2017

Monitoring of performance indicators by the water services regulator using wireless sensor networks

Brenda Lufuso Anzagi
Fuculty of Information Technology (FIT)
Strathmore University

MONITORING OF PERFORMANCE INDICATORS BY THE WATER SERVICES REGULATOR USING WIRELESS SENSOR NETWORKS

BRENDA ANZAGI

A Thesis submitted in partial fulfillment of the requirements of the Degree of Masters of Science in Information Technology at Strathmore University

Faculty of Information Technology

Strathmore University

Nairobi, Kenya

June, 2017

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Brenda Lufuso Anzagi

8th June 2017

Approval

The thesis of Brenda Lufuso Anzagi was reviewed and approved by the following:

Prof. Ismail Ateya
Associate Professor - Faculty of Information Technology,
Strathmore University

Dr. Joseph Orero,
Dean, Faculty of Information Technology,
Strathmore University

Professor Ruth Kiraka,
Dean, School of Graduate Studies,
Strathmore University

Abstract

The Kenyan water services regulator is required to report annually to the public on issues of water supply and sewerage services and hence maintain a national database of the same. This means that the quality of data provided by the utilities should be reliable data, to ensure efficient and effective management. The lack of appropriate tools to precisely measure data among water utilities impacts on the accuracy of data submitted to the regulator for monitoring. The data validation process is also time consuming as it is rigorous and involves corroboration with other data sources. Leveraging on technology is vital to ensure improved water utility management. This thesis seeks to enhance performance monitoring of water utilities using wireless sensor networks. Challenges experienced by the water services regulator are identified and the smart water infrastructure is the proposed solution. A smart water infrastructure integrates sensors, controls, and analytical components to ensure that the data captured within the system is consistent and precise. The smart water data generated provides visual monitoring of utilities data and provides real-time data for analysis, hence providing an accurate and timely basis for performance. The proposed smart water prototype leverages on the Internet of Things concept, which is a system of interrelated computing devices, which have unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. Since a typical smart water infrastructure requires large scale deployment, the use of wireless sensor network architecture is used. It includes data management and processing, actuation and analytics. Primary and secondary data was used to identify the requirements, which then formed an input to the prototype's design, implementation and testing. The prototype captures selected water quality and water volumes data. It is then evaluated and found to be effective in monitoring the performance of water utilities and reduces the time spent on validating utilities' operational data. The prototype also demonstrates timely decision making. It is hoped that this study will be applied practically to aid the Kenyan water regulator's monitoring efforts.

Table of Contents

List of Figures

List of Tables

List of Abbreviations/ Acronyms

AMI	Advanced Metering Infrastructure
AMR	Automated Meter Reading
CSS	Cascading Style Sheets
GIS	Geographic Information System
GPRS	Global Packet Radio Service
GSM	Global System for Mobile Communication
HCI	Human Computer Interaction
HTML	Hypertext Markup Language
HTTPS	Hypertext Transfer Protocol Secure
IoT	Internet Of Things
JSON	JavaScript Object Notation
O+M	Operation and Maintenance Costs
PCB	Printed Circuit Board
UML	Unified Modeling Language
WARIS	Water Regulation Information System
WASREB	Water Services Regulatory Board
WSB	Water Services Board
WSN	Water Supply Network (Singapore)
WSP	Water Service Provider

Definition of Terms

Automatic Meter Reading (AMR) — The technology of automatically collecting consumption, diagnostic, and status data from water meter or energy metering devices (gas, electric) and transferring that data to a central database for billing, troubleshooting, and analyzing (Mutchek & Williams, 2014).

Advanced Metering Infrastructure (AMI) — An integrated system of smart meters, communications networks, and data management systems that enables two-way communication between utilities and customers (Mutchek & Williams, 2014).

Internet of Things (IoT) — A dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual "things" have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network (Vermesan & Freiss, Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems, 2013).

JavaScript Object Notation (JSON) — A lightweight data-interchange format that uses text, which makes it easy for machines to parse and generate. JSON is language independent. (www.json.org)

Supervisory Control and Data Acquisition (SCADA) — A system that deploys multiple software and hardware elements that allow industrial organizations to monitor, gather, and process data, interact with and control machines and devices such as valves, pumps, motors, and more, which are connected through human-machine interface software and record all the events into a log file (What is SCADA?: Inductive Automation, 2016)

Turbidity — In water quality, this refers to the cloudiness or haziness of a fluid caused by large numbers of individual particles that are generally invisible to the naked eye, similar to smoke in air. The measurement of turbidity is a key test of water quality. Turbidity adversely affects the efficiency of disinfection. Turbidity is also measured to determine what type and level of treatment are needed (World Health Organization, 2006).

Acknowledgements

First of all, I am grateful to the Almighty God for giving me good health and strength to enable me to complete this thesis.

Secondly, I would like to thank my supervisor, Professor Ismail Ateya for his dedication and guidance throughout the process.

Thirdly, my employer, the Water Services Regulatory Board, for giving me the opportunity for to undertake the course and understanding the dedication and effort required to complete the project. The team has been amazing and I express my sincere gratitude.

My family and friends have also provided me with overwhelming support and encouragement. Thank you for believing in me.

Last but not least, my dear husband, Victor Sudi, who has walked with me through this journey.

Dedication

To Clive, Cristina and the unborn.

Chapter 1: Introduction

1.1 Background of the Study

The current Kenyan Water Act requires that water utilities provide operational data for monitoring by the regulator, Water Services Regulatory Board (WASREB). The utilities submit the data using an online data collection tool using Water Regulation Information System, WARIS. Once the data is submitted by the utilities to the regulator, it is analyzed. The water utilities are then ranked on the basis of their performance and areas of improvement identified. The recommendations prescribed are expected to ultimately improve the water services provided to the consumers. This provides the basis of the *Impact* report, which is then published. *Impact* is WASREB's annual publication for performance reporting of the Kenya water services sector. The *Impact* report is currently at its 9^{th} Edition. It analyses water utilities based on Key Performance Indicators (KPIs), which are nine in number. These indicators are clustered into three main categories; Quality of Service Indicators, Economic Efficiency Indicators and Operational Sustainability Indicators. The Quality of Service Indicators are Water Coverage, Hours of Supply and Drinking Water Quality. The Economic Efficiency Indicators are Personnel Expenditures as a Percentage of O+M Costs, O+M Cost Coverage and Revenue Collection Efficiency. The Operational Indicators are Metering Ratio, Staff Productivity and Non-Revenue Water. The definitions of these KPIs are on Appendix A.

According to Issue no. 8, the quality of data used in performance analysis of commercialized utilities is the most important ingredient in enhancing the credibility of Impact Report (Water Services Regulatory Board, 2015). Quality data is crucial for effective decision making in the monitoring of the water utilities, for instance, it will ensure adequate planning of investments, as well monitoring the realization of the investments. It ensures that investments are timely and well targeted (Water Services Regulatory Board, 2016).

Leveraging on technology is inevitable for the ultimate improvement on the quality of data that an organization utilizes. According to a presentation by the International Water Association (International WaterAssociation, IWA, 2016), water security will be enhanced by integration of smart water infrastructures, which will provide real-time and reliable data. Future developments will rely on such infrastructure. The Internet of Things (IoT) concept has the potential to enhance an organization's competitiveness. It is an important driver for the development of an information based economy and society. Smart connectivity with existing networks and context-aware computation using network resources is an indispensable part of IoT (Gubbi, Buyya, Marusic, & Palaniswami, 2013). Integrated networking, information processing, sensing and actuation capabilities allow physical devices to operate in changing environments (Vermesan & Freiss, Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems, 2013).

