TOWARDS AN INTEGRATED PLATFORM FOR DECISION SUPPORT IN WATER UTILITY MANAGEMENT

Cédric Grueau[1], André Antunes[1], Bruno Ferreira[2], Miguel Gonçalves[1], João Gomes[1]
and Nelson Carriço[2]

[1]Setúbal School of Technology, Instituto Politécnico de Setúbal
Campus do IPS, Estefanilha, 2910-761 Setúbal, Portugal
[2]Barreiro School of Technology, Instituto Politécnico de Setúbal,
Rua Américo da Silva Marinho, s/n, 2839-001 Lavradio, Portugal

ABSTRACT

The pressure on water supply systems is a real challenge for water utilities, which are entrusted with the responsibility to manage and operate those systems. Water utility engineers and managers make decisions based on data that are obtained from several information systems (IS), such as geographic information systems (GIS), billing, asset operation and maintenance or accounting. Most of the time, these IS result in data silos as a consequence of lack of machine communication protocols, common data formats or common meaning of exchanged content. On the other hand, the diversity of IS is a challenge to water utilities which struggle with data consistency and interoperability. In this context, water utility engineers and managers make decisions based on data which are often incomplete, inaccurate or out-of-date. In this paper, we present an IT platform that integrates data from different information systems, including tools to support decision-making on water supply systems (e.g., performance indicators, water, and energy balances). This platform, developed in collaboration with a representative set of Portuguese water utilities, allows water utilities to integrate their data and perform analysis of their systems. With this approach, we provide water utilities and water supervision agencies with the possibility to compute a unified set of metrics whatever the IS they use.

KEYWORDS

Water Utility Management, Web Services, Data integration, Decision Support System, Information Systems

1. INTRODUCTION

In recent decades the percentage increase in water use on a global scale has exceeded twice that of population growth (Cosgrove and Loucks, 2015). Water utility companies, under pressure from consumers and regulators alike, are transforming and adapting their management models to improve water quality and increase supply while protecting water sources. Those organizations often rely on a vast software eco-system to manage the board set of data generated by utility's activity. Substantial investments are being carried out by water utility companies in several Information Systems (IS), namely, Enterprise Resource Planning system (ERP), Geographic Information System (GIS), Customer Relationship Management (CRM) or Supervisory Control and Data Acquisition System (SCADA). These IS generate data that ranges several areas of utility activity, namely, billing and accounting data, GIS data, time series of water volume, pressure, discharge, tank levels, consumption and water quality parameters, such as chlorine, pH and temperature, infrastructure and user characteristics and meteorological data (Carriço et Al., 2018).

The daily control, operation, management and current or future planning carried by water utility companies often resort on several analysis techniques (e.g., calculation of performance indicators or simulation of hydraulic behaviour, as well as more complex Artificial Intelligence (AI) techniques). These different analysis techniques require data that is produced by utility's multiple IS and is often distributed across several decentralized databases, where little interaction occurs between.

The required interoperability and integration between the different utility's IS habitually face many barriers. Systems often don't share information, creating data silos that leave various departments without a holistic view of operations or customer behaviour. The dispersion and variety of data sources and formats is pronounced, resulting from diverse software, systems and departments, as well as the use of outdated technological trends (i.e., use of old technologies based in spreadsheets).

Considering the lack of machine communication protocols, common data formats or common meaning of exchanged content, the access to data from a particular IS or department can be a time-consuming task, typically involving several individuals, multiple file formats and several intermediate processes. Thus, the application of information and communication technologies in water utilities is often restricted, foreclosing the ability to share data and knowledge, and hence interoperate across people and software involved (Howell et Al., 2016).

2. RELATED WORK

Geographical Information Systems (GIS) play a very important role in managing spatial data within the network infrastructure. (Meyers, 2005). Each GIS uses its own data model in order to represent the water facility and provides toolboxes to perform network analysis tools and planning. For this reason, if one wants to compare different systems, he will eventually need to convert data from a system to the GIS solution to perform the comparison.

Several solutions have been developed to solve the data integration and analytics problem in water utilities. The development of this type of solutions may start within the utility itself, which has the necessary resources to develop a system suited to their needs, or through the development of commercial software.

Two Portuguese utilities developed their data integration and analytics problem through GIS driven solution, as a mean of integration between their diverse IS. As such, through this solution it is possible to easily access data related to infrastructures, fleets, water demands, billing, among others (Silva et Al. 2017, Sousa et Al. 2017). H2Porto is a platform developed by a Portuguese utility aiming at integrating data gathered from different sources (i.e., sensors, GIS, equipment automation signals, among others), as well as results from hydraulic simulation models, aiming ultimately at predictive models of network hydraulic behaviour and decision support tools (Vieira et Al., 2018). Nonetheless, the use of both of these platforms is restricted to the utility itself.

