


## Digital transformation for the water industry: how a data-driven business intelligence platform can improve operations

Alessandro Bettin 

Venezia, VE 30172, Italy

E-mail: alessandro.bettin@gmail.com

 AB, 0000-0002-0372-8852

### ABSTRACT

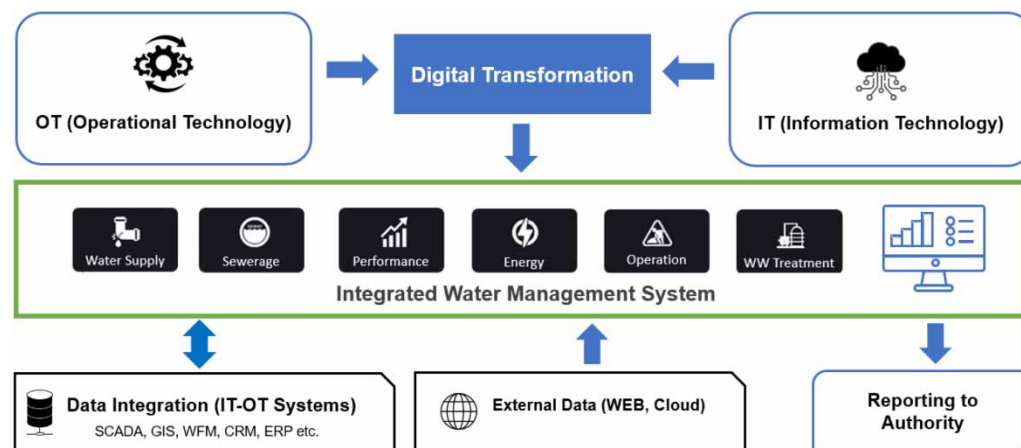
The management of a water network is a complex and challenging job. Nowadays, thanks to the technology available, there is the opportunity to move to a data-driven decision process. To meet the increasingly stringent technical quality requirements from the Authority, water utilities need to increase the operational efficiency. The solution is to integrate and organize operational and business-related data into a common data lake and use these data as a source for a business analytics software able to assist and support daily operations and future investment programs. This paper describes the activities and results of a digitalization case study carried out in Italy where a leading water utility serving over 2 million residents implemented an integrated water management system to enhance the business performance by leveraging a data-driven approach. The project resulted in the establishment of an advanced control room at the company's headquarters, featuring real-time dashboards that are fully integrated with the utility's operational tools and software. The integrated system allows faster responses to operational needs and supports business decisions through streamlined access to relevant data.

**Key words:** business intelligence, data integration, digitalization, utility management, water management system

### HIGHLIGHTS

- Thanks to the technology available, water utilities can move toward a data-driven decision-making process.
- The collaboration between OT and IT is the key to digital transformation.
- The solution is to integrate and organize operations and business-related data into a data lake to allow a holistic business view.
- A business intelligence platform is a tool to support operations and optimize investment.

### GRAPHICAL ABSTRACT



This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

## INTRODUCTION

Digitalization is a process for business development, where digital solutions are used for automation and innovation (Arnell *et al.* 2023). In a challenging environment characterized by water resources deterioration due to climate change, water utilities are facing increasing expectations on the level of service and operation efficiency.

Network failures can bring pollution events, leakage, and interruption of water supply due to unforeseen problems. The use of data as an asset to optimize performance and activate before an incident occurs is becoming increasingly important for water companies. The good news is that there is already a huge amount of data available from operational and IT systems.

To move to the next level, water utilities should find the path to extract value from the data they already have, making intelligent connections and analysis, to improve technical, operational, and financial management. Data can be useless and misleading without the possibility to analyse the right information at the right time.

Data integration is the problem of combining data residing at different sources and providing the user with a unified view of these data (Lenzini 2002). There is a manifold of applications that benefit from integrated information. For instance, in the area of business intelligence (BI), integrated information can be used for querying and reporting on business activities, for statistical analysis, online analytical processing (OLAP), and for data mining to enable forecasting, decision-making, and enterprise-wide planning (Ziegler & Dittrich 2007).