Numerous research and application projects have been done in different application fields. For instance, smart cities, smart energy, smart transportation, smart homes and buildings, smart manufacturing, smart health and food and water security are some of the IoT application areas. The extent of IoT applications will be driven by the diverse needs of potential uses which address societal needs. The research puts into perspective the water regulator's monitoring needs. Application of IoT to ensure food and water security is a prioritized topic. (Vermesan & Freiss, Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems, 2013).

Examples of previous researches that have analyzed the implementation of specific smart technology systems have been on automated meter reading (AMR) and advanced metering infrastructure (AMI) for water infrastructure (Mutchek & Williams, 2014). The stated research examples illustrate how smart water meters can be used for residential and commercial water consumption billing. However, integrative, strategic, and macro-level discussions of smart water grids are lacking in academic and other literatures (Mutchek & Williams, 2014). According to Mutchek and Williams, there exist barriers to the adoption of smart water grids. There has been lack of funding for research and development, economic disincentives as well as the presence of institutional and political structures that favor the current system.

As the practical use of IoT in water networks spans over a large geographical area, the use of Wireless Sensor Networks is proposed. This is due to the small size of the sensor nodes, and their capabilities of transmitting data remotely. The wireless sensors can be deployed at locations that are not easily accessible or are impossible to access. Wireless Sensor Networks are also cheaper and their usage is rapidly growing in the industrial and commercial fields

(International Electrotechnical Commission, 2014). This paper proposes how Kenyan water utilities can be monitored by the water services regulator, using wireless sensor networks.

1.2 Problem Statement

Data collection is any process whose intention is to acquire or assist in the acquisition of data (Sekaran, 2004). The data collection process by the water service regulator should aim to be effective and efficient, and aim to capture reliable data from the utilities that it regulates. Currently, the data that is captured by the water utilities is done through an online system, WARIS. The online system has rigorous validation checks, and ensures two levels of validation before final submission for analysis. However, there is no way of ensuring precise capturing of the actual data. For example, a user may erroneously input inaccurate values. To mitigate this risk, the regulator undergoes various cross-checks with inspections reports and other data sources for consistency. This process is usually time-consuming for all the utilities which are currently over 90. It then becomes necessary to enable real-time data capture.

This research aims to provide a solution on improved data captured by Kenyan water utilities. The lack of appropriate tools to precisely measure operational data among utilities impacts on the accuracy of data submitted to the regulator (Water Services Regulatory Board, 2016). Although there have been efforts to improve the quality of the data by upgrading the data collection tool, WARIS, which has rigorous validation checks, there still lies a gap on the utilities' side in terms of providing precise data.

To address this, a smart water grid, a concept that will network and automate monitoring and control by leveraging on the Internet of Things is proposed. The smart water prototype enables the real-time collection of operational data using wireless sensor networks. With the smart water infrastructure, new data changes are captured in real-time and transferred to the target system, hence removing the need for inefficiency in the processes.

1.3 Research Objectives

i. To investigate the challenges that the regulator has when capturing monitoring data for Kenyan water utilities
ii. To review the use of wireless sensor network technology as a method for capturing data
iii. To propose a prototype using wireless sensor networks to capture data for monitoring
iv. To evaluate the proposed prototype

1.4 Research Questions

i. What are the challenges that the regulator has when capturing monitoring data for Kenyan water utilities?
ii. What are the features of a wireless sensor network?
iii. How can wireless sensor networks be used to capture data for monitoring?
iv. How will the prototype be evaluated?

1.5 Justification

Quality data is what drives an institution. Capturing of this data in the most effective and efficient way is critical. The purpose of a regulator's existence is to make rules, monitoring and enforcement activities. These activities are data driven. Leveraging on information technology to get accurate data is inevitable as it is one of the primary drivers besides policy and money. Innovations for water use – particularly consumption and treatment need to be explored. The embracing of enabling technologies such as nanotechnology, communications, sensors, embedded systems, cloud networking, network virtualization and software is essential to provide the capability to be connected all the time, and everywhere. This research will support future IoT innovations affecting the water sector.

1.6 Scope and Limitations

The scope of the research was limited to selected operational data provided for by the water utilities for monitoring. This selected operational data was water production volumes and water quality. The research did not go beyond all the other probable uses of data that can be generated from a smart water infrastructure.

Chapter 2: Literature Review

2.1 Introduction

The literature review focuses on the research objectives. From the background and the problem statement, it becomes evident that various issues need to be discussed. The first one is the process of data collection by the Kenya water services regulator. This is then enriched by discussing the characteristics of good quality data. The proposed solution leverages on the use of a smart water architecture. The smart water environment and how it works, its benefits and its applications come to light.

2.2 The Data Collection Background and Process

The current Kenyan Water Act requires that water utilities provide operational data for monitoring by the regulator. The utilities submit the data using an online data collection tool, and once the submitted data is analyzed, *Impact* report is published. *Impact* is WASREB's tool for performance reporting. It analyses water utilities based on nine Key Performance Indicators (KPIs). The 9 KPIs are Drinking Water Quality, Hours of Supply, Metering Ratio, Non-Revenue Water, O+M Cost Coverage, Personnel Expenditure as a Percentage of O+M Costs, Revenue Collection Efficiency, Staff Productivity and Water Coverage.

According to Impact Report, Issue no. 8 (Water Services Regulatory Board, 2015) on the performance of utilities, the quality of data used in performance analysis of commercialized utilities is the most important ingredient in enhancing the credibility of the *Impact* Report. The data that is contained in the publication is extracted from the regulator's online data collection tool, the Water Regulation Information System (WARIS). WARIS is the system that WASREB uses to makes requests for data submission from the utilities. The data in the system is organized into datasets. The Figure 2.1 represents the path a dataset goes through the data collection process until it is published.

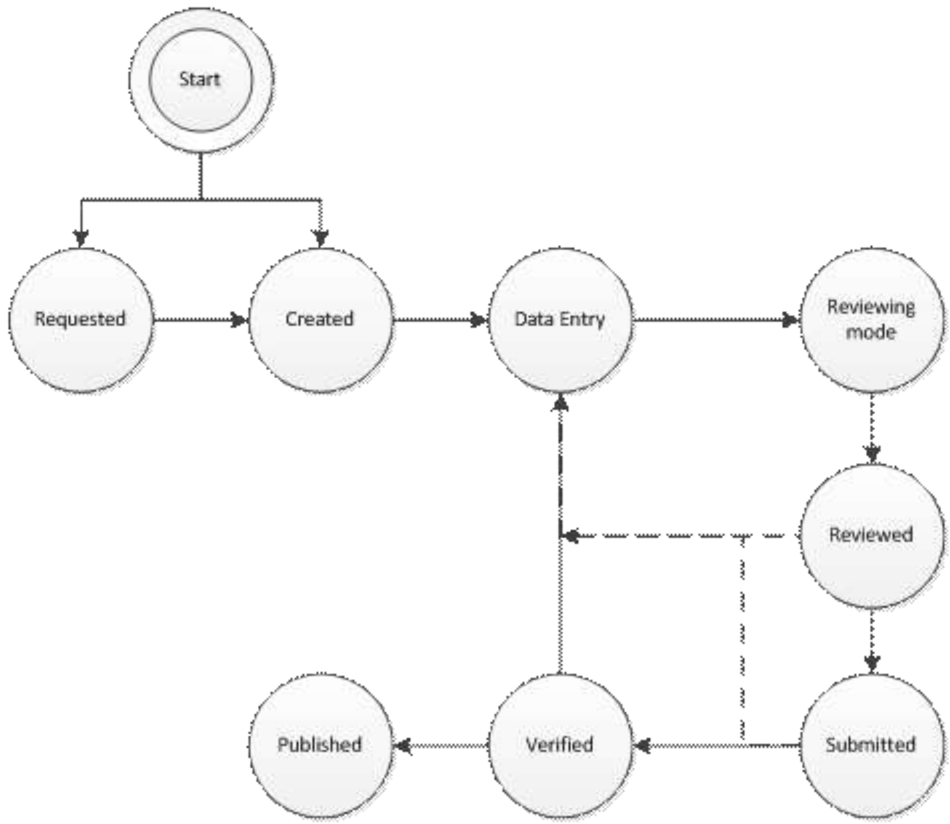

Figure 2:1 Data Set Status Path

(Water Services Regulatory Board, 2016)

