SCOREWATER

Smart City Observatories implement REsilient Water Management

DELIVERABLE D1.1

REQUIREMENT SPECIFICATION (HARDWARE, SOFTWARE, STANDARDS)

REVIEW OF DIGITALIZATION OF THE WATER SECTOR

Dissemination level	Public
Type	Report
Issued by	ICRA
Contributing project partners	ICRA, EURECAT, CIV, Talkpool
Author(s)	Ll. Corominas, E. Rubion, Bas Vanmeulebrouk, Arjen Hof, Stefan Lindgren
Reviewed by	Arocena C., Hallgren F., Lindblom E.
Keywords	Digitalization, water sector, FIWARE
Number of pages	52
Number of annexes	1
Date:	2020-09-01
Version:	V 2
Deliverable number	D1.1
Work Package number:	WP 1
Status:	Delivered
Approved by coordinator (IVL)	2020-09-28

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

REVISION HISTORY

Version	Reason for changes	Name	Date
1	Original release to EU	Corominas L	2019-10-31
2	Revised version based on Project Officer's feedback sent in June 2020: Reference suggested by the project officer have been added to the references list.	Corominas L	2020-09-28

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

CONTENT

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

LIST OF FIGURES

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

LIST OF TABLES

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

ABBREVIATIONS

Abbreviation	Definition
AI	Artificial Intelligence
API	Application Programming Interface
CSO	Combined sewer overflowTLS
DL	Deep Learning
ICT	Information and communication technologies
IoT	Internet of Things
ML	Machine Learning
OUP	Open Urban Platform
SDG	Sustainable Development Goals
SME	Small and Medium-sized Enterprise
Transport Layer Security	TLS
UDS	Urban Drainage System
UV-Vis	UltraViolet-Visible
UWM	Urban Water Management
WP	Work Package

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

PROJECT ABSTRACT

SCOREwater focuses on enhancing the resilience of cities against climate change and urbanization by enabling a water smart society that fulfils SDGs 3, 6, 11, 12 and 13 and secures future ecosystem services. We introduce digital services to improve management of wastewater, stormwater and flooding events. These services are provided by an adaptive digital platform, developed and verified by relevant stakeholders (communities, municipalities, businesses, and civil society) in iterative collaboration with developers, thus tailoring to stakeholders' needs. Existing technical platforms and services (e.g. FIWARE, CKAN) are extended to the water domain by integrating relevant standards, ontologies and vocabularies, and provide an interoperable open-source platform for smart water management. Emerging digital technologies such as IoT, Artificial Intelligence, and Big Data is used to provide accurate real-time predictions and refined information.

We implement three large-scale, cross-cutting innovation demonstrators and enable transfer and upscale by providing harmonized data and services. We initiate a new domain "sewage sociology" mining biomarkers of community-wide lifestyle habits from sewage. We develop new water monitoring techniques and data-adaptive storm water treatment and apply to water resource protection and legal compliance for construction projects. We enhance resilience against flooding by sensing and hydrological modelling coupled to urban water engineering. We will identify best practices for developing and using the digital services, thus addressing water stakeholders beyond the project partners. The project will also develop technologies to increase public engagement in water management.

Moreover, SCOREwater will deliver an innovation ecosystem driven by the financial savings in both maintenance and operation of water systems that are offered using the SCOREwater digital services, providing new business opportunities for water and ICT SMEs.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

SUMMARY

This report includes a review of types of sensors and their communication protocols, and data driven models commonly used in the water sector (more specifically from sewers and urban drainage systems). The specific requirements for the SCOREwater project case-studies can be found in Deliverable 1.3. The review comes from peer-reviewed journal papers and from grey literature, including outcomes generated by former financed EU projects. The review shows that many water quantity and water quality sensors exist in the market which allow for proper monitoring of water characteristics. Recent developments in communication technologies (hardware) make it possible to transfer data from the collected sensors installed underground and in isolated places (e.g. LoRa). A plethora of data-driven models have been applied to the water field to detect abnormal functioning of sensors, to optimize the performance of the system, etc. Even though, several data-driven models exist there are only a few Big Data platforms deployed in the water sector. The widespread implementation is limited by the availability of high-quality data and fit for purpose data-driven algorithms. The digitalization of the water sector is getting into the agenda of the International Water Association and funding is allocated through the H2020 programme.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

1. PURPOSE OF A LITERATURE REVIEW

This report describes the key elements of Big Data analytics and how these have been applied to sewer and urban drainage systems. The report includes a review of water quality and flow sensors, existing communication technologies (hardware and software) potentially applicable to the water sector, a description of existing data frameworks, and a review of data-driven models applied to sewer and urban drainage systems management and operation. The review comes from peer-reviewed journal papers and from grey literature, including outcomes generated by former financed EU projects. An outlook is provided in the last chapter identifying the importance of FIWARE framework application within the SCOREwater, and indicating innovation needs which are of paramount importance during the development of the project. The purpose of this literature review is to set up the scene and provide an introduction to any partner and stakeholder willing to learn about the basics of Big Data analytics in the water sector, specifically in sewers and urban drainage systems. The specific requirements for the SCOREwater project can be found in Deliverable 1.3.

2. INTRODUCTION: BIG DATA AND THE WATER SECTOR

Climate change, mass urbanization and ageing infrastructure challenges cause inadequate system performance. Current urban water infrastructure is vulnerable to excessive rainfall, demands high capital and operational costs and deals with complex technologies and pipe networks. Hence, there is an urgent need for innovation and a change in the current urban water management (UWM) framework to provide more sustainable UWM services (Hering et al., 2013). The promise of collecting and utilizing large amounts of data has never been greater in history of UWM. Big data analytics provide a unique opportunity for the water sector to obtain reliable and relevant information at high spatial and temporal resolution. Big data analytics can be used for proper decision-making processes, as well as be considered as a catalyst for social change. Big data analytics may help to extend the service life of existing urban water infrastructure by proactive maintenance and optimized operation (Karl and Wyatt, 2018) and partially alleviate their investment needs.

The transition towards smart UWM services (using Big data analytics), the so-called digitalization of the water sector, is evident in drinking water systems, with the development of smart water supply networks. A review (executed in 2018) of the EU-funded projects which belong to the ICT4Water cluster shows that the potential of Big data analytics has been mostly demonstrated in drinking water distribution network operation and management and drinking water metering (13 projects out of 42 as can be seen Figure 1). The expansion of Big data analytics to other elements of the urban water cycle is being stimulated by the International Water Association through the newly launched IWA Digital Water Programme, which is a gateway for water utilities to access knowledge on research, technology and innovation in the digital water space (Sarni et al., 2019). The 2019 calls in H2020 on the digitalization of the water sector also include projects which focus on drainage systems and wastewater treatment.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

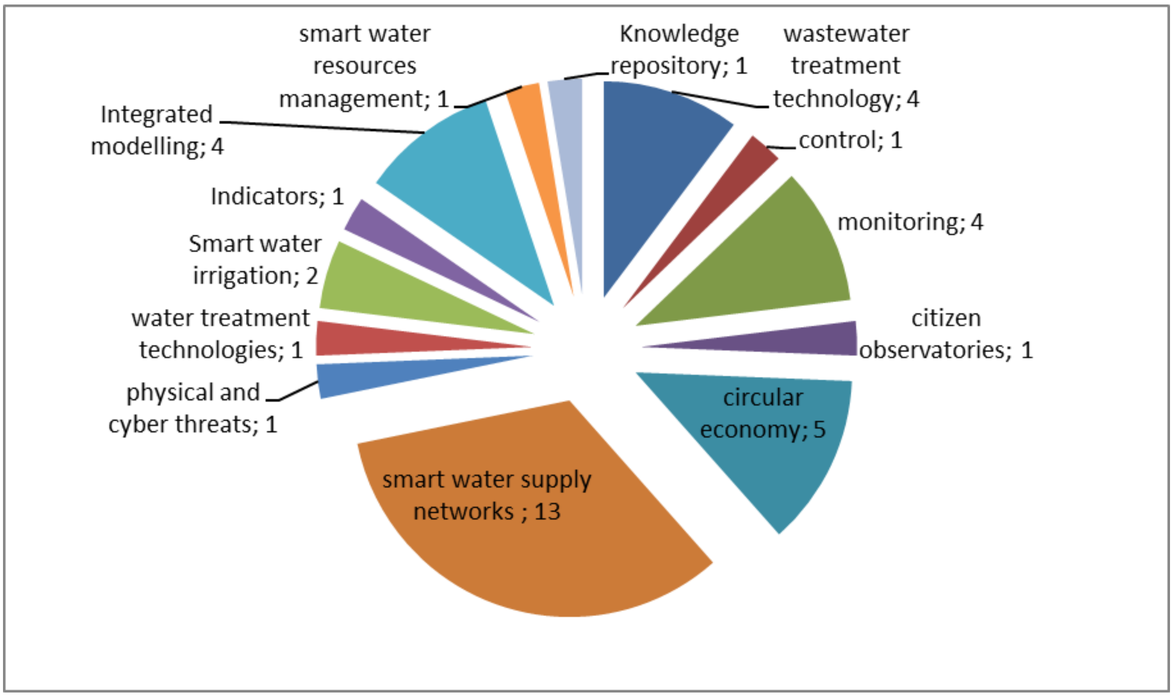

Figure 1. Classification of ICT4Water cluster projects

The transition towards digitalization of UWM cannot be achieved without the development and deployment of Big data platforms which integrate latest developments from the sensor manufacturers industry, remote-sensing, information and communication technologies, with UWM applications. The aim of this report is to describe the different elements required to develop effective Big data platforms and provide examples applied to enhance the performance of urban drainage (UDS) and sewerage systems, which is the scope of the SCOREwater project.

3. BIG DATA AND PLATFORMS

Big data are information assets characterized by high volume, velocity, variety and veracity. Fast advances in sensors, high-resolution remote sensing techniques, smart information and communication technologies and social media have contributed to the proliferation of big data. Big data brings about new opportunities for data-driven discovery, but it also requires new forms of information processing, storage, retrieval, as well as analytics. Overall, requires the development of Big data platforms made of different layers. The first layer includes the different data sources (see section 3.1); these sources can be manually updated to the platform or can be automatically uploaded by means of communication technologies; the latter include a physical layer (section 3.2) and a software layer which includes the IoT backend and the data harvester (section 3.3), which in Europe is deployed using FIWARE). The FIWARE platform brings a number of Deployment tools easing the deployment and configuration of FIWARE or third-party components and their integration with FIWARE Context Broker technology. The Context Broker is included in the Core context management layer (section 3.4). The Context Broker contains information regarding the current context. The Context Broker works according to the publish-subscribe mechanism: other components (e. g. a component to create a time series database) subscribe to updates sent out by the Context Broker. On top of that we have the context processing, analysis and visualization layer (section 3.4).

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

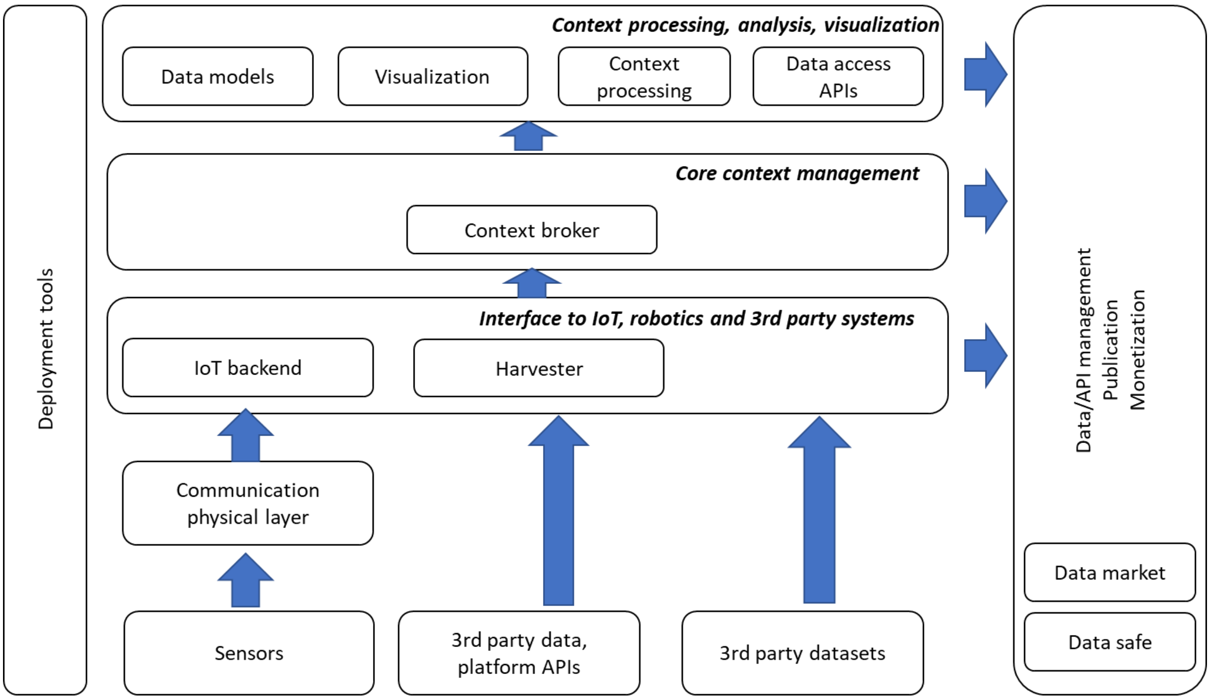

Figure 2. Typical architecture for a Big Data platform applied to the water sector. An interrelation of the elements is shown: Devices/data sources, connectivity/network, data framework intelligence framework and data market.