Gruppo CAP, a leading water utility that manages over 2 million inhabitants in Northern Italy, decided to implement an operational intelligent system through a public tender. The objective was to develop a BI platform that would assist the water utility in making faster, more informed decisions to enhance customer service, deploy maintenance solutions more effectively, and comply more efficiently with industry regulations.

The project has been developed with the aim of completing a journey from 'silos' to an integrated platform. 'Silo' refers to the data held by a department that is not fully visible or accessible to other departments of the same organization. Siloing can be seen as the exact opposite of integration. An integrated data analytics platform works out to be a big plus for any organization. Having data from different departments in an organization be visible and accessible to any department is the goal (Pal 2022).

The solution selected to achieve this challenging objective was to integrate all relevant information into a data lake, namely a container of data from different sources, and to analyse and process these data using a dedicated BI platform. The outcome is a water management system able to drive operations and maintenance with the help of dashboards, key performance indicator (KPIs) visualizations, trends, correlations, reports, and process analytics.

### Benefits from data integration

Data integration provides water utilities with a comprehensive view of their water systems, enabling them to optimize resource allocation, streamline processes, and make informed decisions on operations, maintenance, and customer service. It also allows for early detection of potential issues, which can be addressed proactively. By integrating data from multiple sources, water utilities can improve efficiency and sustainability while reducing costs and improving service delivery. Moreover, data integration and analytics can give substantial help to avoid or manage intermittent water supply and reduce water loss.

To meet the increasingly stringent technical quality requirements from the Regulatory Authority, water utilities need to turn the current business into more efficient processes based on quality measurable standards. Reporting to Authority requires a massive data acquisition, analysis, and elaboration, including safe data storage. An integrated digital platform, with a central database and a set of predefined reports and visualization, allows water utilities to save time and resources in the fulfilment of these requirements and to avoid fines for missing technical and water quality standards.

### Problems with data integration

The process towards data integration is not easy. Several problems can be encountered and must be overcome. First of all, there are different sources, different formats, and different proprietary software. Moreover, a problem of data quality can be present. In fact, many data can be duplicated across software, and it is common that the same single data is present in different software but with different formats and syntax (i.e. a street address). Manual data input can also bring frequent errors causing problems when dealing with geocoding and geographical information system (GIS) representation of events and elements.

In general, information systems are not designed for integration. Thus, whenever integrated access to different source systems is desired, the sources and their data that do not fit together have to be coalesced by additional adaptation and reconciliation functionality (Ziegler & Dittrich 2007).

Another problem is dealing with too much data. This problem is amplified when collecting data from multiple channels without a proper data management system in place. With the sheer amounts of data being created daily, it becomes a big challenge to manage, analyse, and extract value from data when the signal in the noise cannot be found (Campos 2022).

The process of data integration from different software requires combining data from various sources into a centralized database. However, before loading the data into the database, it is necessary to perform preliminary data management of the different sources to select, prepare, and verify the data. When integrating with a GIS, several problems may arise, for instance, if the customer relation management (CRM) system reports low water pressure, while the WFM system shows that a leak has been repaired near the customer's home, the issue is considered resolved. This correlation is only possible if both data sources, the CRM and work force management (WFM), are connected to the GIS using the same data model. Otherwise, the territorial correlation is missing. To optimize the data integration process, these preliminary steps for each data source should be followed.

