**Knowledge and Data Engineering of Large Water Network Systems for Increased Infrastructure Resilience**

*Despite billions of Pounds invested in securing the nation’s precious water resources, mains water cannot be guaranteed even today, and this problem will become more severe[[1]](#footnote-1)”*

The purpose of this project is to investigate and design knowledge and data engineering (KDE) infrastructure for strategic and large scale water and waste water treatment processes (WWTP). By KDE of water treatment operations we mean the building of the essential capabilities for acquisition, analysis and modelling of the data to create the knowledge for large scale (regional, national and international) strategic planning of water supply and consumption. This will be achieved by overcoming some of the existing barriers of interoperability and harmonisation of data and information. For this purpose we will take inspiration and technology currently deployed in Communication Service Providers in Mobile industries who have successfully managed the challenges of interoperability and service reliability for continuous service provision. In specific the 5G as infrastructure that will enable remote telemetry and real time data availability. The proposed technology will link to the climate driven volatility in clean water supply to help long term planning and mitigation plans – creating an event-based logical and objective link between capabilities (infrastructure) and the impact of the environmental changes. For example, the study by [1] links the predicted demand for clean water with the reduction of abstraction licenses, where there will be a potential loss of 5-50% of available water that is supplied by five main UK suppliers.

At plant level the Water and Waste Water Treatment industry is experiencing an evolutionary growth, thanks to the investment in bespoke ICT for plant real-time control and performance measurement technologies. Currently, a substantial level of R&D is being directed towards building the data acquisition and modelling tools to measure and control the industry’s key performance indictors (KPI). In this project we will concentrate on identifying and developing the most appropriate tools for measuring four KPIs for the WWTP. They are: Carbon Footprint, Energy Consumption & Generation, the yield of the By-Products of the processes, and the Overall Performance of the Equipment and Reliability (Productivity) [REFERENCES TO PROJECTS PAPERS AND WORKS]. The approach is to create the information modus-operandi that provides the necessary information from the ground level to higher global systems level. For example, performance and Fault Management Systems used be mobile phone service operators allow for optimal deployment of plant and human resources by focusing on faults that are really performance effecting and ensuring that all capital equipment is deployed optimally. Additionally, KPI reporting in the industry is near real time and is driven towards Experience Assurance. The same concept will be followed with the KPI monitoring and management in the Water Industry and its specified performance indicators.

The proposed KDE for WWTP connects the key resources and equipment in every plant (Machine), processes (Machine-to-Machine), supplier services (Plant-To-Plant), Supply-Demand Network (Supplier Network-Consumer Network), and Governess (Strategic Capability and Infrastructure Planning). This will be achieved under the auspices of Industry 4.0, Internet-of-Things in Smart Cities of the future.

At plant level the application of offline and Real-Time Machine Learning [REFERENCE] for seamless integration of GHG emissions, by-product yield, and plant productivity models through historical and adaptive real-time data acquisition technologies. The output of the KPI information will be standard and uniform format transferrable and comparable to other similar WWTP plants.

Current sensors detect changes and failures and provide sustainability data for WWTP. However, the gap in systematically integrating this knowledge into WTTP operation and real-time control persists. Event-Modelling [REFERENCE] supported and improved by Off-Line Machine Learning [REFERENCE] algorithms will identify the significant dependencies between WTTP KPI; creating the first systematic knowledge and data engineering capability that combines historical (offline) and real-time (online) data; thus bridging the gap between data and its translation into reliable knowledge of system/processes state. Coupled with Communication Systems Provider inspired interoperability protocols for data acquisition and dedicated big data processing techniques (i.e. effective ML and AI) the integration of performance measures for responsive control solutions will be achieved at plant, inter-plant and large scale multi service provider.

The nature and content of the information and knowledge produced from plant and interplant (service provider) level needs to be transformed into strategy and policy oriented models. Here at the multi service provider level (i.e. strategic) the indicators of success will be how the combination of available resources at regional, national or international level fulfils the sustainability, economic and societal (consumer satisfaction & well-being) targets. The models for this level will be based on recently suggested and evolving Sustainability Assessment Tools, Life Cycle Assessment and Cycle Costing [REFERENCE]. A more encompassing and informative water management system will be in a better condition to deal with unexpected environmental impacts such as droughts and floods, each causing severe disruption to supply.