In the following sections, the state-of-the art for each of the elements is described.

3.1. SOURCES OF BIG DATA

Urban drainage and sewerage systems Big Data may originate from a) multi-sensor data collected from ground-based monitoring networks and Internet of things (IoT), b) large-scale datasets collected from field experiments via multiple instruments, c) data simulated by system models, d) high-frequency data products derived from Earth observation systems, and e) crowdsourced data from social media and citizen science (Sun and Bridget, 2019).

This review focuses on the multi-sensor data collected from ground-based monitoring networks. Smart management of sewage networks relies on sensors that provide on-line data and actuators from installed equipment in the network (e.g. pumps, valves, etc.). A review of existing sensors in sewers and urban drainage systems is provided in this section. Real time control of urban water infrastructure is not the focus of SCOREwater; hence, the description of actuators is not included in this review. Although reliable equipment has become available nowadays, the wrong choice of monitoring equipment is one of the main factors causing the improper working of sewers and UDS. In terms of estimating water quantity (i.e. flows) a plethora of sensors is available, with capacitive probes, pressure sensors, ultrasonic probes or microwave sensors, which have been successfully implemented to monitor the water level, and ultrasonic or electromagnetic meters to monitor the water flow.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

On-line sensors measuring temperature, oxidation-reduction potential, pH, dissolved oxygen, conductivity and turbidity have proven their reliability and robustness in drainage systems (Vanrolleghem and Lee 2003). Multi-wavelength spectroscopy in the ultraviolet visible range (UV-Vis) has been used for wastewater and stormwater quality monitoring (Van der Broeke et al. 2006, Lourenco et al. 2008, Torres and Bertrand-Krajewski 2008), since most organic compounds and some soluble minerals (such as nitrate/nitrite) absorb in the UV-Vis region (Vaillant et al. 2002). On-line UV-Vis sensors have already been used for wastewater treatment plant operation (Vanrolleghem and Lee 2003, Schuetze et al. 2004) and robust submersible commercial devices are also becoming available. However, their application in UDS has scarcely been reported (Gruber et al. 2006, Rieger et al. 2006, Torres and Bertrand-Krajewski 2008). In UDS the hydraulic conditions are extremely variable, fouling may occur and there is limited access for maintenance. These circumstances may contribute to the collection of inaccurate data. Additionally, wastewater in UDS may not present a homogenous composition along its cross section. This creates a challenge for spectrophotometry, which can be minimized by choosing a location with a turbulent flow. In drainage systems with storm water connections, flow may change rapidly, modifying the water quality matrix (Maribas 2008), increasing fouling probability and worsening access conditions, which adds pressure to in-line monitoring. Statistical models will then play a key role since they can enhance the scope and significance of acquired spectral data.

Annex Table 1 provides a list of on-line sensors available for the monitoring of flows, levels and water quality in urban drainage and sewerage systems with a detailed description.

Besides the data coming from sensors installed in the different applications the Big Data platform can integrate data coming from other sources; for example it can connect to 3rd parties platforms and exchange data by using an Application Programming Interface (API). In many cases it is not necessary to store all the relevant information in a specific Big Data platform if that information is already available in another platform. Hence, it makes sense to develop APIs which allow for the exchange of data between platforms. The API is executed whenever there is a need to recover the required data.

3.2. COMMUNICATION PHYSICAL LAYER

This subsection describes the physical layer for the data communication. It consists of the electronic circuit transmission technologies of a communication network. It is a fundamental layer and can be implemented through a great number of different hardware technologies with widely varying characteristics (Table 1).

Table 1. Hardware technologies for data communication

Hardware data communication	Pro	Con
Wired connectivity	Low maintenance cost High data throughput	Installation cost Inflexible
3G/4G/5G/LTE	Flexible Mid data throughput	Network coverage Short battery life
WiFi	Flexible High data throughput Cheap	Network coverage impossible Short battery life
LoRa	Flexible Long battery life Cheap Network coverage	Low throughput
NB-IoT	Flexible Cheap Network coverage	Short battery life
Bluetooth	Flexible High data throughput Cheap	Network coverage impossible Short battery life
6LOPAN	Auto configurable mesh network	Inflexible

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

The specificities of sewers and UDS makes the data communication challenging. Sewer networks are widespread; as an example Spain counts on 90,000 Km of sewer pipes, a distance which is twice the the equatorial circumference of Earth. The sensors are normally installed underground and in isolated places with no electricity connections. Hence, the main challenge is to provide sensors with connectivity which can transmit at long distances, with low battery consumption. The main challenges are deployment cost, maintainance cost and flexibility. Hence, the LoRa technology is gaining popularity in the field of monitoring sewer systems and UDS.

3.3. INTERFACE TO IOT AND DATA HARVESTER

3.3.1. COMMUNICATION PROTOCOLS AND DATA MODELS

Despite a promising technological scenario, the water domain is characterized by a low level of maturity concerning the standardization of Big Data limiting the build-up of knowledge across domains and platforms. This is due to the fragmentation of the sector, the heterogeneous water data sources and platforms, and no holistic vision. Therefore, the legacy systems of different stakeholders difficult the integration with third parties due to **non-standard communication protocols** and **data models**. Improving access to data and fostering open exchange of water information is foundational to solving water resources issues. The development of system standards is essential for smart water solutions that should ensure interoperability of solutions.

Existing water Big Data platforms rely on a service-oriented approach, where web services connect with different involved systems using non-standardized XML/SOAP (Anzaldi, 2014). This kind of architecture requires most of the human input to adapt the XML/SOAP interconnection when the network grows due to the lack of discovery services to support plug and play capabilities (Malewski, 2013). To overcome this deficiency, the latest approaches use the OGC stack complemented by data models such as WaterML2 and O&M (Anzaldi, 2014). Also, novel initiatives of European Commision have focused their efforts on creating interoperable platforms like FIWARE (Zahariadis, 2014), which are capable of consuming heterogeneous information across multiple IoT standards (MQTT, AMQP, DDS.) and publishing the knowledge through open and dynamic interfaces like NGSI and NGSI-LD. As well as, standardization organizations such as ETSI, OASIS, OGC, OMA are addressing these challenges providing open interfaces and extending vocabularies for the water sector (Anzaldi, 2018).

Annex Table 2 includes a list of the relevant communication protocols used on water sector is presented. The table includes the field of application of the standard, a brief description and references to the literature where the standard is used.

The OPC-UA standard[1], which supports a request/reply communication with process automation systems, is widely used in the industrial sector, but it delegates the telemetry transfer to publish/subscribe mechanisms in order to optimize the data communications. Therefore, MQTT, AMQP and DDS protocols are responsible for this communication in the OPC-UA standard. Also, it is important to note that Omron[2] and Siemens[3] PLCs, that integrates OPC-UA standard, uses MQTT to transfer data.

[1] HYPERLINK "https://opcfoundation.org/about/opc-technologies/opc-ua/"https://opcfoundation.org/about/opc-technologies/opc-ua/

[2] http://blog.omron.eu/a-practical-illustration-of-iot-and-industry-4-0/

[3] https://support.industry.siemens.com/cs/document/109748872/fb-lmqtt_client-for-simatic-s7-cpu?dti=0&lc=en-WW

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

The communication protocols presented in Annex Table 2 can be used to connect multiple devices in a distributed network through wired and wireless communication technologies (for example, MQTT, AMQP, DDS and CoAP) or Server to Server (for example, JMS, RESTful, OGC SOS, Digital Delta, FIWARE NGSI and NGSI-LD). All of them are available for free thanks to open source licences. AMQP, MQTT, JMS are brokered-based architecture. Therefore, publishers post messages to a trusted message routing and delivery service, or broker, and subscribers register subscriptions with the broker which also performs any message filtering. Moreover, they facilitate the networks' scalability deploying more instances of the broker. Instead, REST/HTTP, Digital Delta and CoAP are based on a typical Client-Server architecture where client invokes the methods of the server. FIWARE NGSI and FIWARE NGSI-LD have a hybrid architecture, offering request-reply and publish-subscribe interfaces. NGIS, NGIS-LD, OGC SOS, DDS, REST/HTTP, Digital Delta and CoAP are interoperable, hence their messages can be exchanged and understood by different implementations. Instead, MQTT, AMQP and JMS are not completely interoperable. MQTT is agnostic to the content of the payload and does not specify the layout or how data is represented in the message. Therefore, the exchange of the messages is sure, but the serialisation of the content requires a shared scheme. AMQP messages adds information about the layout in the "content-type" and "content-encoding", but it is only a convention. Therefore, the data serialisation scheme should to be understood by the publisher and subscriber to ensure that the data payload is interpreted. JMS does not provide a standard for interoperability outside of the Java platform. All messaging technologies have a comparable performance in a simple point-to-point configuration, although broker-based architectures (MQTT, AMQP and JMS) adds an additional overhead in the communications. MQTT, AMQP and JMS do not provide automatic discovery, unlike DDS, NGIS, NGIS-LD, Digital Delta and OGC SOS, this means that configuring a distributed system that uses one of these technologies is through the broker. CoAP, NGIS, NGIS-LD, Digital Delta and OGC SOS support a client/server programming model based on a Service Oriented Architecture, SOAP for OGC SOS and RESTful for the others, in which resources are server-controlled abstractions made available by an application process and identified by Universal Resource Identifiers (URIs). Clients can manipulate resource using HTPP: GET, PUT, POST and DELETE methods. It also provides in built support for resource discovery as part of the protocol. AMQP, JMS, NGIS, NGIS-LD and OG SOS provide transactional modes of operation, hence they can take part in a multi-phase commit sequence. The trusted and fault-tolerance of the messaging technologies is also important. JMS, OGC SOS, NGSI and NGSI-LD do not provide an API for controlling the privacy and integrity of messages. Then, the security is provided by the JMS, OGC SOS vendors, and FIWARE GEs module for NGSI. MQTT v3.1 and AMQP provides authentication facilities and the encryption of data exchanged can be handled using SSL or TLS. DDS defines the Security Model and Service Plugin Interface (SPI). It customizes the behaviour and technologies that the DDS implementations use for Information Assurance, specifically allowing customization of Authentication, Access Control, Encryption, Message Authentication, Digital Signing, Logging and Data Tagging. RESTful uses asymmetric cryptography for authentication of key exchange and symmetric encryption for confidentiality through SSL or TLS. CoAP uses Datagram Transport Layer Security (DTLS) that is equivalent to SSL/TLS over UDP. Finally, all the standards are supported by FIWARE.

Annex Table 3 presents a summary of the highlights of each communication protocol. Data models, such as ontologies and schemas, promote interoperability and play a prominent role in the World Wide Web Consortium. They also play a role in the IoT and linked data fields, where they assist data contextualization, reduce discovery and consistency. Data models describe concepts, relationships, data properties within a domain, in a machine-readable manner, being key in knowledge sharing. Annex Table 4 summarizes the most relevant water data models, including schemas and ontologies identified on the literature review. Table 2 provides an analysis of each identified data model considering: the richness, quantity of entities and properties; the formalization, how the data model is implemented; the standardization organization involved in the development of the data model; the licensing model adopted by the data model and the most relevant dependencies, that is, the data models imported.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Table 2. Analysis of relevant water data models

Data Model / Ontology	Richness	Formalization	Standardization Entity	Licencing Model	External Dependencies
WaterML2	663 elements 19 attributes	XSD Schema	OGC	Open Source	GML OM
HY_Features	452 Individuals 282 Classes 403 Properties 6239 Axioms	OWL/RDF	OGC	Open Source	Geosparql Skos Sf Gml iso19150 iso19115 iso19103
Digital Delta data models	13 elements 66 attributes	RAML	HydroLogiv, Nelen & Schuurmans, Deltares	Open Source	GeoJSON (GRFC 7946)
HydroNET4	6 elements 10+ attributes	Swagger OpenAPI specification		-	NetCDF-CF (http://cfconventions.org/), GeoJSON (GRFC 7946),
RiverML	-	XSD Schema (based on WaterML)	OGC	Open Source	WaterML 2.0 GML OM
SWEET	2147 Individuals 4543 Classes 363 Properties 25057 Axioms	OWL	ESIP Foundation	Open Source	-
GWML2	757 elements 19 attributes	XSD Schema	OGC	Open Source	GeoSCiML OM
O&M	639 elements 19 attributes	XSD Schema	OGC	Open Source	GML GMD
HydroML	139 elements 3 attributes	XSD Schema	National Water Information System of the Water Resource	Open Source	
INSPIRE	717 elements 19 attributes	XSD Schema	European Commission	Open Source	GML 3.2.1
FIWARE Data Models	26 elements 34 attributes	JSON	ETSI	Open Source	-
SAREF4WATR	On development	OWL	ETSI	Open Source	-

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

3.3.2. DATA HARVESTER

Data harvesting uses a process that extracts and analyzes data collected from online sources. The term data harvesting actually goes by other different terms. They include web mining, data scraping, data extraction, web scraping, and many other names. Data harvesting has grown in popularity in part because the term is so descriptive. It derives from the agricultural process of harvesting, wherein a good is collected from a renewable resource. Data found on the internet certainly qualifies as a renewable resource as more is generated every day. Data harvesting can be very beneficial, especially when using a third-party service. The data gathered from websites can provide organizations with helpful information and insights that can inform their business practices and help them reach out to prospective consumers. With so much data available on the web, data harvesting has become a popular and at times necessary tool so companies have a more thorough knowledge of marketplaces, consumers, and competitors.