- **Data profiling:** Analysing the data sources to understand their structure, format, quality, and completeness. This step helps identify any data issues or inconsistencies that need to be addressed before integration.
- **Data cleaning:** Identifying and correcting errors, inconsistencies, and inaccuracies in the data sources. This step may involve removing duplicate records, standardizing data values, and correcting data errors. For example, street addresses and geocoded information must be uniquely determined, and manual input must be cleaned.
- **Data transformation:** Converting data from one format to another or reformatting data to ensure consistency and compatibility with other sources. This step may involve data mapping, data conversion, and data normalization, such as reformatting comma-separated values (CSV) data from an external source before loading it into the database.
- **Data enrichment:** Enhancing the data with additional information to provide more context and value. This step may involve merging data from external sources, performing calculations, or applying business rules to the data. For example, to automatically calculate the water balance of each DMA, the flow meter codes must be the same in the enterprise resource planning (ERP) software (where the monthly water volume is reported), in the supervisory control and data acquisition (SCADA), and in the CSV file that associates each flow meter to a DMA.
- **Data validation:** Data validation involves checking the accuracy and completeness of the data after cleaning, transformation, and enrichment. This step helps ensure that the data is reliable and suitable for integration.

By performing these preliminary data management steps, the data are prepared for integration with other data sources. This helps ensure that the integrated data are accurate, consistent, and of high quality, which can improve decision-making and business outcomes.

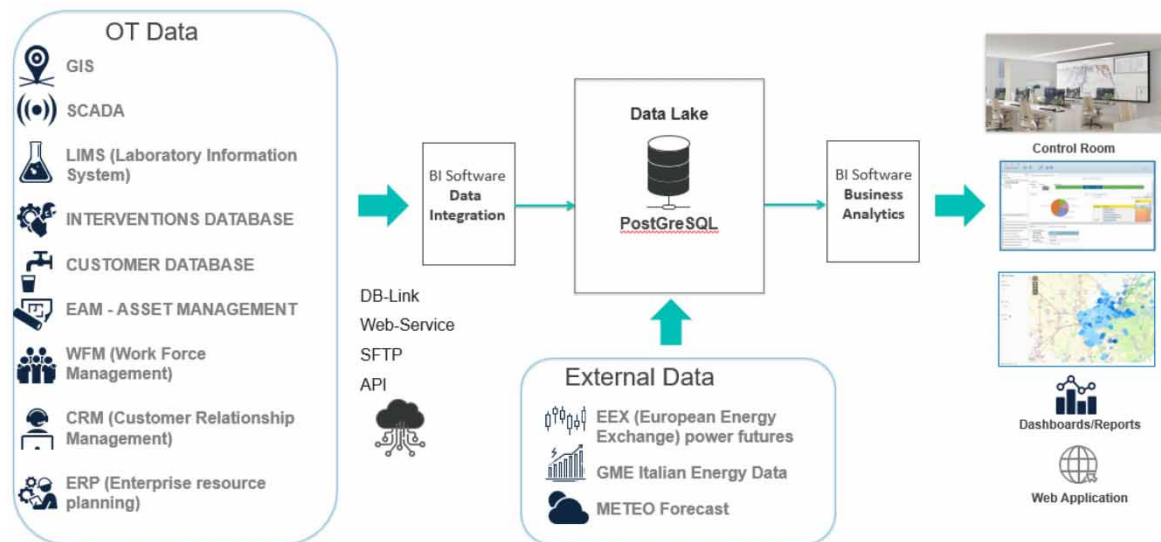
To achieve a successful digital transformation, it is important to follow a holistic approach that combines OT (Operation Technology) and IT (Information Technology). This approach will allow the potential of artificial intelligence and big data to be fully utilized for operations and management. OT experts, such as hydraulic engineers, with industry knowledge and utility experience, are essential for providing technical guidance and specifying data requirements and business visualization goals. The IT team, including data scientists, database experts, and software engineers is necessary for coding and analytics and to ensure that the integrated platform is fully operational. Collaboration between the project team and the water utility business owners, represented by managers and technical staff running different clusters, is also critical for the success of the digital integration project.

### **The implementation of the integrated water management system**

The integrated water management system has been developed using a BI software that integrates open-source components and open standards with a process-driven engine. The BI platform provides an execution framework and services that include logging, auditing, security, scheduling, ETL (Extract, Transform, and Load data into the database), and web services. The end-user BI capabilities include reporting, analysis, workflow, dashboards, and data mining. The data integration module has been used for the integration, preparation, and blending of enterprise data, while the business analytics module has been utilized for dashboards development and data analytics. There are several BI software alternatives in the market. The decision on the best solution should be based on the

comparison of specific competencies in the following categories: product capabilities, licencing cost, service and support, integration, and deployment.