With increased quality of information dissemination capabilities envisaged, the impact sought for this project is to stimulate water-wise communities’ engagement and active participation into decision making (supply chain). The system will allow citizens, public bodies, NGOs, and governments to access a wealth of high quality data including strategically selected environmental, economic and key performance indicators. Such an integration of metricised data will allow key policy makers to update and upgrade industry specific guidance, rules and regulations. And consequently have the instrument to objectively compare the impact of policy and governing interventions at regular intervals.

The final outcome will be Water Industry specific solution that maintains high level of reliable quality of service comparable to the high-tech communication and manufacturing industry. It will enable water suppliers help ensure the right quality of water and its supply at the right time whilst utilising the wealth of knowledge and effective optimisation algorithms to reduce capital expenditure (Capex) and operational expenses (Opex). Such an approach will cover a major consumer demand, i.e. pricing and affordability.

The solution will become the base technology to help largescale transfer of water between companies and regions, supported by storage and new local resources, to offer best and continuous level of services [1].

Figure below provides a simplistic illustration of the main features of the proposed system.



**Industry Challenges and Research Questions:**

This project intends to address the challenges pertaining in the modern water services provision, where according to WssTP vision document [REF] *“the future-proof model for a water-smart society”*:

1. Community Involvement by creating a simple easy to understand information and knowledge about the *“Story of Water”*. The story that informs citizen of where the water comes from? How it is treated? What implications the treatment has? What could help reduce those implication through consumer advice? **Water-Smart Society[[2]](#footnote-2) (from WSSTP vision)**

The specific research questions to be addressed: What are the main attributes of *Water-Smart Society* and how we can satisfy the objectives of creating the society (objective means and methods). The research challenge is to identify those key attributes and translate plant level operational data into simple language understood by all citizens, integrate patterns of consumption and the implications in one the main objectives of this project.

1. Better Strategy for Real time Control and Process Optimisation: Enrichment of the quality of knowledge of the water and waste water treatment requires expanding the capabilities of process monitoring (expansion of the number of influential parameters) that affects the performance. We will suggest the integration of sustainability dimension into the management and control of WWTPs by building the ICT platform to apply smart process monitoring and optimization. The integration of GHG emissions and energy sensors enables the conversion of carbon and energy footprint of target biological processes into timely meaningful performance measures and optimal control. A solution should be developed to enhance existing plants management technologies, increase efficiency and effectiveness of WWTPs, reduce the measurement needs by identifying inter-correlated parameters, identify false values, reduce GHGs emissions, increase real-time accuracy of knowledge for decision making on resource efficiency, reduce operational costs for water utilities and enhance access, interoperability and delivery of data acquisition in an agile decision making environment. It will develop the methodological framework to investigate the effect of the operating parameters on direct GHGs emissions generation and energy consumption in WWTPs. **Smart Water Plants**

The research question here is: What are the criteria and main specifications of a smart plant? How can current plants be upgraded to meet those criteria? Does the suggested hypothesis of expansion and integration of process KPI and GHG Monitoring lead to better operations control and management?

1. WWTP and Network Efficiency: The development of the foundation technology that effectively collects, processes and provides optimal operational performance through timely direct actuation and plans (automated and/or manual). The operational level will implement the most appropriate data acquisition technology to conduct direct readings from sensor networks in plants and environmental sensors (local or wide range). And where sensor technology is not yet reliable and economical, utilise inferential models (e.g. soft sensors) to measure the key operations performance indicators. With the help of our proposed network solution, the measurements will cover individual plants, interplant and wider water service provision network. The solution will contribute to the integration of multiple water systems within the **Multiple Water[[3]](#footnote-3) {REF WSSTP Document]** concept.

The research questions to be addressed are: What are the principles of scaling up and what are the impact of scaling up in the modelling process? How interplant knowledge can help in building a holistic approach to performance optimisation and strategic decision making? Will better insight into collective performance enable the evolution to strategic modelling platform (evolution of operational models to socio-economic models).