The data is usually requested by the regulator. The water utilities receive the request, and create the dataset, which enables the capturing of data. After the data entry is complete, the data is reviewed internally, where it undergoes a verification process within the utilities. The online system has rigorous validation checks, and ensures two levels of validation before final submission for analysis. It is then submitted externally to WASREB, where it is further cross-checked against tariff information submissions, inspection reports and annual licensee reports prior to its analysis and publishing. This process is usually time-consuming for all the utilities which are currently over 90. In addition to that, there is the critical question: What if the cross-checked data sources are also inconsistent? In cases where cross-checks show data inconsistency, WSPs and WSBs are contacted directly to confirm the accuracy or make the necessary corrections (Water Services Regulatory Board, 2015), and in some cases, the data is deemed as ‘not credible’. Figure 2.2 shows the data collection cycle (Water Services Regulatory Board, 2016).

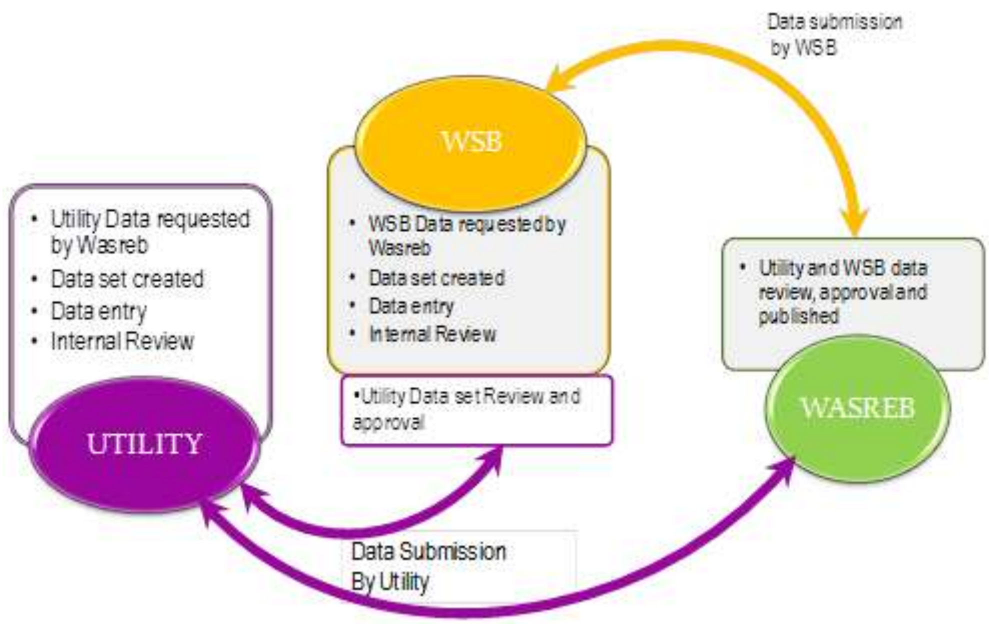

Figure 2:2 Data Collection Cycle

(Water Services Regulatory Board, 2016)

2.3 What is Good Quality Data?

2.3.1 Characteristics of Good Quality Data

As information systems are embraced with the aim of making business operations efficient, companies continue to generate big data. Big data refers to large volumes of data, which can either be structured or unstructured. Big data contains massive, variable and complicated data sets. Emphasis is on how the companies use the big data, and hence make meaningful sense out of it. As organizations use and process big data, extracting real and high-quality data becomes an urgent issue (Li Cai, 2015). Good quality data is data which conforms to the relevant uses and meets user requirements.

Not only has data quality depended on its own features, but also on the business environment using the data. The business environment includes the business processes and the business users, both external and internal. Generally, both data producers and consumers develop data quality standards either directly or indirectly. However, in the age of big data, with the diversity of data sources, data users are not necessarily data producers. Since it is very difficult to measure the quality of data, Li Cai (Li Cai, 2015) proposed the data quality framework, which is composed of five dimensions of data quality: - availability, usability reliability, relevance and presentation quality.

From Li Cai's framework (Li Cai, 2015), availability refers to the degree of convenience for users to obtain data and related information. Data availability is divided into the three elements- which are accessibility, authorization, and timeliness. Usability refers to whether the data is useful and meets users' needs. Reliability refers to whether the data can be trusted and consists of accuracy, consistency, completeness, adequacy, and audit ability elements. Relevance describes the degree of correlation between data content and users' expectations. Presentation quality provides a valid description method for the data, which allows users to fully understand the data. Its dimensions are readability and structure.

2.3.2 Good Quality Data and Decision Making

Data is regarded to be of high quality if it is "Fit for Use" in its intended operational, decision-making and other roles (Herzog, Scheuren, & Winkler, 2007). Fitness of this data is linked to conformance to a set of standards. Many organizations realize that their data is too messy to analyze without cleansing (Herzog, Scheuren, & Winkler, 2007).The seven most cited properties of data are relevance, accuracy, timeliness, accessibility and clarity of results, comparability, coherence and completeness, are well integrated into Li Cai's data quality framework. Process improvement is one of the key practical ways of improving the quality of data.

In order to improve services management and enhance performance accountability, organizations require information that is fit for purpose, for instance the need for good-quality information to make judgments about the efficiency, effectiveness and responsiveness of an organization's services. Any typical organization will need to make complex decisions about its priorities and the use of resources. Apart from the organization itself, the stakeholders, who include the service users and members of the public more widely, need accessible information to make informed decisions. Regulators and government departments must satisfy their responsibilities for making judgments about performance and governance (Audit Commission for Local Authorities and The National Health Service in England, 2007).

The Audit Commission for Local Authorities report for 2007, states that a lot of time and money is spent on the activities and systems involved in collecting and analyzing data. The Kenyan water services regulator also undergoes the same predicament (Water Services Regulatory Board, 2015). The Audit Commission report goes further to say that despite this, there is still lack of confidence in some of the information produced from these data. Since there is an increasing reliance on performance information in performance management and assessment, the need to demonstrate that the underlying data is reliable has become more critical (Audit Commission for Local Authorities and The National Health Service in England, 2007).

Public bodies, such as the water services regulator and the water utilities are accountable for the public money they spend. Public organizations must manage competing claims on resources to meet the needs of the communities they serve, and plan for the future. The financial and performance information they use to account for their activities, both internally and externally to their users, partners, commissioners, and to government departments and regulators, must be based on good quality data (Audit Commission for Local Authorities and The National Health Service in England, 2007).