3.3.3. FIWARE

As FIWARE is the approach followed in SCOREwater we provide a more detailed description in this subsection. FIWARE is an open source initiative defining a universal set of standards for context data management which facilitate the development of Smart Solutions for different domains such as Smart Cities, Smart Industry, Smart Agrifood, and Smart Energy. The European Innovation Partnership for Smart Cities & Communities (EIP-SCC) and the Espresso-project have described a capabilities map with all the different data-components for FIWARE.

The FIWARE Catalogue is a curated framework of open source platform components which can be assembled together and with other third-party platform components to accelerate the development of Smart Solutions. The main and only mandatory component of any “Powered by FIWARE” platform or solution is the FIWARE Orion Context Broker Generic Enabler, which brings a cornerstone function in any smart solution: the need to manage context information, enabling to perform updates and bring access to context. Building around the FIWARE Context Broker, a rich suite of complementary FIWARE components are available, dealing with: a) Interfacing with the Internet of Things (IoT), Robots and third-party systems, for capturing updates on context information and translating required actuations; b) Context Data/API management, publication, and monetization, bringing support to usage control and the opportunity to publish and monetize part of managed context data and c) Processing, analysis, and visualization of context information implementing the expected smart behavior of applications and/or assisting end users in making smart decisions. Existing FIWARE data frameworks are described in Table 3.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Table 3. FIWARE data frameworks

Data framework	Field of application	Description	Reference
FIWARE Orion context broker	Real time data	The Orion Context Broker holds information about the current context from sensors and IoT-devices.	https://www.fiware.org/developers/catalogue/
FIWARE Cygnus/QuantumLeap/Draco	Cygnus, QuantumLeap and Draco subscribe to updates from a context broker and persist these, thus creating a time series database. Depending on the application at hand, the most appropriate one should be selected.	Comparable components to support data persistence mechanisms for managing the history of context	https://www.fiware.org/developers/catalogue/
CKAN	Open data portal platform, services catalogue and metadata registry	For managing static open data and describing metadata. Within FIWARE extended with plugins to support data market and context data.	https://ckan.org/
IDAS	IoT-agents for interfacing with specific systems	Interfaces to make it easier to connect with IoT-devices	https://www.fiware.org/developers/catalogue/
API-management	Monitization, monitoring and access to data. Managing API's	API-management enables policies for managing access to data, including billing, and metering	https://www.fiware.org/developers/catalogue/ https://www.redhat.com/en/technologies/jboss-middleware/3scale
Process, analyze and visualize			

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

3.4. BIG DATA ANALYTICS

Big data platforms include a layer specific for data analytics; data analytics can be based from simple statistics to artificial intelligence (AI). The layer provides improved decision support to stakeholders, enabling real time input and data-driven services that increases the possibilities to ground decisions on valid data. In the literature, the terms Artificial intelligence (AI), Machine learning (ML) and deep learning (DL) are sometimes used interchangeably. AI is a general term referring to the use of computers/machines to imitate human-like behaviors, ML is a branch of AI that aims to train machines to learn and act like humans and to improve their learning in autonomous fashion through data fusion and real-world interactions, while DL refers to a newer generation of ML algorithms for extracting and learning hierarchical representations of input data (Sun and Bridget, 2019).

AI's ability to constantly adapt and process large amounts of data in real-time makes it a promising tool for managing water resources in an ever-changing environment. Data-driven methods are being applied by the research community to the field of water urban management (Newhart et al., 2019). The water treatment field has benefited from model developers and process engineers who have helped produce mechanistic models trying to describe the pollutants transformation processes occurring in the drinking water plants, distribution systems, sewers, WWTPs, and rivers. Such models are being complemented with data-driven models (Eggimann et al., 2017; Newhart et al., 2019). In the field of wastewater treatment, the research community has been testing and developing methods for fault detection, isolation and diagnosis, for process modelling, for control, etc. Artificial Neural Networks and Principal Component Analysis are the methods that have been widely tested in this field. Other examples of methods are Fuzzy logic, Clustering, Independent Component Analysis, Partial Least Squares, self-organizing maps, Regression, Support Vector Machines, and Qualitative features detection (Corominas et al., 2018). However, these methods remain in the scientific domain and has been little knowledge transfer to industry. The main limiting factor for widespread implementation of these methods at full-scale is the limited capabilities to ensure the data is collected is of high quality. Even though there has been large progress in the development of water quality sensors they are exposed to extreme environmental conditions (e.g. extreme hydraulic conditions, large turbidity which can cause clogging, fouling or blocking due to sand). There is no sense in having a great data driven model if the data that feeds the model is of poor quality.

Below, the state of the art of the three main issues addressed throughout the SCOREwater project are reviewed: Flash flood predictions to anticipate disasters in endangered areas, predictive maintenance on Sewer systems to improve urban resilience, and finally prediction of water quality to anticipate pollution problems and stormwater treatment. A summary can be found in Annex Table 5.

Floods are among the most destructive natural disasters, which are highly complex to model. The research on the advancement of flood prediction models contributed to risk reduction, policy suggestion, minimization of the loss of human life and reduction of the property damage associated with floods. These models follow different strategies in order to contribute to the problem, using short-term or long-term flood predictions, or geographical prediction to identify the flood alert areas.

In hydrology, the definitions of short-term and long-term in studying the different phenomena vary. Short-term predictions for floods often refer to hourly, daily and weekly predictions, and they are used as warning systems. On the other hand, long-term predictions are mostly used for policy analysis purposes and often refer to prediction time greater than a week.

Short-term lead-time flood predictions are considered important research challenges, particularly in highly urbanized areas, for timely warnings to residences so to reduce damage (Zhang, 2018). In addition, short-term predictions are beneficial to water resource management. Even with the recent improvements in Machine Learning methods, short-term predictions aren't an easy task (Badrzadeh, 2015). In order to solve this kind of problems, the modeling methods can be single or hybrid. Single methods use one Machine Learning algorithm in order to predict an objective (Yu, 2017), while the hybrid methods use the Machine Learning algorithm with the addition of a statistical algorithm or hydraulic model (Jimeno, 2017).

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Some examples of single short-term prediction through Deep Learning are (Yu, 2017), (Saghafian, 2017) or (Sahoo, 2006), which use ANN and variations in order to solve flow related problems. There are more Machine Learning techniques used, specifically Support Vector Machines (SVM), Ensemble Models (ELM) and Random Forest Regressors (RF) (Yu, 2017), (Cheng, 2009). These examples use environment variables to predict objective flood variables like water level, surge level, streamflow, rainfall, and structural condition.

Concerning to the hybrid short-term prediction, they improve the quality of the prediction, in terms of accuracy, generalization, uncertainty, longer lead-time, speed and computation costs. Some examples are used to flood warning (Doycheva, 2017), peak flow prediction (Jimeno, 2017), rainfall prediction (Chang, 2014) (French, 2017) (Young, 2017), among others.

Long-term flood prediction is of significant importance for increasing knowledge and water resource management potential over longer periods of time, from weekly to monthly and annual predictions. Many notable ML algorithms have been studied in the past decade, but it is still not clear which ML methods are the best for long-term predictions (Mosavi, 2018).

The last type of flood predictions is spatial prediction of flood-susceptible areas. Various hydrological methods have been studied for determining flood-susceptible areas in the watersheds, but these methods face numerous difficulties because of limitations and high costs. For this reason, environmental analyses in the form of GIS is also being researched (Rahmati, 2015). Machine Learning for spatial mapping is a methodology with an evolving research path and with good success rate, (Rahmati, 2016) purpose a Random Forest and Maximum Entropy solution for groundwater mapping with an area under the curve of 83.1% and 87.7%, respectively, becoming effective models. Finally, (Haghizadeh, 2018) explain an approach using frequency ratio and maximum entropy models in order to do spatial prediction of flood-susceptible areas in Iran. Their result was an area under the curve of 74.3% and 92.6%, respectively, meaning the path using machine learning can be innovative.

Sewer networks are among the most critical urban infrastructures, suffering from a huge variety of problems. Studies have confirmed possible solutions to sewer overflow, sanitary sewer condition deterioration, and sewer chokes by fats, grease or other anomalies by applying Data-driven models.

Sewer Overflow is a major problem to be addressed by many cities. While understanding the behavior of the sewer system through proper urban hydrological models is an effective method for management, conventional deterministic methods rely on physical principles which make the computation expensive and not usable for short time prediction. (Zhang, 2018) propose a Deep Learning model that aims at forecasting sewer overflow events from multiple sewer structures simultaneously in near real-time at a Norwegian citywide level. In order to construct the model, detailed information about the studied sewer system (a small number of 8 points in the sewer network), rainfall data and sewer hydrological data such as flow or water level was provided. The research demonstrates that the multi-task approach is generalized better than single-task approach, furthermore, the GRU and LSTM are especially suitable to capture the temporal and spatial evolution of the sewer network event and superior to other methods.

Sanitary sewers, as a part of wastewater infrastructure systems, are designed to collect the sanitary sewage from domestic, industrial, commercial and public users and convey it to a treatment plant. (Mohammadi, 2019) introduce a Machine Learning methodology to predict the condition of sanitary sewer pipes based on historical inspection data from the City of Tampa. Information about pipe materials, physical, environmental and operational variables were utilized to build the deterioration model. Using logistic regression, the results predict with an accuracy of 81.4% and the most important variables were pipe age, material, diameter, length and water levels, as they affect the most to the condition of the pipe over time. (Lin, 2015) propose a Bayesian nonparametric approach, namely the Dirichlet process mixture of hierarchical beta process model for water pipe failure prediction. In this research, the pipe attributes of diameter, length, laid date, material, protective coating, soils, and traffic intersections are used for the predictive model of three regions. Using the algorithm HBP (hierarchical beta process) and some variations of this algorithm achieves an accuracy of 82.67%, 74.51% and 78.37% of each region respectively.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Sewer chokes are blockages typically caused by external factors such as fats, grease, tree roots and foreign items in the pipes such as wet wipes. Chokes may lead to sewage overflows through designated sewer system overflow points or uncontrolled overflows into public or private property. (Cameron, 2017) worked on Machine Learning solutions for the Sydney Water sewer networks, using a predictive model composed by the variables of tree roots, pipe properties, past chokes, climate data, buildings and behavior data in the zone. (Bailey, 2016) present the development of an ensemble of decision tree models produced using the data from the wastewater network of Dwr Cymru Welsh Water to predict the likelihood of blockage and inform the prioritization of proactive maintenance. The data used to model was the sewer diameter, the sewer length, the number of property and food producer connections per sewer, information on the property types and ages present, and the water velocity. The results of this research were around 0.7% area under the curve for the different inputs of information.

Water **pollution** is a serious problem in the world which threatens human health, ecosystems, and plant life. Prediction of water quality is one of the main concerns in water resource and environmental systems, as it helps control water pollution. There is research in urban water quality prediction, water quality around construction areas, prediction of sediment toxicity from sewers and much more.

Urban water quality refers to the physical, chemical and biological characteristics of a water body, and several chemical indexes can be used as effective measurements for the water quality in current urban water distribution systems. (Liu, 2016) deploy several monitoring stations throughout the city's water distribution system to provide real-time water quality reports. In order to identify water quality, turbidity, residual chlorine, and pH are evaluated in the different points of interest, being the Residual Chlorine the final water quality index. To predict these properties, pressure, meteorology, flow, spatial factors like pipe or road networks and pipe properties like age, material, and length were used for the model construction. In order to predict the water quality of a station by fusing multiple sources of urban data, a novel framework using spatio-temporal multi-view multi-task was presented. Related to the investigation, (Zheng, 2015) purpose various spatio-temporal models that combine multiple datasets in order to identify city anomalies in a spatial network. Their research can be used to identify water quality anomalies in different points of interest of a city.

Sediment toxicity from sewers is important, even more, when there are stormwater provoking overflows. (Schertzinger, 2019) explain their research on urban wet weather discharges, which involves sediment pollution from habitats destruction and other pollutants from UWWDs such as metals, polycyclic aromatic hydrocarbons, polychlorinated biphenyls, pesticides or flame retardants. With the help of data from different locations which include sediment samples, maintenance information, sediment contact assays, ecological risk assessment, and statistical analysis. This research ends showing that combined sewer overflows (CSO) have an impact on the toxic potential as well as on the oxygen demand of downstream located sediments, showing the most important variables for future forecasting studies.

With the rapid development of urbanization, increasing impervious areas interrupt stormwater infiltration channels and greatly increase stormwater runoff volume and peak flow. In addition, due to human activities, atmospheric deposition and other factors, a large number of pollutants accumulate and are discharged into the municipal stormwater sewer by stormwater runoff flushing. (Wang, 2015) research shows the impact of adsorption of heavy metals by construction wastes and an effective new way for resource utilization of city construction waste. On the other hand, (Cha, 2017) propose a machine learning method for building demolition waste and develop a demolition waste generation rate. The study uses data from 796 low-rise residential buildings immediately before the building demolition process to acquire the data for characteristics and material quantities for each building. The variables used were region, building type, type of structure, wall material, roof material and base construction materials like mortar or concrete. The prediction of concrete generation was successfully predicted by using a decision tree variation called CHAIN with a 98'9%.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

3.5. DATA MARKET

As the gigabytes, terabytes, and petabytes of unstructured information pile up, most UWS organizations lack actionable methods to tap into, monetize, and strategically exploit this potentially enormous new value. McKinsey research reveals that companies currently underutilize most of the IoT data they collect. One effective way to put IoT data to work and cash in on the growing digital bounty involves offering the information on data marketplaces to third parties. Digital marketplaces are platforms that connect providers and consumers of data sets and data streams, ensuring high quality, consistency, and security. The data suppliers authorize the marketplace to license their information on their behalf following defined terms and conditions. Consumers can play a dual role by providing data back to the marketplace. Third parties can offer value-added solutions on top of the data the marketplace offers. For example, real-time analytics can make consumer insights more actionable and timelier than ever before. The marketplace also has an exchange platform as a technical base for the exchange of data and services, including platform-as-a-service offers.

