exploratory research and future deployment scenarios.

3- Results and Discussion

3.1 Operational Integration of Digital Twin Components

The major outcome of this work is the successful integration of heterogeneous environmental data streams, data analytical models, and several visualization services into one unified operational DT system. The proposed system demonstrates that multiple measurements such as groundwater quantity, air quality, and satellite based atmospheric observations can be efficiently managed within a single unified architecture while preserving application specific characteristics. From the software viewpoint, the proposed design proved to be very effective and modular in isolating different processes, such as data ingestion, storage, analytics, and visualization services. The data manager fetches the data from external API's, transforms it to NGSI (Next Generation Service Interfaces) model and populated it to the database, which were further accessed in real time through web application without any manual intervention. It confirms the feasibility of adopted NGSI-based transformations for multipurpose environmental digital twins. Most importantly, the same architecture foundation supports both water quantity and air quality twin, which validates the hypothesis that a single DT framework can accommodate several environmental subsystems.

3.2 Data Availability, Temporal Coverage and Consistency

The adopted data ingestion strategy Successfully retrieve historical data sets of multiple years for groundwater levels/depths, meteorological observations, air quality pollutants, and satellite observations. The daily time series record of groundwater observations is fetched from hydrological repositories, whereas the meteorological variables are spatially matched to the measurement points to link the weather data ensuring spatial and temporal consistency. For air quality twin, the station-based observations are integrated with satellite-based measurements which are aggregated over specified polygons. This polygon-based data aggregation approach allows satellite observations in alignment with ground-based observations and thereby providing comparable time series data despite of different spatial resolution.

3.3 Trends Visualization

The designed application platform supports interactive analysis of both and temporal pattern of each environmental variable from multiple data sources. The Home page of the application provides the current status view of system state, consisting of maps, scattered plots, and tabular summaries. The homepage provides facility to user to select any station through interactive map of the region, after selection of station, the homepage will be updated with the most recent data of different variables of the selected station. The user can visualize the past records by exploring the historical visualization service. This service allows user to inspect time series behavior, analyze seasonal patterns, recent or past trends.

3.4 Prediction and Forecasting Performance

The prediction and forecasting services, were examined using historical trends retrieved from MongoDB database. The machine learning based prediction service demonstrates prediction at every point by using the future explanatory values [8], while the forecasting service projects future values with

autoregressive assumptions [9]. DARTs based time series forecasting service is also integrated which successfully generates short and long-term forecasts using several models such as Prophet, ARIMA, N-BEATS, and Transformer based architectures. These outcomes from machine learning based models highlights both the strength and weakness of data driven forecasting techniques, nearby, reinforce in the importance of complimentary physical inspired models [10].

The significant functional enhancement of proposed system is achieved with the integration of physical inspired model. The Gardenia model is developed by BRGM which is a process-based model integrated to simulate the groundwater levels, its optimization and forecasts. This service offers a foundation Scenario based analysis, facilitates user to explore the impact of changes in climatic or hydrological conditions and thereby extending the capability of DT from a monitoring platform towards decision support system.

3.5 2D Mapping

The 2D (two dimensional) mapping service enables the construction of interpolated spatial surfaces from the non-uniformly distributed monitoring stations. The resulting maps exhibits spatially structured transitions aligned with known groundwater and atmospheric patterns. In case of groundwater levels highlights zones with lower storage and higher storage, while air quality maps represent zones with increased pollutants concentrations. Despite the constrained interpolation performance by the spatial sensor distribution, the generate maps are effective for qualitative spatial analysis and visualization.

3.6 Discussion and System Limitations

The results demonstrate the practical feasibility of deploying an operational DT offering multiple environmental domains built exclusively using open-source software components. The proposed framework successfully integrates heterogeneous data, intelligent analytics, and interactive platform for visualization and forecasting tasks within

unified and consistent system. Despite of architectural advantages, a number of structural and operational limitations were also observed. The complete real time execution is limited due to the delays in data availability, forecasting uncertainty grows noticeably with the extension of forecast horizon, and the accuracy of spatial interpolation strongly depends on the density of monitoring stations. These limitations are fundamental for complex environmental systems and emphasize the importance of hybrid approaches that combines data driven, physical modeling, and expert informed techniques. Despite of these limitations, the proposed DT system forms a robust experimental platform suited to conduct a research activities, technical demonstrations, and future progressive operational deployment within the JUNON programme.

4- Conclusion

The fellowship has resulted in design enhancement, implementation, and validation of operational DT framework dedicated to groundwater, and air quality assessment within the JUNON programme. The work highlights that the heterogeneous data consisting of groundwater quantity, air quality, and satellite observations can be integrated into unified digital infrastructure, while preserving semantics and specific constraints of each domain. The proposed model ensures the relevance of service-oriented architecture for environmental DT. The separation among data acquisition, storage, and data analytics demonstrated significant for extensibility and scalability. The implementation of standardized data models and interfaces ensures consistent intersection among different components and supports addition of new DT, services following same architectural principles. The developed web interface ensures the practicality of DT by offering interactive access for users to visualize historical data, run prediction and forecast services, generate 2D maps through interpolation, and testing process-based modeling workflows. This work presents a solid foundation for future developments with Lin JUNON. The existing DT is an evolving research platform, capable of supporting

scientific innovations, methodological comparisons, and operational use cases related to efficient and sustainable management of natural resources.