Figure 1 reports the schematic software architecture of the water management system showing the detailed connections between OT data and the BI platform with the dashboard representation in the control room. Several data sources from different software have been integrated: SCADA (more than 100,000 signals), Web-GIS, CRM, WFM for operations, billing system, asset management, ERP, and LIMS (laboratory information management system). Moreover, to allow further analysis and forecast, some external data have also been integrated into the data lake: meteorological data, EEX (European Energy Exchange) power futures data, and energy data from the Italian market.



**Figure 1** | Schematic software architecture of the water management system implemented.

All data sources have been integrated using the most appropriate technology among DB-Link, Web-Service, SFTP (secure file transfer protocol) or API (application programming interface).

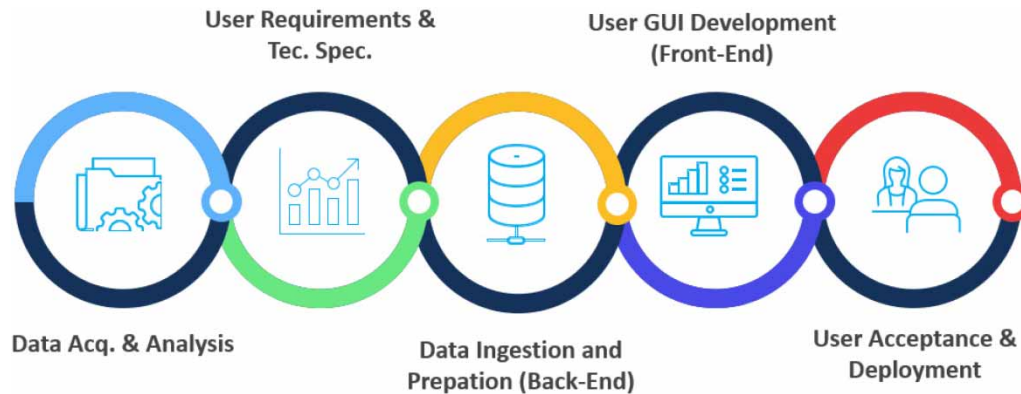
The project outcome is a fully operating intelligent control room that is used 24/7 to manage network operations and prompt intervention, to optimize activities of the operations and maintenance teams and to monitor real-time anomalies and alarms.

The intelligent platform covers the overall system under management: 154 municipalities, 6,400 km of water supply network, 6,600 km of sewerage, and 40 wastewater treatment plants.

## METHODOLOGY

The integrated platform has been developed through five main phases (see Figure 2).

- Data acquisition and analysis: definition of dataset structure and content, understanding of data fields and records, selection of data access procedures. Data management is a prerequisite to organize a suitable data model.
- User requirements and technical specifications: analysis of business requirements, design of dashboards, and tools for visualization and reporting.
- Data ingestion and preparation (back-end). This is the back-end activity aimed at data ingestion, data blending and preparation, creation of the data lake, and preparation of data marts for dashboards implementation. For this task, PostgreSQL has been used. PostgreSQL is an open-source object-relational database management system (ORDBMS) developed at the University of California at Berkeley Computer Science Department (Postgresql.org. 2022).
- User graphical user interface (GUI) development (front-end): dashboard developments, testing of dashboard functionalities with sample data, collection of user's feedback, bug fixing.



**Figure 2** | Process methodology, main steps for data integration, and dashboards visualization.

- User acceptance and deployment: user final acceptance test, dashboards deployment in the production environment. At this stage, the dashboard is ready for user utilization.

The visualization of dashboards can be in real-time or time-delayed, depending on the refreshing time of data sources and the dashboard's characteristics. Some visualizations are updated once per day and some others at the end of each month. Several reports/visualizations have been created to allow a deep historical analysis of KPIs, time-series, and events and to find out correlations and anomalies.