A successful data market requires i) Developer support by means of documentation and examples, ii) API management by means of billing, metering, security and iii) API provisioning by means of NGSI (FIWARE), TMForum Open APIs, or other. Following we provide a more detailed description of security and standardization, which are two key elements for the data market.

3.5.1. SECURITY

There are many levels of security to take into account.

Hardware design of IoT devices and communication. *As the number of Internet of Things (IoT) devices have been predicted to surpass 50 billion by 2020, attackers' attention has also shifted towards tools, techniques, and procedures to exploit IoT networks. Hence, IoT security and privacy are two urgent challenges. The security mechanism ensures the correctness and integrity of the data which is being communicated through the communication devices and gateways, that is, they are sent to its destination without any distortion. Therefore, communication protocols, platform and hardware communication devices requires from security mechanisms such as: (i) Authentication and Authorization to make sure that only trusted clients can connect and don't interfere with each other by using user/password, private/public keys pairs, or X509 certificates and (ii) Transport Layer Security (TLS) to ensure that eavesdropper can't read and intercept the messages for transport layer encryption. It is important to note that FIWARE and the most relevant IoT communication protocols (MQTT and AMQP), supports authentication and authorization mechanisms like user/password, private/public keys and X509 certificates. Moreover, all of them also integrate TLS mechanisms to encrypt the messages.*

Components within the Big data platform. *All components within the Big data platform should be based on well-maintained open source libraries whenever possible. The latest security updates of those libraries should be installed.*

API's. *API's are the corner stone of any platform. There are multiple ways to secure API's and address possible vulnerabilities. The Open Web Application Security Project (OWASP) keeps track of a "cheat sheet" with all kinds of API-vulnerabilities. A REST API should for example always use HTTPS, never store readable passwords in a database or a configuration file, never disclose sensitive information in a URL, restrict access to methods which are not used, use input validation (e. g. always verify the user actually enters an integer when the system expects an integer) to make sure that the user does not enter values which will break down the system, not disclose information of the internals of the system in error messages, use proper HTTP response codes and content types etcetera. This article provides a good overview of best practices for REST API security. A solid API-management and API-provisioning solution should support a wide range of security measures like encryption, authentication, and authorization protocols.*

User interface. *The user interfaces should not allow end-users to enter sensitive information (for example a password) in a human readable fashion.*

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

End user. *End users should be aware of security risks. This involves awareness of phishing and social engineering (do not click suspicious links, inform the IT department of scams), access, passwords and connection (use strong passwords, do not share login credentials with co-workers), device security (do not connect unsecured devices to the company network) and physical security (do not leave devices unattended, always lock your computer when stepping away from your desk) as described by continuum.net.*

4. OUTLOOK

4.1. POTENTIAL VALUE UNLOCKED BY BIG DATA IN THE SEWAGE TREATMENT SECTOR

Big data has the potential to create trillions of dollars of value across the economy (Kato, 2018). It is estimated that the potential value unlocked by ICT in the sewage treatment sector is up to $22.8B in 2017 and will grow by 7.2 % annually from 2017-2021 (Bozalongo, 2017). Tech giants are slowly getting into the water sector by adding data management and big data analytics on top of existing water and wastewater treatment businesses. The Bluemix (from IBM), the Predix (from General electric) and Genesis64 (from Iconics) platforms, amongst others, have already been applied to the water sector (Krause et al., 2018). These platforms provide capabilities in remote asset monitoring, energy analytics, and water security. The integration of all types of data will facilitate critical decision making relative to assets and facilities at the right time and place in order to optimize security and reliability. Large water companies such as Stantec (MWH), Suez-Degremont are investing in digitalization and are generating the demand for it.

4.2. FIWARE APPLICABILITY

Currently, FIWARE is able to manage multiples transport mechanism between the devices and the IoT Agent of FIWARE based on Request/Reply or Publish/Subscribe paradigm. Request/Reply paradigm uses HTTP to connect each device directly to the IoT Agent. Instead, Publish/Subscribe paradigm is event-driven and requires an additional central communication point (broker) which is in charge of dispatching all messages between the senders and the rightful receivers. The dispatching is based on the topic subscription and publication, enabling highly scalable solutions without dependencies between the data producers and the data consumers. Therefore, it is recommended to use publish/subscription paradigm to load data on FIWARE platform throughout the SCOREwater project. Currently, the most relevant publish/subscribe standards previously identified are: MQTT, AMQP, DDS and JMS. It is important to note that MQTT has a relevant position in the market due to German industry is recommending it as telemetry standard. Although the Request/Reply transport mechanisms are not as efficient, they should not be discarded. The integration of legacy systems and/or multi-parametric devices can condition this recommendation due to limited communication capabilities of them. As an example, Digital Delta communication protocol is widely used in Netherlands and therefore is part of the legacy ICT tools, some multi-parametric devices are closed and only offer HTTP-based communication.

Concerning to the standardized communication server to server, FIWARE provides natively NGSI & NGSI-LD standards. Therefore, it is strongly recommended to use one of them as communication protocol between applications, NGSI-LD preferably due to better support to linked data, that is, entity relationships, property graphs and semantics. Despite following a request-reply paradigm, NGSI and NGSI-LD also includes Publish-Subscribe paradigm providing the best of both approaches, dynamic discovery usually related to Request-Reply paradigm and low coupling related to Publish-Subscribe paradigm. In the same way as for communication between devices and FIWARE platform, the rest of the standards should not be discarded until a detailed analysis of the possible interactions with third-party applications. For example, if hydrological models or similar are integrated in the Dutch Case, probably the Digital Delta standard will be required.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Finally, although FIWARE data models include water data information such as WaterQualityObserved, it is very limited. Listing 1 shows a summary of the schema, which does not include information about the data quality, the origin of the data (measured, interpolated, simulated...). Concepts addressed by other data models such as WaterML2, HY_FEATURES and the future SAREF4WATR. WaterML2 is based on schema and hence, it lacks semantic capabilities. Although HY_FEATURES ontology is applied in numerous water projects to flood risk management, data management and smart city services, its formalization hinders the integration in FIWARE platform. FIWARE NGSI and NGSI-LD interfaces support JSON and JSON-LD specifications respectively, instead HY_FEATURES is formalized in OWL language. Concerning SAREF4WATR, which is also developed by ETSI as FIWARE platform, will be formalized with JSON-LD and hence, it will be totally and natively supported by FIWARE.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Listing 1. JSON schema for WaterQualityObserved of FIWARE Data models


```
{
 "$schema": "http://json-schema.org/schema#",
 "$id": "https://fiware.github.io/dataModels/specs/Environment/WaterQualityObserved/schema.json",
 "title": "GSMA / FIWARE - Water quality observed schema",
 "description": "Water Quality data model is intended to represent water quality parameters at a certain water mass (river,  lake, sea, etc.) section",
 "type": "object",
 "allOf": [
  {
   "$ref": "https://fiware.github.io/dataModels/common-schema.json#/definitions/GSMA-Commons"
  },
  {
   "$ref": "https://fiware.github.io/dataModels/common-schema.json#/definitions/Location-Commons"
  },
  {
   "properties": {
    "type": {
     "type": "string",
     "enum": ["WaterQualityObserved"],
     "description": "NGSI Entity type"
    },
    "dateObserved": {
     "type": "string"
    },
    "measurand": {
     "type": "array",
     "items": {
      "type": "string"
     },
     "minItems": 1
    },
    "temperature": {
     "type": "number"
    },
…
    "NO3": {
     "type": "number",
     "minimum": 0
    },
    "refPointOfInterest": {
     "$ref": "https://fiware.github.io/dataModels/common-schema.json#/definitions/EntityIdentifierType"
    }
   }
  }
 ],
 "required": ["id", "type", "dateObserved", "location"]
}
```


This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

SAREF4WATR is an ontology under development by ETSI to meet the water cross-domain information exchange needs resulting from various water infrastructures (for example, smart meters, early warning systems...). The starting point of the SAREF4WATR is the Rioter ontology[4], an ontology built by Eurecat that extends SAREF ontology towards water domain. Rioter ontology links physical and digital water world, the measurements with their corresponding data quality and extends the measurements towards considering different variable indexes and scales. It is important to note that European Commission is supporting and encouraging the SAREF4WATR usage through ICT4WATER cluster. The cooperation of the different projects involved in the ICT4WATER cluster will allow to develop a water vocabulary/ontology fully aligned with the European and International Standards and stakeholders needs. Therefore, SAREF4WATR is a firm candidate to be integrated on FIWARE platform throughout the SCOREwater project and other projects involved in the same call as discussed during the last ICT4Water Cluster meeting. However, despite all the benefits offered by a cross-project collaboration, the SAREF4WATR ontology is out of the control of the SCOREwater, and because of that its integration will be evaluated continuously throughout the project.

Concerning the third party integration, currently the major part of water software does not support any standard, they only use proprietary data models and APIS. For example, MIKE software suite for stormwater and urban drainage modelling uses DFS API and data models and PFS API and data models. Both are totally proprietary and hence, non-standard. To ensure the interoperability of the SCOREwater platform, the interfaces and data models will be standard, and if necessary, specific connectors to support communication standard-proprietary or proprietary-standard must be implemented.

With regard to platforms and their development it is important to look at more generic features too. The development of Open Urban Platforms (OUP) is stimulated by the publication of a DIN-standard (DIN is the German standardization body). This standard builds upon the work done by EIP-SSC and is aimed to promote interoperability, openness and standardization. FIWARE is adopting these principles and being used in several water management projects, like SCOREwater and FIWARE4Water. As mentioned before, FIWARE offers buildings blocks that support the development of cross-sectoral open solutions and has been successfully applied to a wide range of use cases.

4.3. REMAINING CHALLENGES FOR THE WATER SECTOR

The widespread implementation of ICT-driven UWM is hindered by generic challenges related to generic data and technology, which includes data processing (turning data into information and into knowledge), data availability (making useful information available to stakeholders), data quality, data costs (achieving low operation and maintenance costs for sensors and measurements), and the lack of general standards and protocols or data management. IWA has recently launched the digital water programme to facilitate the journey of the water industry towards digital uptake and integration into water services. The IWA Specialist group on instrumentation, control and automation (ICA), has been coordinating international initiatives (including forums for discussions, collecting and exchanging of methodologies, and practical experience). They have embraced significant progress in developing low-cost water-related sensors, models, and control algorithms with a very effective combination of process knowledge and ICA tools. Now, the tech giants come into position, with vast experience in AI tools, but rather limited process knowledge. Either the tech giants bring the water processes knowledge in, or water academics adapt to this reality by embracing new deep learning tools. The differing interests from the water research community and the tech giants might be a limitation to pursue the effective integration of AI tools in the sector. One potential solution would be the training of a new generation of researchers/practitioners trained both in engineering, statistics, and computer science through the creation of multidisciplinary training programmes.

[4] https://rioter-project.github.io/rioter-core-ontology/

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

4.4. SOCIAL AND ORGANIZATIONAL ENABLERS FOR THE DEVELOPMENT AND USE OF SMART WATER MANAGEMENT

One of the risks of widespread implementation of big data platforms (more than lack of technological capabilities) is the reluctance of utility managers to delegate the control of water treatment infrastructure to machines, and such reluctance is slowing down the penetration of tech giants into the water sector. Too often, technical development is made with inadequate interaction with stakeholders, not fully attending to their needs and requirements. As a consequence, regulatory/legal, economic and social barriers to develop useful tools and to the adoption and widespread use of these developments are not adequately addressed. Moreover, data-driven UWM requires a change in organizational practices. Professionals working in UWM are generally not used to dealing with an abundance of data and dynamic systems. In order for various stakeholders to participate and effectively contribute to the development of digital technologies and to use and profit from the improvements in water management that these technologies offer, as well as to identify best practices and guidelines for further exploitation, there is a need to identify salient social and organizational enablers. The successful adoption of new technical services by stakeholders requires a proven business case; it must be clear that the new tools provide opportunities to quantify improvements in resilience and track changes over time. Moreover, having an iterative development process where developers and stakeholders meet recurrently over time is crucial to create useful tools that can be adapted to the needs of different users. The Smart Resilience project also shows that for high efficiency, new tools need to complement and be integrated with existing tools, aiming towards reaching goals or enabling new ways to reach them. However, neither Smart Resilience nor other projects reviewed has validated the models and data developed through actual use or developed process variables for increasing resilience.