1. Modelling the Circular Economy: the grand scale of the proposed water resilience system needs to provide the appropriate data and knowledge platform to accurately translate the integrated data acquired for accurate circular economy modelling. This model will encompass nutrient and energy recovery, water recycling, as well as informing on further added values of the industry beyond biogas. **The Value in Water (WTSSP vision)**
2. Transforming the Water Industry to Industry 4.0 compliance: the nature of the solution is built on the fundamental paradigms of industry 4.0, i.e. smart sensing, automated large scale data and pattern recognition based modelling (Machine Learning), real-time and predictive control (Smart Control), and long term strategic infrastructure planning and decision making. **Digital Water[[4]](#footnote-4) (same document)**

The research questions: What are the main features of Digital Water? What is the role of IoT in achieving the Digital Water Paradigm? How can the final outcome of the project test the hypothesis of an integrated regional/national/international knowledge and awareness of the collective capabilities of the Water Industry can improve efficiency, productivity, environmental footprint and access to good quality of water?

With the help of our industrial partners, regulatory bodies and public (e.g. NGO or customer focus groups XXXX TO DISCUSS WHO NEEDS TO BE MENTIONED HERE) we aim to build a demonstrator technology to support the national, reginal, and international water resilience programme(s). It will be a platform for monitoring and decision making. This platform will encompass the knowledge and data engineering infrastructure that produces meaningful (knowledge) and actionable (decision making) information customised and dedicated at three levels: Engineering, Operations Management and the Strategy/Policy. With its advanced data modelling and fusion capabilities the platform customises information for the involvement of the citizens encouraging smart usage of water. It will also disseminate the information about achievements of smart-plants in recovering and recycling of valuable resources (Sustainable). Through this technology the public (e.g. tax payers), governments, NGO and investors can track the performance of the infrastructure and the impact of strategies and decisions made for water systems resilience. It will build the foundation technology to contribute to the ultimate goal of the European Water framework Directive: *“A single system of water management”*.

The objective function will be to maximise consumer satisfaction [ref] and well-being with respect to local and global Opex and Capex.

The methodology for achieving the objectives set:

1. Water smart society*,* a big data analysis and modelling platform will be developed, so-called the *Ontology of Water Smart Society*. An ontology based approach to logically and digitally assemble the engineering and operational level data into high level information on Sustainability, Economics and Societal will be created. The approach will translate complex multi-facetted data into straight non-specialist language (*Soft Language*).
2. Smart Water Plant, construction of a real-time continuous monitoring and control capability (the industry is feeling the need and willing to invest), that is enriched by historical and expert knowledge. A novel unsupervised feature extraction and machine learning capability to capture the correlation between events that influence the performance of plants will be developed. The processed data will help to construct inferential models and real-time (open control) and feedback control (delayed control) commands for process optimisation.
3. Multiple Water, upon the validation and verification of models (objective) at plant level, the exchange of data from multiple vendors and different hardware/software standards will be dealt with customising the latest interoperability technologies from mobile service providers …
4. Circular Economy Model: development of circular value chains towards more sustainable options by the application of Life Cycle Assessment (LCA) and Life Cycle Costing (LCC) tools. The tools will generate optimum circular economy options and strategies for technologies and recovered resources by integrating technical environmental and economic data, in addition to social factors, citizen engagement and customer satisfaction. The developed critical-mass network of water utilities, stakeholders and user communities (objectives 1-3) will demonstrate synergies (collaborative thinking) that will be directly communicated to the public of the benefits of circular economy implementation.
5. Digital Water, based on the principles of industry 4.0 and generic framework will be developed to encompass the outcomes of objective 1-4. The generic architecture and standards will advise the water industry and its solution suppliers with the blueprint of integrated water systems. With provision of such guideline and advice the water industry, its technology suppliers and the wider service suppliers that engage with the public will be able to plug into the systems and with respect to access privileges be able to actively create, innovate and advise within the framework of digital water.

The expected impacts are:

* Expanding the scope of clean water supply resilience, reliability and protection facilitated by a systemic integration of collective knowledge and available resources.
* Integrated water management system based on the key sources and flow i.e. *river basins management* approach.
* Combined approach of reducing environmental impact, energy consumption and increasing quality of products and services in an objective measurable and comparable format.
* Embedding citizens’ requirements and well-being with respect to quality of products and services, as well as their involvement in protecting water sources/systems through dissemination of simplified and relevant information. Leading to more efficient use of water at home, business and in agriculture.

**Work packages**

**WP1: Feasibility study to ascertain the type of plants/processes that will be used as the industrial cases**

Outcome – to determine whether the existing MSC can provide sufficient information for the (i) process performance and control and decision making (water utilities) and (ii) customers engagement and satisfaction (water wise communities – become participants in the supply chain – not consumers).