The implementation phase, including software and hardware supply and installation, was scheduled to be completed within 20 months. The contractor who won the tender issued by the water utility (the client) has successfully set up the system as described in Figure 2. Since the integrated platform was designed to serve all utility business clusters, support from software development groups working in parallel for ETL and front-end implementation was necessary. Additionally, a team of engineers was fully involved in user requirements, technical specifications, and business consulting with the client. The contractor's project team was complemented by the project management office and technical support staff, including data scientists, software architects, and IT engineers.

The project was developed from the early stages following a co-creation mode together with the client, with the participation of the technical and managerial personnel of the water utility required for tasks such as preliminary data management, user requirements, and acceptance testing. The utility's staff also included the participation of personnel from the IT and software departments.

The project also included the supply and installation of servers and storage to run the BI platform, as well as monitors, software, servers, and equipment for the new control room equipped with a video wall of approximately 12 m<sup>2</sup> (18 monitors of 49 inches each).

Generally, the hardware required can be reduced if a cloud solution is chosen to host the platform, rather than an on-premises implementation. The scope of the project, including the number and complexity of use cases, the type and amount of data to be integrated, and the number and type of software used to collect this data, will all affect the effort and cost required to complete a project of this kind.

The water management system can be implemented in any network, regardless of its size. Thanks to the scalability of the solution, it can be set up in different phases to accommodate the future needs of the water company. The annual maintenance cost for such systems ranges from 10 to 20% of the initial supply cost depending on the level of support required.

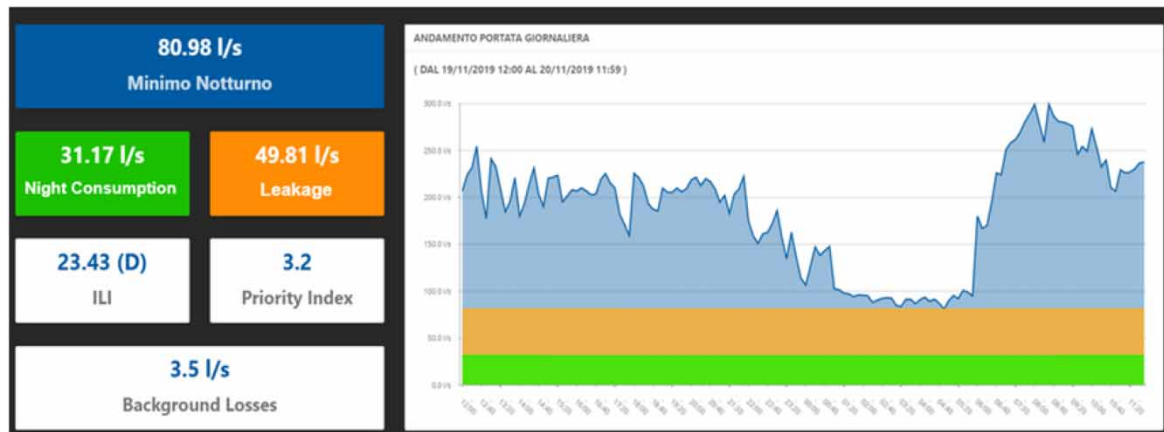
### Cluster web applications

Several dashboards, in the form of web applications available for any browser and media, have been developed to cover the entire business of the water utility: water distribution and leakage management, water production, performance, energy, sewerage and storage tanks, wastewater treatment, prompt intervention, telemetry system.

Regarding the water distribution cluster, the platform supports several features for leakage management: real-time monitoring of flow, pumps operation, data quality and correction, NRW (non-revenue water) monitoring in each DMA (district metered area), the priority of intervention considering cost-benefit analysis (Bettin & Rogers

2012), detailed analyses of leaks repaired and water recovered, leakage trend, KPIs for improving network efficiency, benchmarking analysis, and energy performance indicators. It is also possible to plan maintenance interventions and understand in advance which pipes require replacement or rehabilitation. The system can identify and correct functional anomalies and measurement errors. The benefits are therefore both economic and environmental.