5. CONCLUSIONS

This report provides evidences on the application of Big Data and AI in the management of urban drainage and sewerage systems. The water sector represents some of the most interesting and yet challenging use cases that are out there for Big data, due to its multidisciplinary character. It involves different data sources (e.g. multi-sensor data collected from ground-based monitoring networks and IoT, large-scale datasets collected from field experiments via multiple instruments, data simulated by system models, etc.). For IoT the communication is challenging as some of the sensors are installed underground and the field needs to use the most advanced protocols (e.g. LoRa). We identified 10 types of communication protocols available, and 12 data models (e.g. WaterML) for data standardization. In terms of Big data analytics, we provide a review of data driven models already applied (artificial neural networks, auto-regression, support vector machine, Hybrid ML, Regression trees, Logic regression, Bayesian networks, Principal component analysis, Fuzzy logic, clustering) to solve different problems. Even though, several data-driven models exist there are only a few Big Data platforms deployed in the water sector. The widespread implementation is limited by the availability of high-quality data and fit for purpose data-driven algorithms. The digitalization of the water sector is getting into the agenda of the International Water Association and funding is allocated through the H2020 programme.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

6. REFERENCES

(Adelman, 2017), “Remote Sensing in a Body of Water Using an Adafruit Feather”, Proceedings of Student-Faculty Research Day, CSIS

(Aguado, 2008), “Multivariate statistical monitoring of continuous wastewater treatment plants”, Eng. Appl. Artif. Intell., 21, 1080-1091, doi: 10.1016/j.engappai.2007.08.004

(Aichouri, 2015), "River flow model using artificial neural networks”, Energy Procedia, 74, 1007-1014

(Alvisi, 2019), “Wireless Middleware Solutions for Smart Water Metering”, Sensors 2019, 19(8), 1853; https://doi.org/10.3390/s19081853

(Alwis 2016), "Evaluation of the durability and performance of FBG-based sensors for monitoring moisture in an aggressive gaseous waste sewer environment", JLT.2016.2593260, Journal of Lightwave Technology (Åmand 2013) " Aeration control - a review" Water Science & Technology | 67.11 | 2013 (Brito 2013), "In situ UV-Vis spectroscopy to estimate COD and TSS in wastewater drainage systems", Urban Water Journal, 2014 Vol. 11, No. 4, 261-273.

(Anjana, 2015), “An IoT based 6LoWPAN Enabled Experiment for Water Management”, IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), 1-6, doi: 10.1109/ANTS.2015.7413654

(Anzaldi, 2014), “A holistic ICT Solution to improve matching between supply and demand over the water supply distribution chain”, Journal of sustainable development of energy, water and environment systems", 2 (4)

(Anzaldi, 2018), “Digital Single Market for Water Services Action Plan”, Publications Office of the European Union, doi:10.2759/724173

(Badrzadeh, 2015), “Hourly runoff Forecasting for Flood Risk Management: Application of Various Computational Intelligence Models”, Journal of Hydrology, 529(3), 1633-1643, doi: 10.1016/j.jhydrol.2015.07.057

(Beaufils, 2019), “Geotechnical data standardization and management to support BIM for underground infrastructures and tunnels”, WTC – World Tunnel Congress

(Bello, 2013), “POME is treated for removal of color from biologically treated POME in fixed bed column: applying wavelet neural network (WNN)”, Journal Hazard. Mater., 262, 106-113, doi: 10.1016/j.jhazmat.2013.06.053

(Brodaric, 2018), “Enabling global exchange of groundwater data: GroundWaterML2”, Hydrogeology Journal, 26 (3), 733-741

(Campisano A. 2013), "Potential and limitations of modern equipment for real time control of urban wastewater systems", Urban Water Journal, 2013 Vol. 10, No. 5, 300-311.

(Carstea 2016), " Fluorescence spectroscopy for wastewater monitoring: A review" Water Research 95 (2016) 205e219.

(Cha, 2017), “New approach for forecasting demolition waste generation using chi-squared automatic interaction detection (CHAID) method”, Journal of Cleaner Production, 375-385

(Chang, 2014), “Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information”, Journal of Hydrology, 508, 374-384

(Cheng, 2009), “Comparison of the Extreme Learning Machine with the Support Vector Machine for Reservoir Permeability Prediction”, International Conference on Computational Intelligence and Security, 173-176, doi: 10.1109/CIS.2009.124

(Coble 1996), "Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy", Marine Chemistry 51 (1996) 325-346.

(Cornelissen 2018), "Application of online instrumentation in industrial wastewater treatment plants - a survey in Flanders, Belgium", Water Science & Technology | 78.4 | 2018.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

(Corominas, 2018), “Transforming data into knowledge for improved wastewater treatment operation: A crtitical review of techniques”, Environmental Modelling & Software, 106, 89-103, doi:10.1016/j.envsoft.2017.11.023

(Çoruh, 2014), “The use of NARX neural network for modelling of adsorption of zinc ions using activated almond shell as a potential biosorbent”, Bioresour. Technol., 151, 406-410, doi: 10.1016/j.biortech.2013.10.019

(Cunningham, 2017), "Assessment of Ecosystems: A system for rigorous and rapid mapping of floodplain forest condition for Australia’s most important river”, Land Degradation Development, 29(1), 127-137, doi: 10.1002/ldr.2845

(Danso-Amoako, 2012), "Predicting dam failure risk for sustainable flood retention basins: A generic case study for the wider greater manchester area", Computers, Environment and Urban Systems, 36, 423-433

(Deo, 2015), “Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in Eastern Australia”, Atmospheric Research, 161, 65-81, doi: 10.1016/j.atmosres.2015.03.018

(Dornblut, 2013) “HY_Features: a geographic information model for the hydrology domain”, GRDC Report

(Dovzan, 2011a), “Recursive fuzzy c-means clustering for recursive fuzzy identification of time-varying processes”, ISA Trans., 50, 159-169, doi:10.1016/j.isatra.2011.01.004

(Dovzan, 2011b), “Recursive clustering based on a Gustafson-Kessel algorithm”, Evol. Syst., 2, 15-24, doi: 10.1007/s12530-010-9025-7

(Doycheva, 2017), “Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application to runoff simulations and flood warning”, Advanced Engineering Informatics, 33, 427-439

(Dürrenmatt, 2011), “Data-driven modeling approaches to support wastewater treatment plant operation”, Environmental Modeling Software, 30, 47-56

(Eggimann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy De Vitry, M., Beutler, P., Maurer, M., 2017), “The Potential of Knowing More: A Review of Data-Driven Urban Water Management” Environ. Sci. Technol. 51, 2538-2553. https://doi.org/10.1021/acs.est.6b04267

(Elsafi, 2014). “Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile”, Alexandria Engineering Journal, 53(3), 655-662, doi: 10.1016/j.aej.2014.06.010

(Eriksson, 2018), “Techniques for and consequences of using INSPIRE extensions: a case study with Swedish hydrological data”, International Journal of Spatial Data Infrastructures Research, 13, 172-201

(Fleischer, 2010), “An Integration Platform for heterogeneous sensors systems in GITEWS-Tsunami Service Bus”, Natural Hazards and Earth System Sciences, 10, 1239-1252

(French, 2017), “Combining machine learning with computational hydrodynamics for prediction of tidal surge inundation at estuarine ports”, Procedia IUTAM, 25, 28-35

(Gazendam, 2016). “Integrative neural networks models for stream assessment in restoration projects”, Journal of Hydrology, 536, 339-350, doi: 10.1016/j.jhydrol.2016.02.057

(Geunjeong 2007), "Performance of an electrochemical COD (chemical oxygen demand) sensor with an electrode-surface grinding unit" J. Environ. Monit., 2007, 9, 1352-1357.

(Ghose, 2018), "In Measuring discharge using back-propagation neural network: A case study on brahmani river basin, Singapore”, Springer Singapore: Singapore, pp 591-598

(Gilbert, 2010), “Knowledge discovery with clustering based on rules by states: a water treatment application”, Environ. Model. Soft, 25, 712-723, doi: 10.1016/j.envsoft.2009.11.004

(Gruber 2005) "Continuous monitoring in sewer networks an approach for quantification of pollution loads from CSOs into surface water bodies" Water Science and Technology, 52 (12), 215-223.

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

(Haghizadeh, 2018), “Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models”, Geocarto International, 1752-1762, doi: 10.1080/10106049.2017.1316780

(Hanson 2007), "Comparisons of Optical pH and Dissolved Oxygen Sensors With Traditional Electrochemical Probes During Mammalian Cell Culture" , Biotechnology and Bioengineering, Vol. 97, No. 4, July 1, 2007 (Harremoes 1992), "Wastewater treatment plants under transient loading-performance, modelling and control" Sci. Tech. Vol. 27, No. 12, pp. 71-115,1993.

(Hauser 2019), "Real-time wastewater monitoring as tool to detect clandestine waste discharges into the sewage system" ENVIRONMENTAL FORENSICS https://doi.org/10.1080/15275922.2019.1566295

(Hussain, 2015), “A generic framework to integrate water supply distribution systems using interoperable standards”, E-proceedings of the 36th IAHR World Congress

(Jackson, 2014), "RiverML: a harmonized transfer language for river hydraulic models”

(Jeppsson 2002), "Status and future trends of ICA in wastewater treatment - a European Perspective" Water Science and Technology Vol 45 No 4-5 pp 485-494.

(Jiang 2019) "Submersible probe type microbial electrochemical sensor for volatile fatty acids monitoring in the anaerobic digestion process" Journal of Cleaner Production 232 (2019) 1371e1378.

(Jiang2014) "Determining the long-term effects of H2S concentration, relative humidity and air temperature on concrete sewer corrosion" Water Research 65 (2014) 157e169.

(Jimeno, 2017), “Estimation of Instantaneous Peak Flow Using Machine-Learning Models and Empirical Formula in Peninsular Spain”, Water, 9(5), doi: 10.3390/w9050347

(Ju, 2009), “Division-based rainfall-runoff simulations with bp neural networks and xinanjiang model", Neurocomputing, 72, 2873-2883

(Kamienski, 2019), “Smart Water Management Platform: IoT-Based Precision Irrigation for Agriculture”, Sensors, 19(2), 276, doi: 10.3390/s19020276

(Kanwar,2010), “Web Service based hydrologic data distribution system”, Computers & Geosciences, 36(7), 819-826, doi: 10.1016/j.cageo.2007.10.017

(Kim, 2001), “Quantitative Flood Forecasting Using Multisensor Data and Neural Networks”, Journal of Hydrology, 246, 45-62

(Kim, 2016), “A real-time forecast model using artificial neural network for after-runner storm surges on the tottori coast, japan”, Ocean Engineering, 122, 44-53.

(Korak 2014) "Critical Analysis of Commonly Used Fluorescence Metrics to Characterize Dissolved Organic Matter" Water Research 2014.

(Kourgialas, 2015), “Statistical analysis and ann modeling for predicting hydrological extremes under climate change scenarios: The example of a small mediterranean agrowatershed”, Journal of Environmental Management, 154, 86-101

(Kumarjha 2018) "Smart Water Monitoring System for Real-time water quality and usage monitoring" Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018).

(Lee 2019), "Optimizing pre-treatment conditions for anaerobic co-digestion of food waste and sewage sludge" Journal of Environmental Management 249 (2019) 109397.

(Lee, 2018), “Applying CoAP for real-time device control over public networks”, International Conference on Electronics, Information, and Communication (ICEIC), 1-2, doi: 10.23919/ELINFOCOM.2018.8330562

(Lin, 2006). “Using support vector machines for long-term discharge prediction”, Hydrological Scienses Journal, 51(4), 599-612

(Lin, 2012), “Monitoring and control of UV and UV-TiO2 disinfections for municipal wastewater treatment plants by a multivariate statistical process control and neural networked-based software sensors”, Process Biocherm, 43, 1107-1113, foi:10.1016/j.procbio.2008.06.002

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

(Liu, 2016), “Urban water quality prediction based on multi-task multi-view learning”, IJCAI'16 Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2576-2582

(Liukkonen, 2013), “Advanced monitoring platform for industrial wastewater treatment: multivariable approach using the self-organizing map”, Environmental Modelling Software, 48, 193-201, doi: 10.1016/j.envsoft.2013.07.005

(Lohani, 2014), “Improving real time flood forecasting using fuzzy inference system”, Journal of Hydrology, 509, 25-41

(Looser, 2014) “International Standardization of water information exchange: Activities of the WMO/OGC Hydrology Domain Working Group”, 11th International Conference on Hydroinformatics (HIC)

(Lopez-Morales, 2019), “An Interoperable Platform for the Digital Transformation of the Agricultural Sector”, Global IoT Summit, 1-6, doi: 10.1109/GIOTS.2019.8766382

(Lorraine, 2015) “Global Storm Surge Forecasting and Information System”, Geophysical Research Abstracts, 17

(Luccarinno, 2010), “Formal verification of wastewater treatment process using events detected from continuous signals by means of artificial neural networks. Case study: SBR plant”, Environmental Modelling Software, 25, 648-660, doi: 10.1016/j.envsoft.2009.05.013

(Ma, 2009), “Inductive data mining: automatic generation of decision trees from data for QSAR modelling and process historical data analysis”, Comput. Aided Chem. Eng., 25, 581-586, 10.1016/S1570-7946(08)80102-2

(Malewski, 2013), “Semantic Matchmaking & mediation for sensors on the sensor web”, 7(3), 929-934

(Marsili-Libelli, 2008), “Intelligent monitoring system for long-term control of sequencing batch reactors”, Water Science Tehcnology, 57(3), 431-438, Doi: 10.2166/wst.2008

(Mcknight 2001) " Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity" Limnol. Oceanogr., 46(1), 2001, 38-48.

(Mesquita 2017) "Monitoring biological wastewater treatment processes. Recent advances in spectroscopy applications" Rev Environ Sci Biotechnol DOI 10.1007/s11157-017-9439-9.

(Min, 2016), “Moving Towards Sustainable and Resilient Smart Water Management”, ACECC

(Mosavi, 2018), “Flood Prediction Using Machine Learning Models: Literature Review”, Water, 10(11), doi: 10.3390/w10111536

(Nanda, 2016), "A wavelet-based non-linear autoregressive with exogenous inputs (wnarx) dynamic neural network model for real-time flood forecasting using satellite-based rainfall products”, Journal of Hydrology, 539, 57-73

(Nebbioso 2012) " Molecular characterization of dissolved organic matter (DOM): a critical review" Anal Bioanal Chem DOI 10.1007/s00216-012-6363-2.