A number of wastewater treatment processes will be selected and evaluated in terms of the existing real time control and monitoring strategies. The selection criteria include: similar type and stages of applied treatment processes (main wastewater and sludge line), size of the plant (capacity), the existing automation, the real-time monitoring and control strategy and the type of the gained information, common plant service considering regulatory and legislative criteria (e.g. discharge of effluent in water bodies, reuse for irrigation, resource recovery), availability of energy, operational data that can be used for the development of the inventories. The feasibility study for the selection of the target plants will be performed in collaboration with SWT and AW in order to form the baseline scenario for the development of the MCS architecture and the transferability to other cases. This Task will give create the foundation for WP2- 4.

**WP2: Determine system parameters for measuring key performance indicators of the plants (Data analysis and Data mining)**

**Task 2.1:** Analysis of existing historical data from WATER-MINING systems. The gathered data contains a wealth of information regarding the plant operation with respect to weather conditions, which will enable us to identify behavioural patterns. This data is collected from the monitoring systems installed in the plants, including operational and environmental data. The next step is to utilise data mining techniques to identify the interrelationships between existing data about performance and stability linked to extreme events. Identification of key indicator that can describe and monitor the systems. The WATERMINING MCS will be able to capture this change and react immediately (e.g. by changing operation parameters) securing the system’s stability energy & cost savings.

**Task 2.2:** The integration of sustainability indicators (GHG and Energy) with the WATER-MINING systems performance for the purpose of better control. The GHGs will be measured using a accurate gas chamber and on line gas analyser that has been developed within the SMART-Plant and C-FOOT-CTL (REF). Energy meters will be installed following the ENERWATER methodology (REF). The GHGs emissions, energy profiles, real time monitored parameters, operating, and infrastructure data will be recorded during the systems operation (*specify*). Utilisation of statistical and real-time event-based analysis learning technique will identify the causal relationships between system input-output and control parameters (correlate environmental risks and monitored). This task will establish the minimum data points required for accurate monitoring-control.

**Task 2.3:** Integration and Testing of the MCS in the WATER-MINING plant’s monitoring and control system (on line detection of anomalies, process and infrastructure monitoring and control under various conditions, on line prediction of plant’s behaviour under certain environmental and operational factors). The model will be derived from a combination of multivariate statistics (data driven methods) and event-based analysis that captures causal relations, triggers of events and system delays.

**WP3: Building the plant level models for measuring performance indicators (Smart Water Plants)**

**Task 3.1:** is to analyse existing historical data from our WWTP (e.g. Minworth). The accumulated historical data contains wealth of information about plant operation under various operational and weather conditions. The data will be used to extract key patterns and correlations between system parameters. Advanced statistical, data clustering and event-based [REF] data analysis will provide meaningful data to help build inferential models. The raw data is provided by monitoring and control systems, infrastructure operations management, and weather data (precipitation, temperature, flood etc.). The pattern recognition algorithms will be enriched by real-time data feed into the data management systems, producing dynamism and a scenario base structure (i.e. causal relations that lead to specific state of the system). A Scenario-based analysis emerges where interrelationships between monitored data, weather data, and infrastructure performance under various scenario (e.g. extreme events) are captured.

The output will be the key indicator that describe system under a certain environmental extreme scenario (e.g. precipitation) causing changes to influent characteristics, process performance and stability of assets. Provided this knowledge becomes available, the plant MCS can react immediately (e.g. by changing operation parameters) securing system’s stability and saving energy and operational costs.

**Task 3.2:** enhancing the plant level monitoring and control capabilities by introducing sustainability indicators measurement sensors (extension of existing sensor technologies installed in WWTP). The measurements include GHG emissions using a prototype, accurate gas chamber and on line gas analyser that operates in Minworth WWTP (STW). Installation of energy meters to accurately measure the energy consumption patterns under various scenarios. The collection of 1 year GHGs emissions, energy profiles, real time monitored parameters, operating data, infrastructure data under normal operation and extreme environmental events are envisaged. This period combined with some of the useable historical will allow the validation and verification of the pattern recognition and inferential models. Statistical technique coupled with real-time event based [REF] learning techniques (liaise with Vasileia) will determine the correlations between plant level operational data with environmental risks (development of signature for each environmental risk and identification of key parameters / indicators that can describe the risk, monitor and control real time).