The dashboard depicted in Figure 3 provides a tool to conduct a daily analysis and measurement of leakage within any DMA under management. The dashboard allows for quick visualization of critical areas requiring intervention based on the economic level of leakage (ELL) (Tripartite Group 2002).



**Figure 3** | Dashboard for leakage analysis DMAs. All metrics to identify leakage are calculated. The priority index highlights DMAs where a leakage reduction intervention is needed based on the ELL.

Dashboards and reports have been developed to monitor and analyse the trend and the forecast for the following macro-indicators selected by ARERA, the Italian Regulator, to monitor the performance of water companies determining their minimum targets (ARERA 2017):

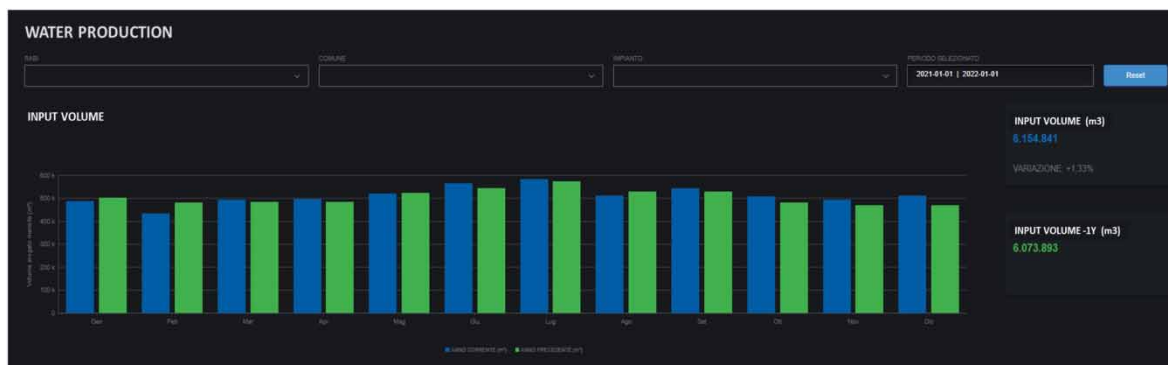
- M1 'Water losses', associated with the objective of containing physical and commercial losses, are defined by taking into account both linear water losses and percentage losses.
- M2 'Service interruptions', associated with the objective of maintaining service continuity, are defined as the ratio between the sum of annual interruption durations and the total number of customers served by the water utility.
- M3 'Quality of water supplied', associated with the objective of providing adequate water quality for human consumption, is defined considering the incidence of non-potability ordinances, the rate of non-compliant samples, and the rate of non-compliant parameters.
- M4 'Adequacy of the sewer system', associated with the objective of minimizing the environmental impact from sewerage, was defined as considering the frequency of flooding and/or overflow from sewers and the regulatory adequacy and control of combined sewer overflows.
- M5 'Disposal of sludge in landfill', to minimize the environmental impact of the wastewater treatment with regards to sludge line, defined as the ratio between the amount of dry sludge disposed of in landfills and the total amount of dry sludge produced.
- M6 'Quality of treated water', also related to wastewater treatment, measures the percentage of discharged treated wastewater samples that exceed legal limits for selected water quality parameters.

Figure 4 displays a dashboard that has been designed to assist in the implementation of a water safety plan. This module analyses the potable water quality in the water supply system, calculates the related macro-indicator M3 and its trend, and highlights any possible anomaly of noncompliance.

Several visualizations for the analysis of the water supply from wells and sources have been created. Figure 5 provides an example of a dashboard designed to analyse the monthly water production for each distribution area, from the DMA level to the overall system. This tool enables the operator to forecast the production amount considering seasonal trends and possible correlations with meteorological conditions.