(Newhart, K. B.; Holloway, R. W.; Hering, A. S.; Cath, T. Y, 2019) “Data-Driven Performance Analyses of Wastewater Treatment Plants: A Review” Water Res., 157, 498–513.

(Nguyen 2015) "Automatic process control in anaerobic digestion technology: A critical review" Bioresource Technology (2015).

(Olsson 2015) " Instrumentation, Control and Automation in Wastewater Systems" Scientific and Technical Report No. 15.

(Pan, 2010), “Comparison of the extreme learning machine with the support vector machine for reservoir permeability prediction”, Computer Engineering & Science, 2, 037

(Parygin, 2017), “A convergent model for distributed processing of Big Sensor Data in urban engineering networks”, Journal of Physics, 803

(Pereira, 2006), “Modeling a densely urbanized watershed with an artificial neural network, weather radar and telemetric data”, Journal of Hydrology, 317, 31-48

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

(Ponomareva 2011) "Microbial Biosensors for Detection of Biological Oxygen Demand (a Review)" Applied Biochemistry and Microbiology, 2011, Vol. 47, No. 1, pp. 1-11.

(Rahmati, 2015), “Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis”, Geomatics, Natural Hazards and Risk, 7(3), 1000-1017, doi: 10.1080/19475705.2015.1045043

(Rahmati, 2016), “Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran”, Catena, 137, 360-372, doi: 10.1016/j.catena.2015.10.010

(Rezaeian-Zadeh, 2012), “Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions”, Arabian Journal of Geosciences, 6(7), 2529-2537, doi: 10.1007/s12517-011-0517-y

(Rezaeianzadeh, 2014), “Flood flow forecasting using ann, anfis and regression models”, Neural Computing and Applications, 25, 25-37

(Robles 2015) "Instrumentation, control, and automation for submerged anaerobic membrane bioreactors" Environmental Technology, 2015 Vol. 36, No. 14, 1795-1806.

(Rooney, 2013), “The Netherlands Looks to Big Data to Tackle Flooding”, The Wall Street Journal, https://blogs.wsj.com/tech-europe/2013/06/25/the-netherlands-looks-to-big-data-to-tackle-flooding/

(Rubión, 2019), “Towards an Open, Low-Cost and Enhanced Standards-Based IoT Architecture for Autonomous and Smart Water Quality Control and Monitoring”, 17th International Computing & Control for the Water Industry Conferences

(Sadler, 2016), “A recipe for standards-based data sharing using open source software and low-cost electronics”, Journal of Hydroinformatics, 18 (2), 185-197

(Saghafian 2017), “Effect of ENSO on annual maximum floods and volume over threshold in the southwestern region of Iran”, Hydrological Sciences Journal, 62(7), 1039-1049, doi: 10.1080/02626667.2017.1296229

(Sahoo, 2006), “Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu”, Journal of Hydrology, 327(3), 525-538, doi:10.1016/j.jhydrol.2005.11.059

(Schertzinger, 2019), “Predicted sediment toxicity downstream of combined sewer overflows corresponds with effects measured in two sediment contact bioassays”, Environmental pollution, 782-791

(Shamim, 2016), “A comparison of artificial neural networks (ANN) and local linear regression (LLR) techniques for predicting monthly reservoir levels”, KSCE Journal of Civil Engineering, 20(2), 971-977, doi: 10.1007/s12205-015-0298-z

(Sheng, 2015), “Recent Advances in Industrial Wireless Sensor Networks Toward Efficient Management in IoT," IEEE Access, 3, 622-637, doi: 10.1109/ACCESS.2015.2435000

(Simons, 2015), “Improving access to groundwater data using GroundWaterML2”, 36th Hydrology and water resources symposium: The art of science of water, 609-616, ISBN: 9781922107497

(Singh, 2013). “Indian summer monsoon rainfall prediction using artificial neural network”, Stochastic Environmental Research and Risk Assessment, 27(7), 1585-1599, doi: 10.1007/s00477-013-0695-0

(Soleimani, 2013), “Experimental investigation, modeling, and optimization of membrane separation using artificial neural network and multi-objective optimization using genetic algorithm”, Chemical Engineering Res. Des., 91, 883-903, doi: 10.1016/j.cherd.2012.08.004

(Srihari, 2018), “Intelligent Water Distribution and Management System Using Internet of Things”, 2018 International Conference on Inventive Research in Computing Applications, 785-789. doi: 10.1109/ICIRCA.2018.8597325

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

(Stasch, 2017), “Coupling Sensor Observation Services and Web Processing Services for Online Geoprocessing in Water Dam Monitoring”, International Journal of Digital Earth, 11, 64-78, doi: 10.1080/17538947.2017.1319977

(Sun and Bridget, 2019), “How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions”, Environ. Res. Lett., 14, 073001.

(Thomas 1999) "TOC versus UV spectrophotometry for wastewater quality monitoring" Talanta 50 (1999) 743-749.

(Toran 2016) "Water Level Loggers as a Low-Cost Tool for Monitoring of Stormwater Control Measures" Water 2016, 8, 346; doi:10.3390/w8080346.

(Tsertou, 2015), “ICT Infrastructure for Efficient Water Resources Management: Towards Harmonization of a Fragmented Field”, Environmental Engineering & Management Journal (EEMJ), 14(9), 2007-2011

(Vacariu, 2015), “An INSPIRE-Compliant System for Pollution Detection for Somes River”, 20th International Conference on Control Systems and Computer Science, 803-808, doi: 10.1109/CSCS.2015.25

(Van der Broeke 2006) "On-line and in-situ UV/vis spectroscopy for multi-parameter measurements: a brief review" Spectroscopy Europe, 18 (4),15-18.

(Vanrolleghem, P.A 2003) "On-line monitoring equipment for wastewater treatment processes State of the art" Water Science and Technology Vol 47 No 2 pp 1-34.

(Villez, 2016), “Shape constrained splines with disontinuities for anomaly detection in a batch process”, Comput. Aided Chemical Engineering, 91, 365-379

(Vinoj, 2018), “Smart City Underground Water Leak and Theft Detection System with IoT”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3(6), 632-640

(Wang, 2015), “Adsorption characteristics of construction waste for heavy metals from urban stormwater runoff”, Chinese Journal of Chemical Engineering, 1542-1550

(Wu 2019) "Anaerobic digestion. A review on process monitoring" Renewable and Sustainable Energy Reviews 103 (2019) 1-12.

(Xu 2019) "Response of water quality to land use and sewage outfalls in different seasons" Science of the Total Environment 696 (2019) 134014.

(Yang 2018) "Characterization of effluent organic matter from different coking wastewater treatment plants" Chemosphere 203 (2018) 68e75.

(Yang, 2018), “An Observation Task Chain Representation Model for Disaster Process-Oriented Remote Sensing Satellite Sensor Planning: A Flood Water Monitoring Application”, Remote Sensing, 10, 375, doi:10.3390/rs10030375

(Yestilmezsoy, 2011), “Artificial intelligence-based prediction models for environmental engineering”, Neural Networks Wold, 21, 193-218

(Yoo, 2003), “Nonlinear modeling and adaptive monitoring with fuzzy and multivariate statistical methods in biological wastewater treatment plants”, Jounal Biotechnology, 105, 135-163, doi:10.1016/S0168-1656(03)00168-8

(Young, 2017), “A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events”, Applied Soft Computing Journal, 53, 205-216

(Yu, 2013), “Possible control approaches of the Electro-Fenton process for textile wastewater treatment using on-line monitoring of DO and ORP. Chemical Engineering Journal, 218, 341-349”

(Yu, 2017), “Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting”, Journal of Hydrology, 552, 92-104, doi:10.1016/j.jhydrol.2017.06.020

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

(Zahariadis, 2014), “FIWARE lab: managing resources and services in a Cloud Federation Supporting Future Internet Applications”, IEEE/ACM 7th International Conference on Utility and Cloud Computing, 792-799, doi: 10.1109/UCC.2014.129

(Zhang, 2010), “Effects of plant diversity on microbial biomass and community metabolic profiles in a full-scale constructed wetland”, Ecol. Eng., 36(1), 62-68

(Zhang, 2018), “Operating characteristic information extraction of flood discharge structure based on complete ensemble empirical mode decomposition with adaptive noise and permutation entropy”, Journal of Vibration and Control 24(17), 5291-5301, doi: 10.1177/1077546317750979

(Zheng, 2015), “Detecting Collective Anomalies from Multiple SpatioTemporal Datasets across Different Domains”, Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems

(Zulkifli 2017) "ArticleDetection of contaminants in water supply: A review onstate-of-the-art monitoring technologies and their applications" Sensors and Actuators B.

(Zyrianoff, 2018), “Scalability of an Internet of Things Platform for Smart Water Management for Agriculture”, Proceeding of the 23rd Conference of Open Innovations Association FRUCT, 58, 432-439

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

ANNEX 1. COMPENDIUM OF LONG TABLES

Annex Table 1. Detailed explanation of sensor types

Type of sensors	Field of application	Description	Reference
Thermistor, Thermocouples, Thermometers	Anaerobic digesters, Sludge management	Temperature is a well-known parameter. It is an important factor in different wastewater treatment processes and it must be controlled especially in anaerobic processes, and sludge reactors. A **Thermistor** is a powered metal oxide whose measures the temperature related to the resistance. There are two different Thermistor, NTC (resistance decreases as temperature rises) and PTC (resistance increases as temperature rises). A **thermocouple** is an electrical device consisting of producing temperature-dependent voltage as a result of the thermoelectric effect. The voltage is interpreted to measure temperature. A **thermometer** measures temperature gradient using a mercury-in-glass bulb, with a numeric scale that shows the numeric gradient. Its functioning is about thermal expansion of the metal, so when the temperature rises, the metal expands.	(Lee 2019) (Das 2017) (Vanrolleghem 2003) (Harremoës 1993)
Pressure sensors **Pressure transmitter** **Pressure transducer**	Aeration processes Anaerobic digesters Alarm functions	A pressure sensor is used to control the pressure measurements of liquids and gases. They can be called pressure transducers or transmitters, which are widely used to control several wastewater treatment processes the pressure. A **pressure transmitter** is a probe that converts pressure into an analog electrical signal. The most used pressure transmitter is strain-gauge based transducer. The strain gauge is deformed into the diaphragm of the probe. The quantification of pressure is related to the physical deformation of this gauge. The electrical resistance produced by the strain gauge is proportional to the pressure.	(Hauser 2019) (Vanrolleghem 2003) (Harremoës 1993)
Liquid Level sensors	Monitor water level, Alarm functions, Sewer level control, Storm water control, Debris control	The measurement of liquid level is an important concern of wastewater management and control. The liquid level can change rapidly, and for that, specific sensors should be used: i) Capacitive probes, ii) Pressure sensors, iii) Ultrasonic probes, iv) Microwave sensors. As a summary, liquid level sensors work in two different ways: **Point level measurement sensors** are used to mark a single discrete liquid height. This kind of sensors functions as alarm, showing overfill or low level conditions **Continuous level transmitters** can measure a fluid level range, providing a liquid level control of an entire system.	(Hauser 2019) (Toran 2016)(Campisano 2013)(Vanrolleghem 2003)(Campisano 2009)(Harremoës 1993)
Gas flow rotameters, Thermal mass flow meters; Electromagnetic flow meter	Wastewater treatment, Alarm functions	A **flow meter** (or flow sensor) is an instrument used to measure linear, nonlinear, mass or volumetric flow rate of a liquid or a gas.	(Hauser 2019) (Vanrolleghem 2003) (Harremoës 1993)

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Type of sensors	Field of application	Description	Reference
pH electrodes	Wastewater quality control Alarm functions	**pH** is the negative logarithm of ion hydrogen concentration. **pH sensor** has a measuring electrode and a reference electrode. The differential voltage of both submerged electrodes gives the pH value of a solution. The temperature is critical because of the differential voltage could change. pH electrodes can be necessary to detect abrupt changes in the sewage network. The control of these probes can spare treatment plants of cost overruns.	(Hauser 2019) (Jha 2018) (Das 2017) (Vanrolleghem 2003)
Conductivity	Wastewater quality control Alarm functions	There are three different conductivity sensors. **Two electrodes sensor, or absolute measuring probe:** An Alternating Current (AC) is applied and it generates an electric current. Its intensity will depend on the number of free anions and cations contained in the liquid. The more anions and free cations the liquid contains, the greater electrical conductivity will be. **Four electrodes sensor or differential measuring probes:** An ion high concentration liquid causes a current reduction, the so-called *Polarization effect*. This effect can influence the measurement accuracy of the conductive probes. This kind of probe has 4 electrodes, 2 without current, so without affectation by the *Polarization effect*. These two electrodes measure the electric potential difference in the liquid. **Inductive/toroid sensor:** Toroid probes contain a transmission and a reception coil, and measure the conductivity in several steps: An oscillator generates a magnetic field in the transmission coil, which induces a voltage in the liquid. The cations and anions of the liquid start to move generating an alternating current. In this way, an alternating magnetic field and, consequently, a current flow in the receiving coil is induced. The current intensity and conductivity are directly proportional to the number of free ions in the liquid.	(Hauser 2019) (Jha 2018) (Das 2017) (Lim 2017) (Vanrolleghem 2003)
Gaseous products(H2, CH4, CO2, H2S)	Sewer WWTPs	H2 and H2S are critical components related to corrosion in wastewater major infrastructure assets. Hydrogen and wet hydrogen sulfide gas are the precursors of aggressive acidic ambient that destroys conventional electronic sensors. **Specific hydrogen analyzers** have been changing to become more robust and reliable. 3D printed epoxy resin packaging and Polyether ether ketone packaging probes, previously calibrated with zero-maintenance, are being the most evaluated.	(Alwis 2016) (Jiang 2014) (Vanrolleghem 2003)