Baseform and WaterSmart are commercial software that claim to provide comprehensive, actionable analytics to assist in water utilities specific efficiency issues and key business outcomes. Nonetheless, even if integration of real sensor data is possible with these solutions, integration of raw data from the scattered data sources may not be possible and the processing treatment required to import and integrate data may not be justified. On the other, if these solutions allow the analysis of a single system, they are not prepared to integrate utilities with different levels of maturity and encompassing different management models.

A dashboard is visual display of the most information needed to achieve one or more objectives which fits entirely on a single computer screen so it can be monitored at a glance (Few, 2006). Various providers such as Tableau or DataHero provide solutions to build dashboards. These solutions implement predefined powerful visualisations that can be combined into a dashboard. To get more flexibility, it is also possible to use visualisation libraries such as D3 JS or Bokeh to create building blocks for a dashboard. These libraries are more suited for the creation of specific visualisations that can cope with the specificity of our domain.

3. METHODOLOGICAL CONCEPT AND SYSTEM DESIGN

In the context of the DECiDE project, a new platform has been designed from scratch. To cover this wide range of utilities, a set of five Portuguese water supply utilities was selected, ranging from small to medium size and from countryside to coastal utilities, encompassing different management models (i.e., Direct municipal management and Autonomous water services of a municipality). These five utilities form a representative sample of the diversity of IS and data gathering methods among Portuguese utilities. As an example, one of these utilities is able to collect consumption data at the costumer level every hour, while the remaining obtain the consumption by manual readings every two months.

3.1 Concept

For improved decision making, water management companies need to compute a set of metrics that help to manage the physical system and the commercial operation. The Portuguese Water and Waste Services Regulator (ERSAR) published several technical guides (ERSAR, 2019), outlining a water supply service quality assessment system based on metrics. A set of sixteen metrics was selected by the five case studies, reflecting the real needs of utilities for systems' performance assessment. The calculation of these metrics resort on data from different IS. Water and energy balances are also carried on a regular basis by the utilities, in order to assess the water supply system's efficiency, both in terms of water losses (Alegre et al., 2000) and energy use (Mamade et al. 2017). Nonetheless, these balances resort on a broad set of data (i.e., input volumes and energy consumption, billed and unbilled consumptions, elevations, amongst others). If sectorial balances are meant to be calculated, the complexity of data collection and management carried by the utility manager can be overwhelming.

As such, water utilities need a system to assess and compare the performance of its water supply systems. The basic question therefore to answer is: how can we provide a unified approach to integrate and compare data organized and stored in a plethora of IS?

Thus, the aim of this platform is to implement a set of services to upload water utilities data and parse those data into a unified database where data from all water utilities are stored. Using web services capabilities, the system design in Figure 1 creates a platform for Spatial and Consumption Data through the use of webservice RESTful API. One of the API is dedicated to serve the values of management metrics for each utility. Another of the API serves the dashboard application that is used by decision makers to visualize and analyse metrics from each water supply system.
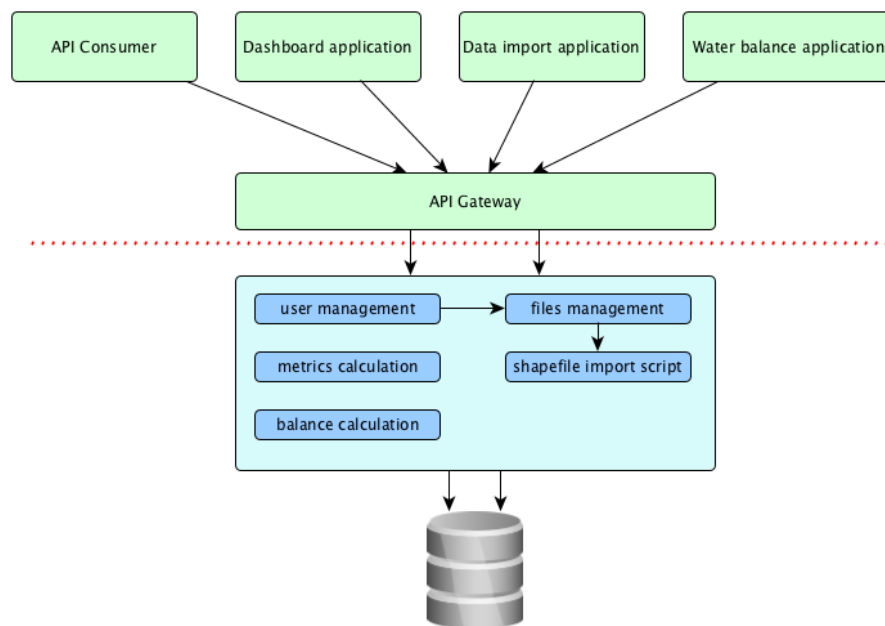