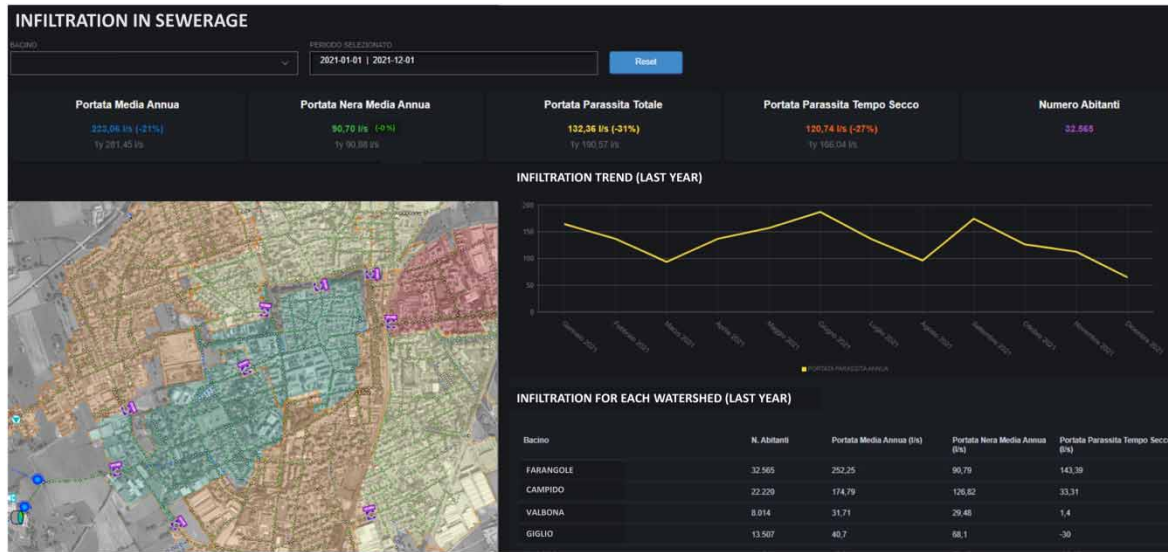
**Figure 4** | The dashboard shows water quality KPIs in water networks related to sampling activity and laboratory analysis. Based on the number of not acceptable water samples and water quality parameters, it calculates the class for each municipality (from A to E with A being the best).



**Figure 5** | Water production with monthly input volume chart and comparison with the previous year.

The production of drinking water involves significant energy costs for water companies mainly due to water pumping. To reduce these costs, it is important to maintain a high level of efficiency in the pumping stations with optimal operating levels. For this purpose, a dashboard has been set up for continuous monitoring of the main operating parameters of the pumps to identify possible anomalies in advance in a predictive maintenance perspective.

For sewerage and wastewater treatment clusters several dashboards have been deployed: discharge forecasts based on rain events, optimization of wastewater treatment processes, monitoring of spillways and floods, analysis of the probability of spill in wet and dry weather, analysis of storage tanks operation, and analysis of sewerage infiltration. Sewerage infiltration is one of the most important threats to the correct operation of the wastewater treatment plant (Beheshti *et al.* 2015). Groundwater can infiltrate through defective pipe joints or from surface water getting into the sewer system through dual manholes (WRc 1994). As reported in Figure 6, a dashboard showing the amount of infiltration in each watershed and its monthly trend has been developed. Infiltration is calculated as the difference between the measured flow volume (from permanent equipment connected to the SCADA) and the rate of users' consumption, elaborated from the billing system, that is discharged in the sewer system. The outcome is the list of pipelines with higher infiltration where further investigations and/or rehabilitation interventions are recommended.



**Figure 6** | Infiltration analysis for the combined sewerage system. Based on the data coming from the SCADA, the dashboard analyses infiltration trends for each watershed.

An additional application pertains to the analysis of the energy market (refer to Figure 7). The platform includes the charts of the main EEX (european energy exchange) power futures contracts with different expiration dates, coupled with graphical tools to analyse the market trend and momentum like moving averages (21, 50, 200 periods) and historical volatility of daily price variation (Natenberg 1994) plus futures volume and open interest. This application aims to assist the water utility in identifying the most favourable price for forward buy contracts mitigating the risks associated with energy cost fluctuations.