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Type of sensors	Field of application	Description	Reference
Alkalinity sensors **Titrimetry principle** **Titrimetric sensor** **Pressure sensors** **Gas flow meters**	Anaerobic Digestion control Total Alkalinity (TA) Partial Alkalinity (PA) pH Ammonia Bicarbonate Dissolved Carbon Dioxide (DCD)	A bacterium in anaerobic digesters requires a range of pH from 6.5 to 8. Stabilization of the pH-value, maintaining it within this range is obtained by the release of large amounts of bicarbonate (a buffering compound). The control of alkalinity is especially controlled in order to obtain a good response from anaerobic digesters. **Titrimetric sensors** are widely used to control the pH of the liquor and to know the concentration of bicarbonate. This technique is conducted titrating the sample down and leads the pH value to 3.5. Other on-line probes have the same principle but using also, a mathematical model interpretation with a titration process that uses a whole pH-range (3-11) control. **Gaseous carbon dioxide quantifiers** as widely used to detect and control the acidification of samples. This measure can be performed by two ways: the first one is **measuring the overpressure** in constant volume vessel. The second one uses a **gas flow meter** to measure the produced volume of gas.	(Campisano 2013) (Vanrolleghem 2003) (Harremoës 1993)
UV-Vis Absorbance **Infrared Spectroscopy (IR)** **Optical sensors** **Ultrasound sensors**	Influent/Effluent Quality Contaminant control Total Organic Carbon (TOC) Chemical Oxygen Demand (COD) Total Suspended Solids (TSS)	**UV-VIS absorbance** measurement is an important organic matter checkpoint. Beside of this, it is significant to monitor and control the removal of micropollutants. Different wavelengths are indicators of high toxicity components as hydrocarbons, pesticides or analgesics. Other important uses of UV-Vis absorbance is the characterization capacity of different organic matter. **Spectrophotometer** works relying on the fact that electromagnetic radiation (EMR) interacts with atoms of a liquid in discrete ways to produce characteristic absorption of emission paths. Some parameters as HS-,TOC, COD and TSS can be determined using the measures extracted from a spectrophotometer combined by mathematical models (MLR, PLSR, PCR, etc.). **Infrared probes** are less common than UV-Vis but they are gathering the interest in different WWT processes, polyhydroxyacanoates (PHA) quantification, among others. The IR spectrum can be obtained using absorbance, transmittance, and reflectance methods.	(Yang 2018) (Thomas 2017) (Mesquita 2017) (Altman 2016) (Taylor 2014) (Campisano 2013) (Vanrolleghem 2003) (Harremoës 1993) (Van der Broeke 2006) (Lourenco 2008) (Torres 2008) Gruber 2006) (Rieger2006)

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Type of sensors	Field of application	Description	Reference
Fluorescence sensors **Ultrasonic sensors** **Biomass sensors** **Electrochemical sensors** **Three-chamber microbial desalination cell (MDC)** **Two-chamber microbial electrolysis cell (MEC)**	Anaerobic digestion Aerobic systems VFA (Volatile Fatty Acids) Biogas flow	**EEM** has widely used to characterize several types of dissolved organic matter in water. This procedure can provide a great and accurate schema of essential intermediates in biological reactions at particular wavelengths. **Fluorescence sensors** have two optical fibers, the first one is to excite the culture using a light beam and the second one is to carry out the information of light collected into the detector. Depending on the element that light pass through, one or another wavelength will be monitored. This is a really good routine because these probes can control the state of the culture and determine the biomass concentration. Due to the submersion of the probes several problems can occur: bubbles can give signal error, fouling related to blocking particles can give erroneous measures. Ultrasonic sensors are used to quantify concentrations. These kinds of sensors measure the ultrasonic velocity sound between a blank sample and a wastewater sample. The principal drawbacks are the temperature changes and bubbles. Biological cells are widely used to determine the viable biomass applying an electrical field. This electrical field creates a polarization throughout the cell membrane that can by measured using capacitance of the suspension. Beside of that, other microbial electrochemical sensors have been developing using similar principles. Three-chamber microbial desalination cell (MDC), two-chamber microbial electrolysis cell (MEC) have been developed in order to monitor and control VFAs concentration.	(Jiang 2019) (Yang 2018) (Carstea 2016) (Korak 2014) (Nebbioso 2012) (McNight 2001) (Coble 1995) (Harremoës 1993)
Fourier Transform Infra-Red (FT-IR) spectrometer **Titrimetry**	VFA (Volatile Fatty Acids) Total Organic Carbon (TOC) Chemical Oxygen Demand (COD) Anaerobic digestion Suspended Solids	Volatile Fatty Acids are an important intermediate part of anaerobic digesters. Several full-scale applications have been developed in order to control and maintain this process in perfect conditions. **Fourier Transform Infra-Red spectrometer** (FT-IR) is a probe that can measures lot of parameters (COD, TOC, VFA, PA and TA) uses a reference of each compound. By comparing a sample of each compound reference, the composition can be quantified using the Beer-Lambert Law. This methodology has a heavy calibration effort, which is a handicap to use it. Some new and robust sensors have been developed using a **two-step titration** with a minimum volatilization. The ratio is obtained using this probes gives a clear idea about the relative amount of buffer is needed to neutralize VFAs and to control the digesters.	(Jiang 2019) (Wu 2019) (Vanrolleghem 2003) (Pind 2002) (Rozzi 1991)

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Type of sensors	Field of application	Description	Reference
Electrochemical sensors (Galvanic and Polarographic probes), and Optical probes	Activated Sludge kinetics Wastewater characterization Aeration processes Control oxygen rates Dissolved Oxygen (DO) Biological Oxygen Demand (BOD)	The supervision of solved oxygen process is really important and widely controlled. Different probes are used in several applications: in the industrial water treatment DO is managed in order to control the possible corrosion of equipment installed. In aquaculture, it is especially important so as to control the marine life. In wastewater management DO is extensively used because it is an indicator about how good the sludge kinetics performance is. There are three important probes to measure DO: **Galvanic and Polarographic probes:** they have an electrochemical measuring cell. DO diffuses from the liquid matrix across the gas-permeable membrane into the cell sensor. Once inside the sensor, the oxygen suffers a chemical reduction reaction and produces an electrical signal. Galvanic probe is self-polarizing due to the material properties of the anode (zinc or lead) and cathode (silver). Polarographic DO sensor requires a constant voltage to be applied to it to be polarized. **Optical probes:** This kind of sensor measures the DO throughout luminescent red and blue dyes in the probe cap. Oxygen interferes with the luminescent dyes in an effect called *quenching*.	(Jiang 2019)(Cornelissen 2018)(Campisano 2013)(Nebbioso 2012)(Hanson 2007)(Vanrolleghem 2003)(Harremoës 1993)
Standard methods **Respirometric probes** **Microbial biosensors** **Electrochemical sensors** **Fluorescence** **UVspectrophotometry**	Biological Oxygen Demand (BOD)	**Standard off-line method** to measure BOD has been widely used. The BOD5 is a measure of consumed oxygen in 5 days (or more). A direct measurement of oxygen can be performed using a potentiometric or bioluminescent electrodes). Continuous measurements of BOD can be made using **respirometric methods**. The oxygen mass balance is calculated in a respiration chamber once the wastewater has been added. **Microbial biosensors** give a rapid efficient analysis. A biosensor is an integrated apparatus that gives quantitative analytical information using a transducer. Another good continuous measurement is using an **electrochemical sensor** (explained above). **Fluorescence and UV-spectrophotometry** are great on-line monitoring methods. Using several widely examined wavelengths (λ220, λ254, and λ260) with different weights and relations, a great approach of BOD, COD and TOC can be achieved. The major drawback of these kinds of sensors is fouling, resulting in a loss of sensitivity, needing recalibration and, in some occasions, an automatic cleaning device.	(Carstea 2016) (Nebbioso 2012) (Ponomareva 2011) (Vanrolleghem 2003)

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Annex Table 2. List of relevant communication protocols

Communication Protocol	Water field of application	Description	Reference
MQTT[5]	Telemetry data Assets control	MQTT is a message-centric wire protocol designed for M2M communications that enables the transfer of telemetry-style data in the form of messages from devices, along high latency or constrained networks, to a server or small message broker. Devices may range from sensors and actuators, to mobile phones, embedded systems on vehicles, or laptops and full scale computers. It supports publish-and-subscribe style communications and is extremely simple.	(Rubión, 2019) (Parygin, 2017) (Vinoj, 2018) (Srihari, 2018)
AMQP[6]	Telemetry data Server to Server communication	AMQP is a message-centric protocol for sending interoperable messages between two or more clients. AMQP depicts the behaviour of the messaging provider and client ensuring that implementations from different vendors are truly interoperable. AMQP is a binary, application layer protocol, designed to efficiently support a wide variety of messaging applications and communication patterns. It provides flow controlled, message-oriented communication with message-delivery guarantees, and authentication and/or encryption based on SASL and/or TLS It assumes an underlying reliable transport layer protocol such as Transmission Control Protocol (TCP).	(Adelman, 2017) (Alvisi, 2019)
DDS[7]	Telemetry data Assets control	The DDS standard, Data Distribution Service, is a data-centric publish-and-subscribe technology to address the data distribution requirements of mission-critical systems. It enables scalable, real-time, reliable, high performance and interoperable data exchanges between publishers and subscribers. Moreover, it is both language and OS independent. DDS is used on business-critical applications like financial trading, air traffic control, smart grid management, and other big data applications. Also, it is used in a wide range of Industrial Internet applications. DCPS provides a set of APIs that present a set of standardised “profiles” targeting real-time information-availability for any domain. Moreover, the protocol also supports automatic “Discovery”. These APIs have been implemented in a range of different programming languages (Ada, C, C++, C#, Java, JavaScript, CofeeScript, Scala, Lua, and Ruby) and helps to ensure that DDS applications can be ported easily between different vendor’s implementations.	(Tsertou, 2015)

[5] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=mqtt
[6] https://www.oasis-open.org/committees/amqp/
[7] https://www.omg.org/omg-dds-portal/

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Communication Protocol	Water field of application	Description	Reference
JMS[8]	Server to Server communication	JMS (Oracle, 2013) is a message-centric protocol for sending messages between two or more clients. It is one of the most widely used publish-and-subscribe messaging technologies, but it also allows point-to-point messaging. Its specification, JSR 914, was developed under the Java Community Process and it is part of the Java Platform Enterprise Edition (Java EE). Mainly, JMS offers capabilities to create, send, receive and read messages to application components based on Java EE encouraging the coupling loss, the reliability and the synchrony. It is important to note that JMS is only a Java API and does not define a wire protocol, hence JMS implementations from different vendors will not interoperate.	(Fleischer, 2010)
RESTful	Server to Server communication Assets control	REST has emerged as the predominant Web API design model. RESTful style architectures conventionally consist of clients and servers. Clients initiate requests to servers; servers process requests and return appropriate responses. Requests and responses are built around the transfer of representations of resources. A resource can be essentially any coherent and meaningful concept that may be addressed. A representation of a resource is typically a document that captures the current or intended state of a resource. REST was initially described in the context of HTTP, but it is not limited to that protocol. RESTful architectures may be based on other Application Layer protocols if they already provide a rich and uniform vocabulary for applications based on the transfer of meaningful representational state.	(Sheng, 2015)
CoAP[9]	Telemetry Data Assets control	CoAP is a document transfer protocol. Mainly, it was designed to communicate over the Internet for very simple electronic devices. CoAP is being standardised by the Internet Engineering Task Force (IETF) Constrained Restful Environments (CoRE) Working Group. CoAP is focused on providing communication capabilities to small low power sensors, switches, valves and resource constrained internet devices such as Wireless Sensor Networks (WSNs). Moreover, it is designed to easily translate to HTPP for simplified RESTful web integration. CoAP is lightweight, simple and runs over UDP (not TCP) with support for multicast addressing. CoAP is based on RESTful architecture and hence, it supports a client/server programming model where the resources are server-controlled abstractions made available by an application process and identified by Universal Resource Identifiers (URIs). It is important to note that CoAP supports resource discovery.	(Anjana, 2015), (Lee, 2018)
OGC SOS[10]	Server to Server Communication	C SOS provides a standardised interface to manage sensors in an interoperable way. The ndard defines a Web Service interface, based on SOAP, which allows querying ervations, sensor metadata, as well as representations of observed features. Further, standard defines means to register new sensors and to remove existing ones. Also, it ines operations to insert new sensor observations.	(Stasch, 2017), (Hussain, 2015), (Yang, 2018)

[8] https://www.oracle.com/technetwork/articles/java/introjms-1577110.html
[9] https://tools.ietf.org/html/rfc7252
[10] https://www.opengeospatial.org/standards/sos