2.3.3 Capturing of Quality Data

Data requirements should ideally be tightly defined around an organization's service and purpose, taking into account the people's needs. Arrangements for collecting and recording the data, and reporting them as performance information, should be integrated as far as possible into the business planning and management processes of the organizations (Audit Commission for Local Authorities and The National Health Service in England, 2007).

Data should be collected and reported once only, on the principle of 'getting it right first time', with clear and simple actions. The manual intervention, if any, should be limited. The aim should be to avoid wastage of time and money spent on duplicated recording, cleansing data, interfacing between different information systems, matching and consolidating data from multiple databases and developing or maintaining multiple, often outdated, systems. This helps to reduce the burden of administration as well as helping to ensure more accurate and timely data (Audit Commission for Local Authorities and The National Health Service in England, 2007).

All systems and processes supporting a body's key data requirements should be reviewed periodically to ensure that data is collected according to the relevant policies and definitions. This will ensure that the processes used remain fit for purpose and are applied consistently (Audit Commission for Local Authorities and The National Health Service in England, 2007).

Challenges in data quality in the big data era according to Li Cai, is the fact that there is a wide range of data sources. This diversity introduces complexity due to the abundant data types and complex data structures and thus, increases the difficulty of data integration. Data volume is also tremendous, and it is difficult to judge data quality within a reasonable amount of time. Data changes very fast and the "timeliness" of data is very short (Li Cai, 2015). This then, will necessitate higher requirements for processing technology.

2.3.4 Data Monitoring Tools and Techniques

Real-time data collection refers to the ability to collect data automatically on demand, and have the data delivered and analyzed quickly enough to effect monitoring and control decisions. Team Ray Technologies (Team Ray Technologies, 2015), a company that focuses on automation design and integration services, data transformation services, and custom software applications to the automotive, government, and manufacturing sectors provides the future trends in real-time monitoring of data. Among the trends listed, are the use of Geo referenced SCADA offsite controls, the direct integration of data collection systems with enterprise systems, the increase in wireless networking, and the increase in the demand for real-time data collection systems.

2.4 The Smart Environment

2.4.1 The Internet of Things (IoT) Concept

Potential applications of the IoT are numerous and diverse, pervading into practically all areas of every-day life of individuals. This has been referred to as the "smart life". The applicability of IoT has also been evident in enterprises and the society as a whole (Vermesan &

Freiss, Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems, 2013).

According to the Cisco Annual Security Report (Cisco, 2016), 500 billion devices are expected to be connected to the Internet by 2030. Each device includes sensors that collect data, interact with the environment, and communicate over a network. The Internet of Things (IoT) is the network of these connected devices. These smart, connected devices generate data that IoT applications use to aggregate, analyze, and deliver insight, which helps drive more informed decisions and actions.

The Internet of Things comprises a digital overlay of information over the physical world. Objects and locations become part of the Internet of Things in two ways. Information may become associated with a specific location using GPS coordinates or a street address. Secondly, sensors and transmitters can be embedded into objects. This will enable them to be addressed by Internet protocols, and to sense and react to their environments. The objects will also communicate with users or with other objects as well.

With the IoT concept, there is an attempt to connect different technologies and applications in one coherent network. The different objects or devices exist in different silo applications. The IoT Architecture, IoT-A, provides a framework that enables the different silo applications to communicate by offering a common language, that is, integration of protocols and interoperable technologies by identifying the basic components of each application silo, then breaking all those components into smaller elements. The IoT-A then arranges those elements into a more general framework for building a never-ending variety of IoT applications (Internet of Things Architecture, 2016). The Figure 2.3 illustrates the comparison between IoT interoperability and non-interoperability applications.

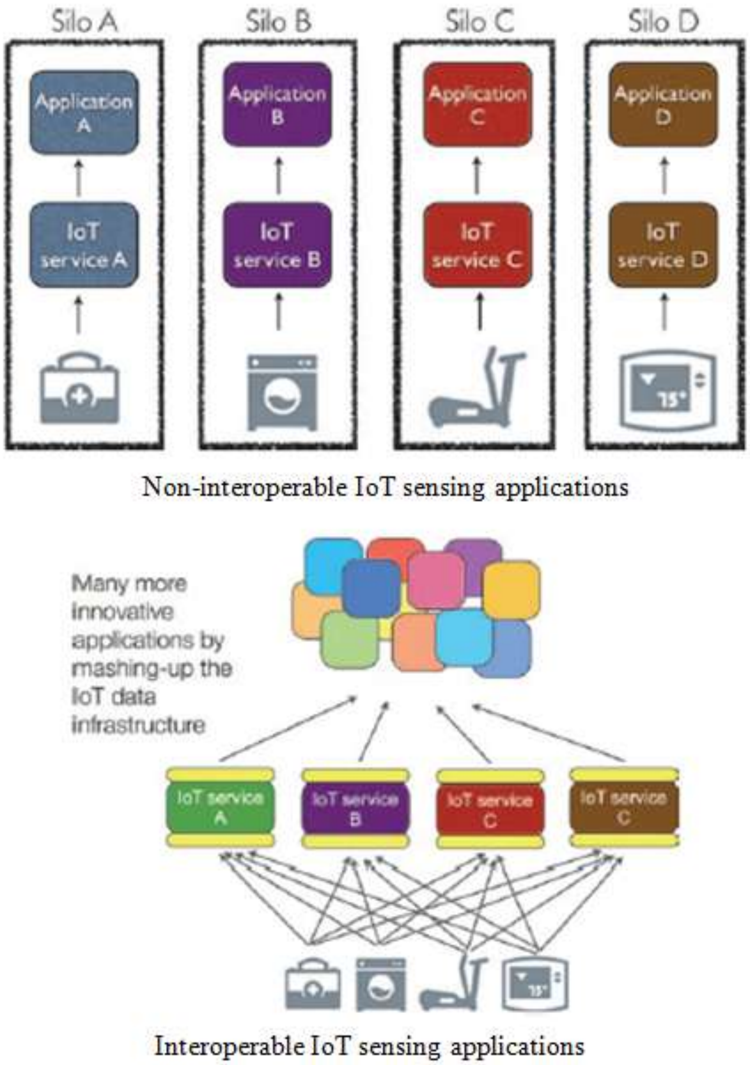

Figure 2:3 IoT: Interoperability vs. Non-Interoperability

(Vermesan & Freiss, Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems, 2013)

While there has been significant progress in IoT, there remain significant challenges in terms of interoperability and standardized modular systems architecture. There have also been privacy, security and user safety. Challenges have also emerged with how users interact with, manage and control an ensemble of devices in this connected environment.

2.4.2 IoT and Smart Water Networks

In smart water networks, IoT helps in providing actionable information to help usage be more efficient and less wasteful. Ovidiu Vermesan and Peter Friess (Vermesan & Freiss, Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems, 2013) state the Internet of Things as one of the strategic technology trends for the next 5 years. The applications areas include the domain of industrial internet where intelligent devices, intelligent systems, and intelligent decision-making represent the primary ways in which the physical world of machines, facilities, fleets and networks can more deeply merge with the connectivity.

There are many benefits of smart water grids, as described the next section of this document. Michele Mutchek and Eric Williams (Mutchek & Williams, 2014) argue that integrative analysis of multiple benefits for larger-scale smart water grid systems could provide future direction. For example, smart water grids must not just be viewed purely as a way to reduce water losses but also drought management. In addition to this, as the benefits of smart water grids accrue to a number of different actors, e.g., utilities, homeowners, and society as a whole, there are important questions as how to distribute costs among beneficiaries. There is hence a need for research and development is to enable such work.