**Figure 7** | Analysis of EEX power futures prices for different futures expiration.

One of the most important objectives of the water company is to achieve a high level of operational efficiency and process automation. For this purpose, a module for prompt intervention monitoring and optimization, fully integrated with GIS and traffic data, has been developed. The dashboard supports daily operation and maintenance with real-time monitoring of prompt interventions grouped by the following categories: anomalies from the telemetry system, requests for interventions made to the call centre, alarms related to noncompliance in potable water quality, ongoing interventions being carried out by the water company teams, and ongoing construction sites managed by external contractors.



## CONCLUSIONS

The presented tools and methodology aim to shift the water utility business towards an integrated decision support platform as demonstrated by a real case study. This project is likely the first and most comprehensive application of its kind in Italy.

Among the project phases, data analysis, selection, ingestion, and preparation are the most demanding tasks. This activity requires a significant effort and full collaboration between the project team and the water company business owners. To succeed in the digital transformation process, it is also mandatory to activate an effective collaboration between hydraulic engineers (industry experts) and experts from the IT world such as software engineers and data scientists.

Due to climate change and increasing demand, water companies face additional difficulties and higher costs to maintain a good level of services and high standards. In this challenging environment, the acceleration toward a fully digital transformation will help water utilities to optimize operations and reduce costs through the automation of repetitive tasks. The integrated platform will also support Gruppo CAP in the planning of investments through the prioritization and selection of the most effective solutions.

## DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

## CONFLICT OF INTEREST

This paper has been produced with the consent of Gruppo CAP.

## REFERENCES

- ARERA 2017 RQTI Regolazione della Qualità Tecnica del Servizio Idrico Integrato [RQTI Technical Quality Regulation of the Integrated Water Service]. Available from: <https://www.arera.it/allegati/docs/17/917-17slide.pdf> (accessed 10 December 2022).
- Arnell, M., Miltell, M. & Olsson, G. 2023 Making waves: a vision for digital water utilities. *Water Research X* **19**(2023), 2.
- Beheshti, M., Sægrov, S. & Ugarelli, R. 2015 Infiltration/inflow assessment and detection in urban sewer system. *Vann*, **01**, 24–34.
- Bettin, A. & Rogers, D. 2012 Autoleak evaluating leakage economically and environmentally – conference proceedings. In *Water Loss Europe 2012 IWA Specialist Conference*, 23–25 May 2012, Ferrara, IT.
- Campos, L. 2022 *Common Data Integration Challenges (And How to Solve Them)*. Blog.hubspot.com. Available from: <https://blog.hubspot.com/website/5-common-data-integration-challenges-and-how-to-solve-them>.
- Lenzini 2002 Data integration: a theoretical perspective. In *Conference Proceedings of the Twenty-First ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, June 3–5, Madison, WI, USA.
- Natenberg, S. 1994 *Options Volatility and Pricing*. McGraw-Hill. ISBN 1-55738-486-X. Printed in the United States of America.
- Pal, K. 2022 *Destroying Silos with Integrated Data Analytics Platforms*. Techopedia.com. Available from: <https://www.techopedia.com/data-management/destroying-silos-with-integrated-data-analytics-platforms/2/31431> (accessed 15 July 2022).
- Postgresql.org. 2022. Available from: <https://www.postgresql.org/files/documentation/pdf/13/postgresql-13-A4.pdf> (accessed 15 July 2022).
- Tripartite Group 2002 *Best Practice Principles in the Economic Level of Leakage Calculation, Report*.
- WRc 1994 *Sewer Rehabilitation Manual Third Edition*, WRc plc – ISBN-1-898920 001.
- Ziegler, P., Dittrich, K. R., 2007 Data Integration – Problems, Approaches, and Perspectives. In: *Conceptual Modelling in Information Systems Engineering* (Krogstie, J., Opdahl, A. L. & Brinkkemper, S., eds). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-72677-7\\_3](https://doi.org/10.1007/978-3-540-72677-7_3).

First received 27 September 2022; accepted in revised form 30 May 2023. Available online 12 June 2023