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Communication Protocol	Water field of application	Description	Reference
Digital Delta[11]	Server to Server Communication	The Digital Delta is a public-private initiative for cooperation in ICT in the water sector that works to create a chain of services to improve water management with online digital tools such as flood forecasting, water-stress warning and analytic hydrology tools. Companies like HydroLogic or Nelen & Schuurmans, work together with other water authorities for its implementation.	(Min, 2016), (Rooney, 2013)
FIWARE NGSI[12]	Server to Server Communication	VARE NGSI Context Management specifications are based in the NGSI Context nagement specifications defined by OMA (Open Mobile Alliance). They take the form of ESTful binding specification of the two interfaces, namely NGSI-9 and NGSI-10, one used exchanging information about the availability of the context information (discovering ts, subscriptions for context availability and registration of context information) and other to about entities and their attributes, respectively.	(Zyrianoff, 2018)
FIWARE NGSI-LD[13]	Server to Server Communication	VARE NGSI-LD is the evolution of FI-WARE NGSI to better support linked data (entity ationships'), property graphs and semantics (exploiting the capabilities offered by JSON- . and the main difference is the addition of a new type of attribute, "Relationship" ended to link one Entity with to another Entity.	(Kamienski, 2019) (Lopez-Morales, 2019)

11 https://github.com/DigitaleDeltaOrg/dd-api-spec
12 https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/FI-WARE_NGSI_Open_RESTful_API_Specification
13 https://fiware-datamodels.readthedocs.io/en/latest/ngsi-ld_faq/index.html

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Annex Table 3. Communication protocols characteristics

Standard	Abstraction	Architecture	Interoperability	Performance	Subscription Control	Data Serialization	Standardization Entity	Licensing Model	Dynamic Discovery	Transactions	Security	FIWARE Support
MQTT	Publish / Subscribe	Brokered	Partial	1000 messages per second and broker	Topics with hierarchical matching	Undefined	OASIS	Open Source & Commercial	N	N	Authentication based on username/password SSL or TSL data encryption	Y[14]
AMQP	Publish / Subscribe	Brokered	Partial	1000 messages per second and broker	Exchanges, Queues and bindings	AMQP type system or used defined	OASIS	Open Source & Commercial	N	Y	SASL authentication, TLS for data encryption	Y[15]
DDS	Publish / Subscribe	Global Data Space	Yes	1000 messages per second	Partitions, Topics with message filtering	CDR	OMG's RTPS and DDSI standards	Open Source & Commercial	Y	N	Vendor specific but typically based on SSL or TLS with proprietary access control	Y[16]

14 https://github.com/Fiware/tutorials.IoT-over-MQTT
15 https://fiware-iotagent-ul.readthedocs.io/en/latest/installationguide/index.html
16 https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/Middleware_Open_API_Specification

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Standard	Abstraction	Architecture	Interoperability	Performance	Subscription Control	Data Serialization	Standardization Entity	Licensing Model	Dynamic Discovery	Transactions	Security	FIWARE Support
JMS	Publish / Subscribe	Brokered	No	1000 messages per second and broker	Topics and Queues with message filtering	Undefined	JCP JMS standard	Open Source & Commercial	N	Y	Vendor specific but typically based on SSL or TLS. Commonly used with JAAS API.	Y[17]
RESTful	Request / Reply	Client / Server	Yes	100 messages per second	N/A	N	Is an architectural style rather than a standard	HTTP available for free on most platforms	N	N	Typically based on SSL or TLS.	Y
CoAP	Request / Reply	Client / Server	Yes	100 messages per second	Provides support for Multicast addressing	Configurable	IETF CoAP standard	Open Source & Commercially	Y	N	DTLS	Y[18]
Digital Delta	Request / Reply	Client / Server	Yes	100 messages per second	-	Y	HydroLogic, Nelen & Schuurmans, Deltares	Open Source	Y	N	Not provided natively	N

[17] https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/FIWARE.Feature.Cloud.ServiceManager.QueryManager
[18] https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/FIWARE.OpenSpecification.IoT.Backend.DeviceManagement_R5#IoT_Agent:_LWM2M.2FCoAP

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Standard	Abstraction	Architecture	Interoperability	Performance	Subscription Control	Data Serialization	Standardization Entity	Licensing Model	Dynamic Discovery	Transactions	Security	FIWARE Support
OGC SOS	Request / Reply	Client / Server	Yes	100 messages per second	Not provided natively. Provided through OGC Publish/Subscribe	Y	OGC	Open Source & Commercial	Y	Y	Not included in the standard. Provided by specific vendors	Y[19]
Fiware NGSI & NGSI-LD	Request / Reply	Client / Server	Yes	100 messages per second	Based on entity creation or update and enhanced with notification rules (filtering attributes, multiple parameters...)	Only XML as data serialization	OMA	Open Source	Y	Y	Do not provide native authentication nor any authorization mechanisms. It is provided by FIWARE GEs	Y

19 https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/IDAS

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Annex Table 4. List of relevant water data models

Data model	Water field of application	Description	Reference
WaterML2[20]	Water Supply and Distribution Data Management and Smart City Services River Basin Management Water Reuse and Recycling Water Scarcity and Droughts Flood Risk Management	Common exchange format for hydrological time-series) including information regarding quality, validity/interpolation and remarks. Developed within the OGC Hydrology Domain Working group. Existing standards like GML and Observations & Measurements has been used to build it. Fully supported to OGC communication protocols such as OGC SOS and OGC WPS.	(Hussain, 2015) (Sadler, 2016) (Lorraine, 2015)
HY_Features[21]	Water Supply and Distribution Data Management and Smart City Services Sustainable Development, Circular Economy, & Ecosystem Services Water-Energy Nexus River Basin Management Water Reuse and Recycling Management of the Water Cycle in Industry Flood Risk Management	OGC implementation Standard to define a common conceptual feature model for use in identification of typical features of the hydrology domain. The model describes types of surface hydrologic features by defining fundamental relationships among various components of the hydrosphere. Moreover, it also links hydrologic observations to their feature of interest, enables systems to ambiguously link data between systems and domains, allows aggregation of cross-referenced features into integrated data sets and data products, enable cross-domain or multidiscipline services to communicate.	(Dornblut, 2013) (Looser, 2014) (Anzaldi, 2014)
Digital Delta Data models[22]	Flood Risk Management Water Scarcity and Droughts	The Digital Delta is a public- private initiative for cooperation in ICT in the water sector that works to create hydrological observations data models for flood forecasting, water-stress warning and analytic hydrology tools. Companies like HydroLogic or Nelen & Schuurmans, work together with other water authorities for its implementation.	(Rooney, 2013)

[20] http://www.opengeospatial.org/standards/waterml

[21] http://www.opengeospatial.org/projects/groups/hydrofeatswg

[22] https://github.com/DigitaleDeltaOrg/dd-api

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Data model	Water field of application	Description	Reference
HydroNET4	Flood Risk Management	1D: Timeseries, modeltimeseries, ensemble timeseries for hydro & meteorological observations and model predictions 2/3D: Grids, modelgrids ensemblegrids for hyddro & meteorological observations and model predictions	
RiverML[23]	River Basin Management	RiverML, currently under development, is a one-dimensional standard built on WaterML 2.0 (OGC) that by providing a common transfer data model language for conveying a description of river channel and floodplain geometry and flow characteristics.	(Jackson, 2014)
SWEET[24] **(Semantic Web for Earth and Environmental Terminology)**	Wastewater and Storm Water Collection (including Flood Risk Management) River Basin Management Water Scarcity and Droughts Sea Water Sustainable Development, Circular Economy, & Ecosystems Services	SWEET is a suite of ontologies of environmental domain that can be used for water management. SWEET consists of nine top-level concepts/ontologies (Representation, Process (microscale), Phenomena (macroscale), Matter, Realm, Human Activities, Property (observation), State (adjective, adverb), and Relation (verb)).	
GWML2[25]	Groundwater Management	OGC standard to exchange groundwater related information including conceptual and logical. It is supported by Web Feature Service (WFS) and Sensor Observation Service (SOS). GWML2 captures the semantics, schema and encoding syntax of key groundwater data, to enable information systems to interoperate with such data.	(Brodaric, 2018) (Simons, 2015) (Beaufils, 2019)

[23] http://tools.crwr.utexas.edu/RiverML/index.html

[24] https://esipfed.github.io/stc/sweet_lode/reprDataModel.html

[25] https://www.opengeospatial.org/standards/gwml2

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Data model	Water field of application	Description	Reference
O&M[26]	Water Supply and Distribution Data Management and Smart City Services River Basin Management Water Reuse and Recycling Water Scarcity and Droughts Flood Management	The O&M Standard defines XML schemas for observations, and for features involved in sampling when making observations. These provide document models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities.	
HydroML[27]	Water Supply and Distribution River Basin Management Water Reuse and Recycling Flood Risk Management Water Scarcity and Droughts	HydroML, a product of the National Water Information System of the Water Resource, defines XML implementation to exchange hydrologic data between persons and organizations, data collection devices and data bases and to be served, received, and processed on the Web.	(Kanwar,2010)
INSPIRE[28]	Water Supply and Distribution River Basin Management Water Reuse and Recycling Flood Risk Management Water Scarcity and Droughts	INSPIRE-Hydrography provides a data specification to facilitate the interoperability of hydrographic information between member states, including the description of the sea, lakes, river and other waters, with their phenomena. It concerns with the network of water bodies and relating structures and objects. Reference Systems, units of measure, data quality and metadata are also taking to account in the data models.	(Eriksson, 2018) (Vacariu, 2015)
FIWARE Data models	Water quality monitoring Weather information Flood and Water Pollution Alert	Data models harmonized to enable data portability for different applications including, but not limited, to Smart Cities. The data models are used together with FIWARE NGSI. It is important to note that the new data models can be created.	(Kamienski, 2019) (Lopez-Morales, 2019) (Zyrianoff, 2018)

[26] http://www.opengeospatial.org/standards/om

[27] https://water.usgs.gov/XML/NWIS/nwis_hml.htm

[28] https://inspire.ec.europa.eu/id/document/tg/hy

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Data model	Water field of application	Description	Reference
SAREF (Smart Applications REFerence Ontology)	Water Quality Monitoring On development the water extension	SAREF ontology provides the semantic interoperability necessary to share the information related to: (i) smart city including water quality monitoring; (ii) industry and manufacturing including production equipment, batches and material; (iii) automotive; (iv) eHealth/Ageing-well; (v) Wearables; (vi) Smart agriculture and food chain.	-

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Annex Table 5. Data driven models applied to flood, sewage, pollution and urban water management

Data driven model	Field of application	Description	Reference
ANN (Artificial Neural Network) including BPNNM (Back Propagation Neural Networks), MLP (Multi Layer Perceptron), BFGSNN (Broyden Fletcher Goldfarb Shanno Neural Network) among others.	Short term flood prediction Long term flood prediction Sewer management Wastewater Treatment Operation	Real-time flood prediction Hourly stage level, streamflow and peak flow, water surge level and flood prediction Daily rainfall-runoff and flood prediction Weekly stream prediction Monthly stream, precipitation, reservoir levels and discharge prediction Seasonal water levels and heavy rainfall prediction Wastewater Treatment Process performance prediction Wastewater Treatment Control Soft Sensing	(Ghose, 2018) (Zhang, 2018) (Gazendam, 2016) (Kim, 2016) (Shamim, 2016) (Kourgialas, 2015) (Aichouri, 2015) (Deo, 2015) (Çoruh et al. 2014) (Lohani, 2014) (Elsafi, 2014) (Bello, 2013) (Singh, 2013) (Danso-Amoako, 2012) (Rezaeian-Zadeh, 2012) (Rezaeian-Zadeh, 2012) (Yestilmezsoy, 2011) (Luccarinno, 2010) (Ju, 2009) (Lin, 2006) (Pereira, 2006) (Shoo, 2006) (Kim, 2001)
AR (Auto-regressive)	Short term flood prediction	Hourly stage level & streamflow Prediction	(Pereira, 2006)
SVM (Support Vector Machines) & SVR (Support Vector Regression)	Short term flood prediction Long term flood prediction Wastewater Treatment Operation	Daily streamflow prediction Monthly stream prediction Soft Sensing Membrane separation design optimization Wastewater control purposes	(Soleimani, 2013) (Yu, 2013) (Lin, 2012) (Pan, 2010) (Lin, 2006)
Hybrid ML methods	Short term flood prediction	Real-time flash flood, flood quantile estimation and rainfall-runoff prediction Hourly water level, watershed, rainfall-runoff and flood area prediction Daily rainfall-runoff, stream flow and flash floods prediction	(Doycheva, 2017) (French, 2017) (Jimeno, 2017) (Young, 2017) (Nanda, 2016) (Chang, 2014) (Rezaeianzadeh, 2014)
RT (Regression Trees) & DT (Decision Trees)	Long term flood prediction Sewage management Wastewater Treatment Operation	Annually floodplain forests prediction Likelihood pipe blockage prediction	(Cunningham, 2017) (Bailey, 2016) (Ma, 2009)

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

Data driven model	Field of application	Description	Reference
LR (Logistic Regression)	Sewage management	Prediction condition of sanitary sewer pipe	(Mohammadi, 2019)
BN (Bayesian Networks)	Sewage management	Water pipe failure detection	(Lin et al., 2015)
PCA (Principal Component Analysis)	Wastewater Treatment Operation	Fault detection Process understanding	(Villez, 2016) (Liukkonen, 2013) (Dürrenmatt, 2011) (Zhang, 2010) (Yoo, 2003)
Fuzzy Logic	Wastewater Treatment	Control and prediction purposes	(Marsili-Libelli, 2008)
Clustering	Wastewater Treatment Operation	Increase process understanding and feature engineering	(Gibert, 2010) (Dovzan, 2011a) (Dovzan, 2011b) (Aguado, 2008)

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751

scoreWATER

This project hase recieved funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 820751