According to a case study on the implementation of a smart water grid in Singapore (Allen, Preis, Iqbal, & Whittle, 2011), there is restricted on-line monitoring and analyzing capabilities within water distribution systems. The monitoring is frequently carried out at the water system inlets. However, there is an increasing need to reduce water wastage through leakages within the distribution network, as well as reduce other operational inefficiencies. Operational efficiencies in this context refer to issues such as power consumption. Reviewing the latest Impact report, Issue Number 9 (Water Services Regulatory Board, 2016), 43% of the water produced by the utilities is lost. This water is lost either due to technical losses which could be from water leakage. The losses can also be due to commercial losses, which include meter-reading errors, illegitimate connections and unbilled authorized consumption.

The idea of a smart water grid is gaining popularity and is considered as the best option towards improving operational efficiency. It improves the understanding and prediction of demand and consumption in water distribution systems. With a smart water grid, a utility will be able to improve the quality of the analytics done by transforming the enormous raw data into valuable and reliable information. This information must be acted on in an appropriate and timely manner. A smart water grid enables sensing, analytics and operational data measurements to be made and delivered more frequently. However, there has been integration issue. The remote sensing systems in place are not typically well integrated with analytics and modeling software, leading to non-standard, hard-to-support solutions and inefficiencies in system operation, including management of overwhelming amounts of data.

The Nairobi Water and Sewerage Company (NWSC), the largest water services provider in Kenya with over 300,000 active connections has a plan to install smart meters. Such meters are intended improve the company's billing efficiency and enable it to tap the revenue it needs to improve supply. These smart meters form part of a smart water grid. The company budgeted to spend an average of Kshs 300 million annually for three consecutive years for the smart meters project, which is targeting approximately 90,000 of the top water consumers. The smart meters project estimated cost was at $1,000 per meter, compared to the current $100 for the average 0.5-inch meter. An end to end smart infrastructure and possible reworking of the network will address the cost issue. With the exponential increase in the number of interconnected devices (Cisco, 2016), costs of integrating such a network may reduce.

There are a number of options for water utilities to consider as they evaluate Automated Meter Reading (AMR) or Advanced Metering Infrastructure (AMI), including the integration of computer information systems, Supervisory Control and Data Acquisition (SCADA), asset management or work order management, Geographic Information Systems (GIS), and other analytic engines. In any combination, water utilities have a great opportunity to improve overall performance through advanced technologies broadly captured under the banner of AMI.

2.4.3 Smart Water Networks and Regulation

According to the US Environmental Protection Agency, EPA, cost reduction and improving techniques for water monitoring is one of the water technology's top ten market opportunities. Making smart decisions about water management depends on collecting accurate and timely data. New monitoring and sensing technologies represent both an opportunity and a necessity for responsible stewardship of water systems (Innovation, 2016). Quality data is crucial for effective decision making in the planning and monitoring which forms regulations' core mandate.

With a smart water grid, the Water Services Regulatory Board (WASREB) will be able to monitor data more precisely. Smart systems are agile, capable of learning, and customer-centered. WASREB will benefit from precise operational data, which will enable the board to discharge its mandate in an effective manner (Kenya Gazette, 2016).

The smart water infrastructure will enable the regulator to effectively evaluate and recommend water and sewerage tariffs to the county water services providers and approve the imposition of such tariffs in line with consumer protection standards. In addition to that, effectively monitor compliance with standards including the design, construction, operation and maintenance of facilities for the provision of water services by the water works development bodies and the water services providers. This will in turn cascade to effectively monitor progress in the implementation of the Water Strategy and make appropriate recommendations. The regulator will be able to precisely maintain a national database and information system on water services, continue reporting annually to the public on issues of water supply and sewerage services and the performance of relevant sectors, and publish the reports and make recommendations on how to provide basic water services to marginalized areas. In addition to all this, the regulator will be able to smartly capture the regulatory levy due by the utilities.

2.4.4 Smart Water Networks and Water Utilities

Some of the benefits (Greg Weeks, 2014) that water utilities can anticipate through the business case and financial analysis, include reduction in revenue loss, water scarcity and conservation through more accurate metering and billing, enhanced water management capabilities such as enhanced tracking of water consumption and flow patterns, ability to track and predict changes in trends and demands, ability to shift water consumption to different time periods, identify anomalies in distribution or customer usage (leaks, pressure management and water quality), greater bill accuracy and enhanced customer service through enhanced billing accuracy, efficient resolution of bill inquiries with real-time data, timely notification of customer premise leaks, customer information (web) portals and innovative payment options such as pre-payment.

2.4.5 Application of Smart Water Systems

In cities around the globe, the drinking water distribution infrastructure is aging rapidly and there is a noted increasing frequency in failures. The result has been significant water losses caused by imbalances between water entering and leaving the system, inefficiencies in system operation, and concerns about the quality of drinking water that is provided to the consumer (Allen, Preis, Iqbal, & Whittle, 2011).

WaterWiSe is a platform for real-time monitoring in Singapore of water distribution systems that is used to improve system management and operation by providing integrated measurement and analytics (Allen, Preis, Iqbal, & Whittle, 2011). WaterWiSe can operate as a self-contained system with its own analysis and management interfaces, or can be integrated into a water utility's existing infrastructure and geographical information platforms. The core WaterWiSe platform has two key components: The Integrated Data and Electronic Alerts System (IDEAS) and the Decision Support Tools Module (DSTM).

The Figure 2.4 represents an overview of the WaterWiSe platform's functionality with a selection of applications enabled by both IDEAS and DSTM as well as a selection of benefits seen by the utility (Allen, Preis, Iqbal, & Whittle, 2011).

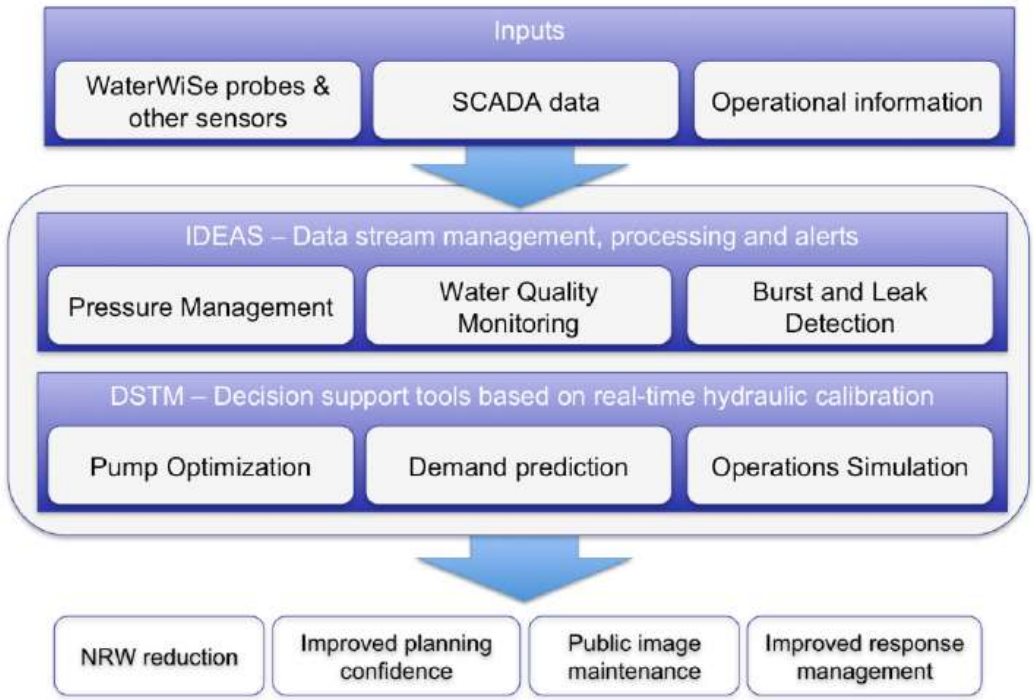

Figure 2:4: WaterWiSe Platform's Functionality

(Allen, Preis, Iqbal, & Whittle, 2011)