5- Perspectives of Future Collaborations with the Host Laboratory

The outcomes of this fellowship open various concrete perspectives for future collaboration with the host laboratory and also with the JUNON ecosystem. The existing model offers solid basis for extending current functionalities with additional environmental twins and services such as soil moisture, surface water, land atmospheric interactions following same architectural principles. The future collaborations could focus on scientific advancement by designing and integrating physics-based models, and machine learning techniques [11]. The robustness, interpretability, and long-term prediction performance can further be improved by integrating hybrid modeling strategies. The platform also offers opportunities for methodological development, including comparative analysis of forecasting approaches, sensitivity analysis across various models and data sources. These analyses could directly support high impact scientific publications and possibilities to contribute towards international research on environmental Digital Twins. Additionally, the design could be progressively adapted for decision support and operational contexts in collaboration with industrial partners. The continued collaboration would naturally support the presentation of results through scientific research articles, conferences, as well as the educational activities reinforcing the visibility of host laboratory, Le Studium and JUNON programme. Therefore, the work carried out during this fellowship represents both a completed mission and a starting point for joint research and innovation efforts.

6- Articles published in the framework of the fellowship

Published articles

1. Sharma, A., **Sharma, A.**, Raju, K. N., Keshta, I., & Guo, K. (2025). Synergistic Integration of 5G-enabled Cloud Native Infrastructures With Advanced Technologies for Urban Safety Enhancement. *IEEE Transactions on Consumer Electronics*, 2025. (SCI journal, Impact Factor: 8.6)

Articles in preparation

1. Sharma, A., Ros, F., et al. A Modular Digital Twin Architecture for Integrated Groundwater and Air Quality Monitoring Using Multi-Source Environmental Data.

(Target journal: IEEE Transactions on Geoscience and Remote Sensing / IEEE Transactions on Artificial Intelligence)

2. Sharma, A., Ros, F., et al. From Environmental Data to Operational Digital Twins: A Review of Architectures, Models, and Challenges in Subsurface Systems

(Target journal: Earth-Science Reviews / Environmental Modelling & Software)

7- Acknowledgements

The author would like to express sincere thanks to **Le Studium - Loire Valley Institute for Advanced Studies**, for funding and supporting this research fellowship, and for providing stimulating research environment in the Centre-Val de Loire region. A special, thanks to **Prof. Frederic Ros**, for his continuous guidance and active involvement throughout the development. His expertise played a key role in structuring Digital Twin framework and guiding technical evolution within the JUNON programme. The author also gratefully acknowledges **BRGM (Bureau de Recherches Géologiques et Minières)** for providing access to data, infrastructure, and expertise essential for the development and validation of Digital Twin. Particularly, sincere thanks are extended to **Hoang Vy Thuy Lynh, Yann Dantal, and Sébastien Dupraz** for the collaboration, several discussions and their valuable support during the integration and deployment phases. A special acknowledgement is due to **Félix**, whose initial architectural design and first version of

Digital Twin prototype laid the solid foundation upon which the present work was built and extended, notably through the integration of additional environmental twins and more advanced analytical services. Author, also thanks **Sophie, Marie-frédérique** and **Aurélien** from Le Studium for their administrative support and assistance throughout the fellowship period. Finally, author acknowledge the JUNON ecosystem, including the PRISME laboratory, and partner institutions for providing a collaborative framework that made this interdisciplinary work a success.

8- References

- [1] Grieves, M., & Vickers, J. (2016). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In *Transdisciplinary perspectives on complex systems: New findings and approaches* (pp. 85-113). Cham: Springer International Publishing.
- [2] Li, X., Feng, M., Ran, Y., Su, Y., Liu, F., Huang, C., & Guo, H. (2023). Big Data in Earth system science and progress towards a digital twin. *Nature Reviews Earth & Environment*, 4(5), 319-332.
- [3] Mylonas, G., Kalogeras, A., Kalogeras, G., Anagnostopoulos, C., Alexakos, C., & Muñoz, L. (2021). Digital twins from smart manufacturing to smart cities: A survey. *Ieee Access*, 9, 143222-143249.
- [4] de la Hidalga, A. N., Hardisty, A., Martin, P., Magagna, B., & Zhao, Z. (2020). The ENVRI reference model. In *Towards Interoperable Research Infrastructures for Environmental and Earth Sciences: A Reference Model Guided Approach for Common Challenges* (pp. 61-81). Cham: Springer International Publishing.
- [5] Anderson, M. P., Woessner, W. W., & Hunt, R. J. (2015). *Applied groundwater modeling: simulation of flow and advective transport*. Academic press.
- [6] Song, Y., & Wu, P. (2021). Earth observation for sustainable infrastructure: A review. *Remote Sensing*, 13(8), 1528.
- [7] Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2020). Integrating physics-based modeling with machine learning: A survey. *arXiv preprint arXiv:2003.04919*, 1(1), 1-34.
- [8] Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of economic surveys*, 37(1), 76-111.
- [9] Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical transactions of the royal society a: mathematical, physical and engineering sciences*, 379(2194).
- [10] Shen, C., & Lawson, K. (2021). Applications of deep learning in hydrology. *Deep Learning for the Earth Sciences : A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*, 283-297.
- [11] Batty, M. (2018). Digital twins. *Environment and Planning B: Urban Analytics and City Science*, 45(5), 817-820.