Figure 1. System architecture for the DECiDE web-based decision support platform for water utility management

The dispersity of data in the various utilities' information systems, as well as the high number of performance metrics related to water supply systems, motivate the need of a system that can graphically represent this type of data in order to facilitate the decision-making process.

The dashboard is one of the tools that is intended to facilitate the decision-making process, based on user-defined filtering (e.g. analysis area, data type and time period), it is possible to represent a set of data with different types of graphical representation (e.g. bar chart, pie chart, timeline chart, bullet chart, map). The system evaluates the best types of representation for each data set.

3.2 System Design

The first challenge to overcome was to design a domain model that could represent all the concepts present in the different water utilities. Due to the heterogeneity of IS and data models used by those utilities, we followed a Domain Driven Design in order to map the domain concepts into software. It was necessary to intensively involve domain expert (researchers and engineers from water utilities) into the process of domain modelling, in order to resolve ubiquitous language, establish a common vocabulary and draw the context boundaries. The result of the expert participation is a domain model that has been enriched with a set of sixteen management metrics about water supply systems, as well as a set of recommendations regarding data management in order to assess the performance of their systems.

3.3 System Overview

After a registration process, water utilities upload their infrastructural data from their own GIS data and their consumption data into the cloud platform using a Web application. The file format required for geographical data is shapefile, a format used as a standard by GIS community. Data are parsed and stored into a database. A set of services, built in the platform, uses data from the physical network and from clients' consumption to compute performance and sustainability metrics about the network, as well as water and energy balances. This information is organized and made available to the decision maker of each water utility through a dashboard. Using this dashboard, managers can assess the health and performance of their systems by comparing metrics for different time periods. The values of those metrics are also made available to public through a REST API. This makes it possible for the regulator to compare the performance of different water utilities at a regional scale, increasing the relevance of the proposed platform.

For the implementation, we decided to use MySQL technology for the database. We wanted to build a flexible system that could be used by utilities with different levels of maturity. Consequently, it was a requirement to be able to upload incomplete data to the platform. With relational databases it is easier to check if a given data is missing, i.e., what columns (attributes) exist for a given row. The API server technology was implemented using Django, a free and open source Python Web framework, in order to obtain a fast developed, scalable and secure platform. We used Django REST framework toolkit that supports Authentication to build the first version of the Web API. The API exposes the endpoint for the frontend application. The client application on the frontend of the framework uses D3 JS library. D3 offers a wide range of possibilities to define the specific visualization that were needed to represent the water and energy balances as well as the dashboard components to represent the metrics.

The platform architecture offers a set of directly accessible services. Figure 2 represents the requests reaching the backend platform services. In the figure, there are three servers (implemented with docker containers), which contain an Apache Service that runs a frontend client application (i.e. import app and dashboard app); the Django backend API application and the MySQL database.
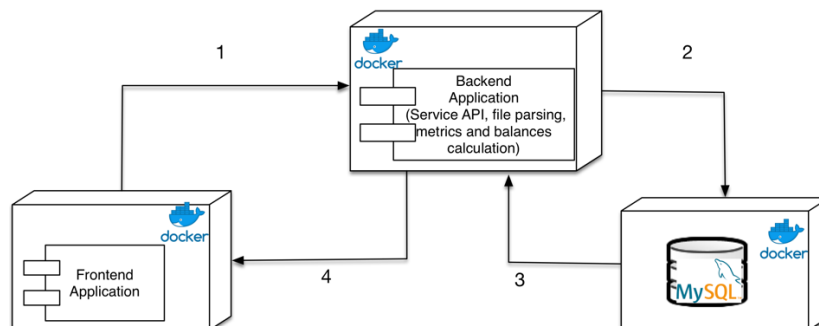

Figure 2. Platform Architecture

3.4 Data Import into the Platform

The data import process is challenging due to the disparity between utilities own information systems and their data storage methods and formats.