The WaterWiSe system has been deployed to monitor pipeline leakages. Some sensor nodes are mounted on poles, while others sensor nodes are inserted into the pipes. The data of the sensor nodes are transmitted by 3G modems (Du, Wireless Sensor Networks in Smart Cities: The Monitoring of Water Distribution Networks Case, 2016). Since 2009, WaterWiSe has been useful in several key areas: pressure anomaly detection and localization, post-event analysis, understanding pressure characteristics, real-time feedback and validation, troubleshooting, real-time modeling and sensor placement and real-time operations (Allen, Preis, Iqbal, & Whittle, 2011).

Benefiting from the study of wireless sensor networks in a water distribution network, other monitoring systems have already been implemented in different cities apart from Singapore. In Ann Arbor (Michigan), a distribution system is built for online contamination monitoring. The system applies probabilistic analysis to assess the water distribution network. In Boston, a system called PipeNet has been developed to detect and localizing leakages based on pressure, acoustic and vibration data. The sensor nodes which are powered by batteries need to transmit the data by short range communication to the gateways nodes, which are powered by the grid. Then the data are transmitted, by long range communication, to the monitoring center (Du, Wireless Sensor Networks in Smart Cities: The Monitoring of Water Distribution Networks Case, 2016).

The data transmission of the above mentioned systems are based on electromagnetic waves. On the other hand, a system called MISE-PIPE uses magnetic induction based wireless communications. Coils are winded on the pipelines, and form a magnetic induction waveguide to relay the data. However, the data rate of the system is limited due to the small bandwidth (Du, Wireless Sensor Networks in Smart Cities: The Monitoring of Water Distribution Networks Case, 2016).

Similar studies have also been done within East Africa. A prototype was developed using wireless sensor networks for water quality monitoring and control for Lake Victoria Basin (Faustine, et al., 2014). Mwangi (Mwangi, 2016) also proposed the use of a Hall Effect Sensor for smart water metering.

2.5 Architectures, Frameworks and Designs

2.5.1 IoT and Wireless Sensor Networks

As the internet has evolved from www- which are web pages, to web2 which is the social networking web, to web3 which is the ubiquitous computing web, the need for data-on-demand using sophisticated intuitive queries increases. This then means that context aware systems must be in place and this can be accomplished, according to Gubbi et al, in three main ways. Firstly, by having a shared understanding of the situation of end users and their appliances. Secondly, by having software architectures and pervasive communication networks to process and convey the contextual information to where it is relevant and thirdly, by using analytics in the IoT environment to determine smart behavior.

Therefore, there is a need to deploy a large-scale, platform independent, wireless sensor network infrastructure that includes data management and processing, actuation and analytics. A wireless sensor network is a self organized network composed of a large number of sensor nodes that interact with their environment and communicate wirelessly. The target is to transfer the processed data to a remote base station.

The network components of a wireless sensor network include a Gateway, Relay Node, Leaf Node and a Sensor or an Actuator. A gateway is an interface between the application platform and the wireless nodes on the wireless sensor network. All information received from the wireless nodes is aggregated or manipulated by the gateway and forwarded to the application. This aggregation or manipulation can be the translation between network packet formats. A Relay Node is considered a full-function device (FFD), and it is used to extend network coverage area, route around obstacles and provide back-up routes in case of network congestion or device failure. A Leaf Node is considered as a reduced-function device (RFD). It is sometimes called an endpoint. It is designed to provide the physical interface between the wireless sensor network and the sensor or actuator that it is wired to. A Sensor or Actuator is the device used for interaction with the physical system that one wishes to monitor and/or control.

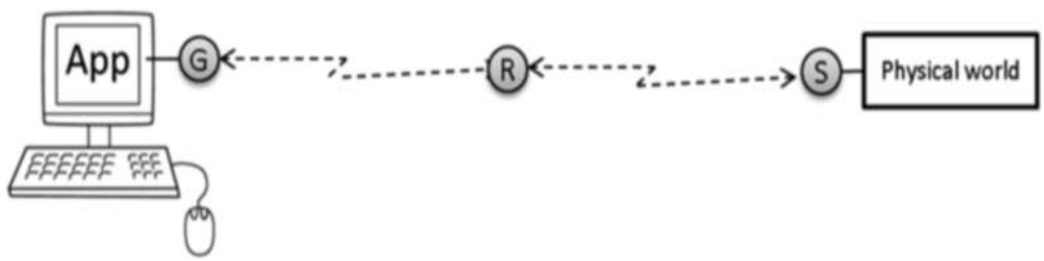

Figure 2:5: Wireless Sensor Network Components

(Springer, 2016)

The three main building blocks of a wireless sensor network are wireless sensor nodes, wireless sensor gateway nodes and the application software. The wireless sensor node is equipped with sensors, micro controller units, GPS module, power supply and the RF transceiver. The Wireless sensor gateway node consists of a microcontroller unit, a GPRS module, RF, memory card and power supply. The application software contains the system database to store received data and provide visualization information.

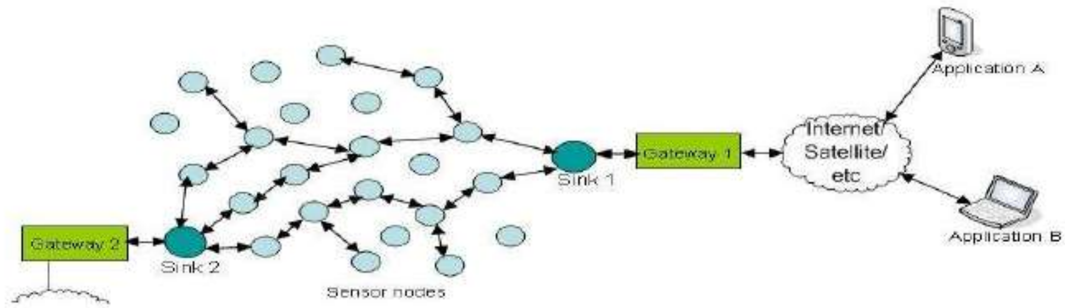

Figure 2:6: Wireless Sensor Network Platform

(Gubbi, Buyya, Marusic, & Palaniswami, 2013)

2.5.2 The Wireless Sensor Networks Platform

A wireless sensor network typically has to provide signal conditioning and data acquisition for different sensors, storage of data (including the configurations), processing capabilities, analysis of the processed data for alert generation, actuation, scheduling and execution of the measurement tasks. Management of node configuration e.g. changing the sampling rate and reprogramming of data processing algorithms, reception, transmission and forwarding of data packets and scheduling and execution of communication and networking tasks is also critical (Springer, 2016).