**Task 3.3:** Design, Integration and Testing of the MCS in the plants’ monitoring and control system (on line detection of anomalies, process and infrastructure monitoring and control under extreme events, on line prediction of plant’s behaviour under certain environmental risks). This implies the development of software and hardware platform (based on the outcome of Task 1 & 2). The historical process data, sustainability indicators and mathematical algorithms are coupled with the process understanding and infrastructure operation under extreme environmental events. The application will integrate real-time readings from the sensor & actuation network and feed the offline multivariate statistics (data driven methods – soft sensors approaches). A one year period for implementation and validation of the system will prove the plant operations management models (KPI models) as well the transfer of models into second tier models the socio-economics models (sustainability, economics, societal).

The output of the T3.3 is an MSC that is able (i) to estimate in real-time those variables that are crucial to system operation under an extreme event and (ii) Develop and apply a data driven MSC that is able (i) to estimate in real-time those variables that are crucial to system operation under an extreme event and (ii) monitor and control the plant, thus providing operators with timely information about malfunctions or chancing process states under certain environmental risks.

**Task 3.4:** evaluate the transferability of the MSC and its on-line methodologies to other water utilities. The adaptation process to other WWTPs and create a standard blueprint of standardisation (the minimum shared base cross utilities) and the adaptation guidelines. Also train operators on how to use the technology and utilise its capabilities.

**Task 3.5:** Prepare the web-based application to communicate operational data to stakeholders and target communities (data representation and levels of simplification expert to lay) strategies will be achieved.

**Main innovation:**

Development of a novel data driven on-line approach that uses multivariate techniques for Water Infrastructure monitoring and control under environmental risks. The MCS integrates historical data, sustainability indicators (GHGs and energy) and mathematical algorithms with the process understanding and infrastructure operation under extreme environmental events.

Design of MCS based on the real-life data of a large municipal WWTP and transferability to other cases. The application of the MCS real time will improve the information acquisition for infrastructure and operation control purposes and allow for safer plant operation under extreme environmental events, tight regulations and economic constraints. The adaptation of the proposed framework will provides plant operators, regulators and other stakeholders involved in the water sector with new perspectives and flexible designable MCS tool. The MCS complement the traditionally used in order to quantify and control impacts of extreme events on water infrastructure, manage risks and develop solutions.

**WP4: Transfer of the plant level measurements to higher level Water Networks (Water Networks)**

The standards and protocols for interoperability and harmonisation of data for the water industry will be established. For example a number of relevant standards can be directly transferred from SDN and new water industry specific protocols will be produced. It will enable the exchange of data between multiple plants and multiple companies for cross referencing and exchange-share of resources.

**WP5: Resource Planning and Strategic decision making (Circular Economy)**

The data fusion step enables the evolution of factory control and operations level models (i.e. figure 1, Productivity, Environment, Energy, and Quality) to second level socio-economical models (i.e. Sustainability, Economy, and Societal/wellbeing). It implies the transition from *Descriptive (inferential and physical models at plant level)* to *Prescriptive* modelling, leading to inclusive strategic planning models using Operational Research techniques. With the objective function to maximise Sustainability/Economical Gain/Well-being or maximising gains in LCA, subject to a accurately measure set of operational constraints produced from (WP-WP3).

**WP6: Project Output Dissemination of**

1. Water Resources long-term planning framework (2015-2065): Summary Report, Water UK, 2016. [↑](#footnote-ref-1)
2. A society in which the true value of water is recognised and exploited, and all available water sources are managed in such a way that water scarcity and pollution of groundwater is avoided, and water and resource loops are closed to a large extent to realise a circular economy and optimal resource efficiency, while the water-system is resilient against the impact of climate change events. [↑](#footnote-ref-2)
3. important underpinning concept of the WssTP water vision, picturing a future in which different alternative water sources and qualities (fresh ground and surface water, rain water, brackish water, saline water, brines, grey water, black water, recycled water) will be available in our society, and applied for different functions by multiple users. [↑](#footnote-ref-3)
4. important underpinning concept of the WssTP vision, based on the predicted development of a world where all people, “things” and processes are connected through the “Internet of everything” leading to capillary networks and sensors, meters and monitoring of the water-system up to the individual user, as such generating large amounts of valuable data (big data) for innovative Decision Support and Governance systems [↑](#footnote-ref-4)