The DECiDE platform allows registered entities to upload their updated data files. Each file is registered in the system and placed in a distinct entity folder. Files can be in multiple formats and content organization is not consistent. Our platform must be able to parse uploaded files and update the database with the parsed data. The challenge was to implement a file parsing mechanism allowing a common interface, semi-automatic process and minimum user intervention. File parsing starts by trying to determine file content. Identification is based on file extension and content analysis. Previous knowledge obtained by analysis of partner entities sample data files is included in the detection algorithm. After detecting file type and content, we are able to identify the related data model elements, database tables and formats to target.

An initial mapping is created between data model elements and file content. The user is allowed to correct and/or validate the detected mapping. The validated mapping is registered in the system and used the next time a similar file from that entity is uploaded. After validation, the importation process starts. Data can be converted as needed to be validated in the platform data model.

The data model is prepared to maintain an historical record of infrastructure and other time-sensitive data. When updating such data elements, the previous state is timestamped and kept in database allowing historical analysis of data evolution. To validate the data import process, a specific web-based application has been developed. It includes the calculation of control values about the GIS imported from a shape file.

4. DISCUSSION

Many services and API for the DECiDE platform have already been implemented. The application built to validate the import process have allowed us to highlight the necessity of implementing a two-steps import process. The first step will consist in the importation of all the necessary files to integrate the GIS data and the data related to water consumption. After the parsing of all these data, the user will be informed about the possible missing information, necessary to calculate the set of metrics. He will be notified about the metrics that can be calculated from the uploaded information and the one that cannot. He will then have the possibility to validate or to complete the missing data. This two steps validation process will help water utilities with fewer data to continue to take advantage of the platform even though they cannot provide all the information.

5. CONCLUSIONS AND FURTHER WORK

We have presented the first step towards a platform for decision support in water utility management. Web service technologies were utilized to develop applications on both client and server side, as well as to integrate data from different IS. With our approach, data integration and metric calculation, required to support the decision-making process, is easily achieved by the utility. This approach will also be useful for any regional or national utilities resorting on data integration for combined analysis, such as sanitation services or state irrigation projects. Future work will be carried out towards improving the dashboard application with water utilities' users and developing the water balance and energy balance applications.

ACKNOWLEDGEMENT

REFERENCES

Alegre, H., Hirner, W., Baptista, J.M., Parena, R. (2000). *Performance indicators for water supply services*, 1st edition, Manual of Best Practice Series, IWA Publishing, London, ISBN: 1 900222 272 (160 p.).

Carriço N., Ferreira B., Monteiro L., Mendes A., Grueau C., Covas D., Santos J., Figueira, J., Baio M., Barreira R., Silvestre I., Isidro R., Mestrinho J., Faím J., Traitolas S., Chaveiro P., 2018. "Multicriteria decision platform for urban water infrastructures", In: *Proceedings of 18th ENASB*, Porto, Portugal.

Cosgrove, W. J., Loucks, D. P., 2015. Water management: Current and future challenges and research directions. *Water Resources Research*, 51(6), 4823–4839. American Geophysical Union (AGU). Retrieved from http://dx.doi.org/10.1002/2014WR016869

Few, S., 2006, *Information Dashboard Design*, O'Reilly Media, CA, USA.

Howell, S.K. , Rezgui, Y. , Beach, T., Zhao, W. , Terlet, J., Li. H., 2016. Smart Water System Interoperability: Integrating Data and Analytics for Demand Optimized Management through Semantics, *ICCCBE 2016*, pp. 1-9.

Mamade, A., Loureiro, D., Alegre, H., Covas, D. (2017). A comprehensive and well tested energy balance for water supply systems, *Urban Water Journal*, DOI: 10.1080/1573062X.2017.1279189

Meyers, J. R., 2005. GIS in Utility, In: *Geographical Information Systems: Principles, Techniques, Management and Applications*, 2nd Edition, Abridged, [available online, last accessed on 3/1/2019]. http://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch57.pdf,

Silva M. H., Vilela J., Morais M. (2017). Interconnection of information systems in SMAS Sintra - contribution in providing quality public service (In Portuguese). In: *Proceedings of ENEG 2017*, Évora, Portugal.

Sousa C., Marques A., Fernandes J. (2017). The contribution of GIS information for an integrated and efficient management systems (In Portuguese). In: *Proceedings of ENEG 2017*, Évora, Portugal.

Vieira P., Costa F., Cunha R., Marques R., Fernandes M. (2018). Technological platform for the urban water cycle. From concept to implementation (In Portuguese). In: *Proceedings of 18thENASB,* Porto, Portugal.