A Fully Functional Device node can be an embedded device or a more powerful computer and will be able to provide all the functionalities stated above. A Reduced Function Device, on the other hand, is typically an embedded device that provides only part of the functionalities stated above. Typical computer platforms run on the operating systems developed that are equipped with standard LAN communication. Such devices have high processing capabilities and allow for flexible configurations. However, they are usually big, too expensive and require external power. Therefore, the use of embedded devices is preferred for a wireless sensor network deployment. Embedded devices, which are smaller, allow for limited resources and are could be battery operated. They allow for constrained resources in terms of processor, radio and memory capabilities, hence are cheaper. In terms of processing for instance, the microcontroller unit is mostly programmed in C, which enables the development of a tight code that fits a limited memory size. Different wireless sensor operating software exist, among them, TinyOS and Contiki, which are the most popular (Springer, 2016).

Since a wireless sensor network may consist of tens, hundreds or thousands of devices, network topologies must be considered in its design. The Table 2.1 illustrates the different topology types in relation to the power usage, communication usage and if time synchronization is required.

Table 2:1 Network Topologies

(Springer, 2016)

Topology	Power usage	Communication range	Requires time synchronization
Star	Low	Short	No
Tree	Low	Long	Yes
Mesh	High	Long	No
Hybrid	Low (typically)	Long	(Depends on the configuration)

The data models for wireless sensor networks describe the interaction between the sensors and the application. Unlike the topology, which is a function of the network protocol, the data model is a function of the application. There are different models for monitoring applications, where the data flows primarily from the sensor node to the gateway, and for control applications, where the data also flows very frequently from the gateway to sensor nodes. The model may be based on periodic sampling, event driven or on the basis of store and forward.

Periodic sampling is ideal in a situation where certain conditions or processes need to be monitored constantly. The sampling period will depend on how fast the condition or process varies and what intrinsic characteristics need to be captured. The event driven model is for monitoring crucial variables immediately following a specific event or condition. The store and forward model is used in many applications, and enables data to captured and stored, or even processed by a sensor node before it is transmitted to the gateway. This enables the improvement of the overall network performance in both power consumption and bandwidth efficiency.

The routing techniques in a wireless sensor network describe how the data is forwarded to the gateway. A wireless sensor network will rely on its network's routing algorithms to discover routes and deliver packets from sources to destinations. Routing algorithms which are data centric, hierarchical and location based have been used for routing data in sensor networks. There have also been some that route the data based in network flow or Quality of Service awareness. The Table 2.2 presents the different routing algorithms and classes (Springer, 2016).

Table 2:2 Routing Algorithms and Classes

(Springer, 2016)

Data centric	SPIN
	Direct Diffusion
	Energy aware
	Reliable Energy Aware Routing (REAR)
	Rumor
	MCFA
	Link quality estimation based
	Gradient based
	Information driven
	Acquire
Hierarchical	LEACH
	EWC
	PEGASIS
	TEEN/APTEEN
	Energy-aware cluster based
	Self-organized
	Minimum energy communication network
	Small minimum energy communication network
Location based	Geographic Adaptive Fidelity
	Energy Aware Greedy Routing (EAGR)
	Geographic and Energy Aware
QoS aware	SPEED
	MMSPEED
	Sequential Assignment
	Real-Time Power-Aware
	DCEERP
	Energy Efficient with Delay Guaranties (EEDG)

2.5.3 Wireless Sensor Networks in Distribution Systems

Wireless Sensor Networks are preferred due to a number of reasons. They are energy efficient as they ensure low energy consumption and are usually battery powered for mobility. They ensure wireless communication; hence, they offer easy installation as no cabling is required, thus making them less costly. Wireless Sensor Networks also ensure distributed data processing in each node. This means that data filtering and aggregation is not only done by the main node, which is relieved with the processing requirements (Kolenc & Zajc, 2007).

Once pollution is detected in a water distribution network, some pipelines must be isolated from the unpolluted area, and anti pollutants need to be dosed into these pipelines. This is done to isolate the contamination extent. Therefore, sensors that measure flow rate, oxygen level, pH level, and other parameters; and actuators such as pumps, valves, have to be deployed in the water network for active real-time monitoring and dynamic control (Du, 2016).

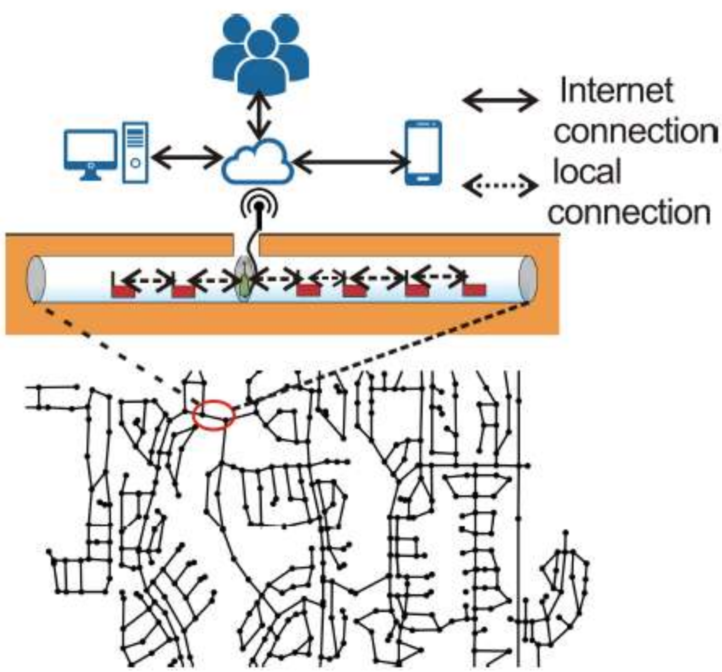

Figure 2:7: Monitoring Pipelines Using Wireless Sensor Networks

(Du, 2016)

The Figure 2.7 illustrates how water can be monitored in a distribution network using Wireless Sensor Networks. The water distribution network consists of pipelines and junctions.

The sensor nodes are placed along the pipelines to monitor the parameters such as flow rate, pipeline vibrations, oxygen level, and pH level of the water, among others. The measurements are sent to the sink nodes which are located at the junctions, and further transmitted via the network to the shared database for access (Du, 2016). In Singapore's WaterWiSe system, some sensor nodes are mounted on poles, whereas some other sensor nodes are inserted into the pipes. The placement of the sensor nodes is determined according to proximity and in some cases by the use of algorithms.

Mobile sensor nodes can also be deployed in distribution networks. The sensors will be covered in a water proof casing to protect the circuit. Figure 2.8 illustrates a mobile sensor node.

Figure 2:8: Mobile Sensor Nodes

(Du, 2016)

2.5.4 IoT and Data Storage

With the increased need to process higher volumes, velocities and varieties of data at a rapid rate, there have been changes in storage needs. In order to enable this scenario, NoSQL databases that enable storing unstructured and heterogeneous data at scale have gained in popularity (Microsoft Azure).

NoSQL is a category of databases distinctly different from the traditional SQL databases. NoSQL is often used to refer to data management systems that are "Not SQL" or an approach to data management that includes "Not only SQL". There are a number of technologies in the NoSQL category, including document databases, key value stores, column family stores, and graph databases, which are popular with gaming, social, and IoT application (Microsoft Azure).

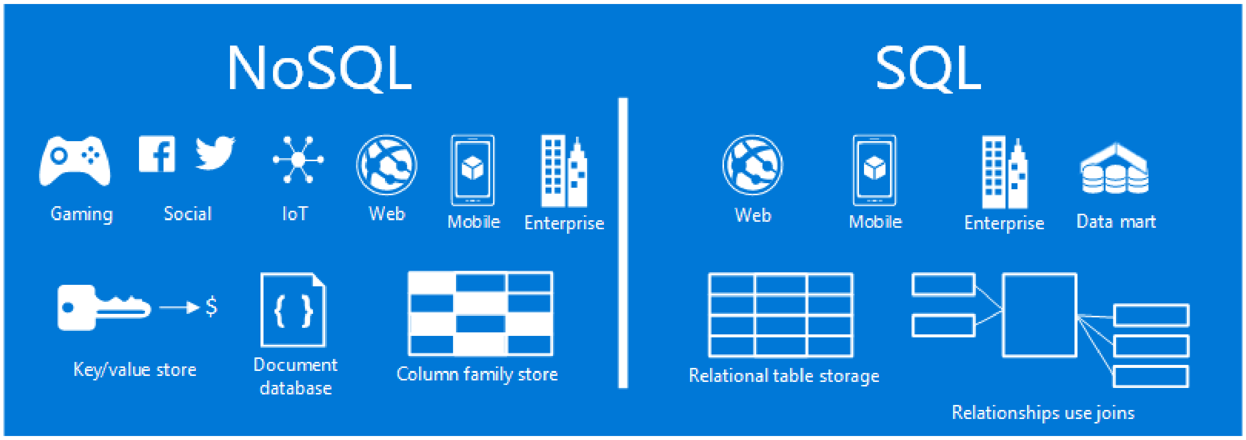

Figure 2:9: NoSQL and SQL Application Comparison

(Microsoft Azure)

2.5.5 IoT and Smart Water Architecture

Architecture of smart water grid based on Internet of Things was proposed, mainly including hierarchy framework, technical system and function framework. The hierarchy framework consists of real-time active sensing layer, interconnecting layer of water information interconnection layer, and smart decision layer. The technical system mainly contains the techniques, for example the multi-carrier network composing and real-time transporting, intelligent storage of mass data and the seamless docking of multi-modules of the system, the coupled simulation of multi-dimensional water model, the evaluation and diagnosis of water security risk ,the fast and dynamic forecasting model building of water pollution incidents, the multi-objective optimization and control technology of water quality and quantity of cascade reservoirs, the dynamic early warning technology in the unexpected water pollution incidents, the emergency disposal technology of the multi-scale, multi-type and multi-objective unexpected water pollution incidents. The functions framework includes the modules, e.g., intelligent sensing, intelligent simulation, intelligent diagnosis, intelligent early warning, and intelligent regulation, intelligent disposal, and intelligent control (Ye, Liang, Zhao, & Jiang, 2016).

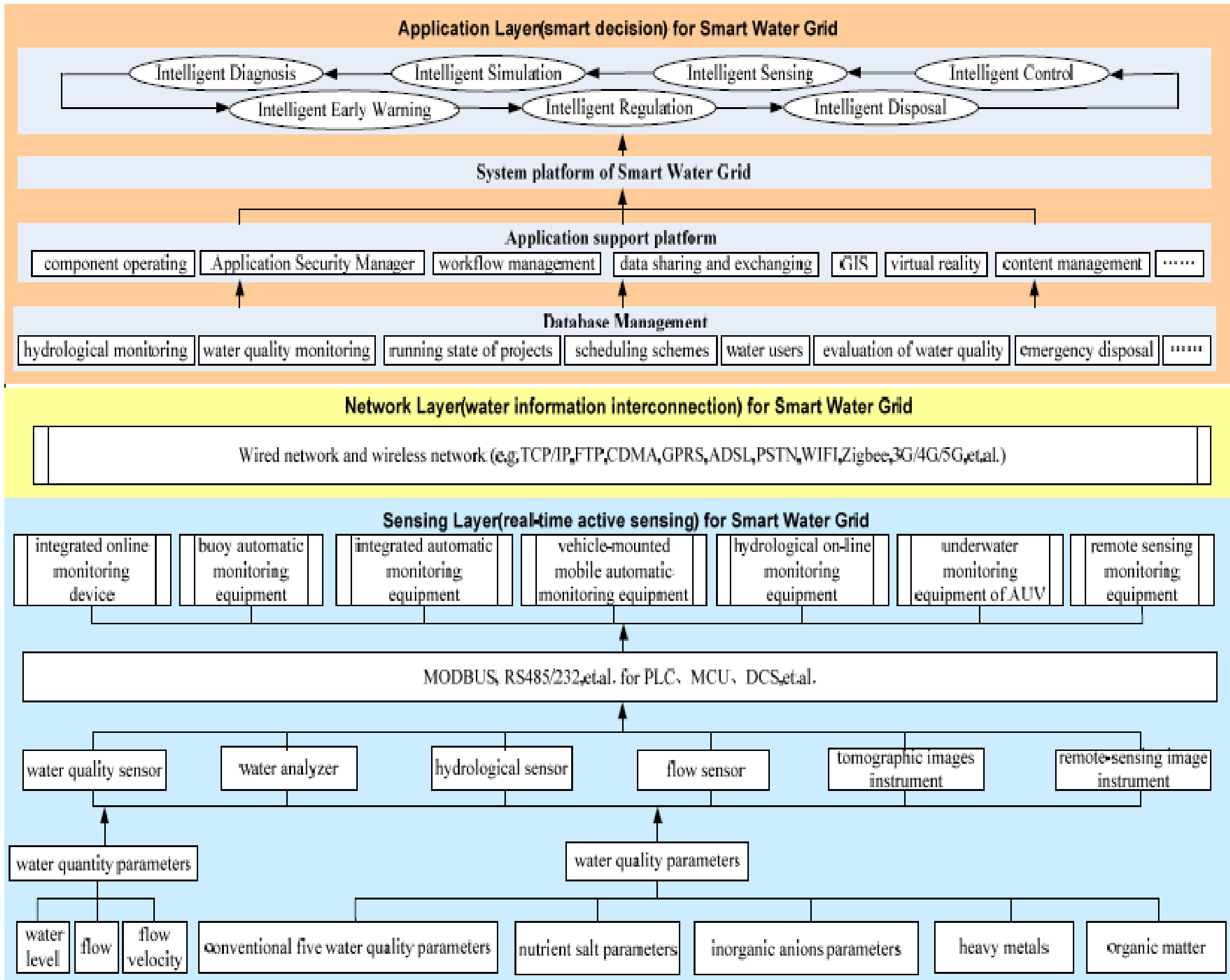

Figure 2:10: The Hierarchy Framework of the Smart Water Grid for Water Security

(Ye, Liang, Zhao, & Jiang, 2016)

The Figure 2.10 shows the hierarchy framework of the smart water grid for water security (Ye, Liang, Zhao, & Jiang, 2016). It focuses on water quantity and water quality, which represents the scope which is covered in this research.

2.5.6 Visualization of Data

According to Dan Potter's presentation at the Internet of Things (Datawatch, 2015) the 6 Requirements for IoT Visualization are: data discovery, streaming data visualization, time series data, predictive & advanced analytics, complex file formats, real-time geospatial & location.

Data discovery is not just a dashboard, it makes it easier for users to author, customize and share data. Anomalies and outliers will also be detected faster. Streaming data visualization will enable faster speeds and faster insights, hence alerts to be done faster as the system is connected directly to data in motion. Time series data is continuous and enables complete situational awareness. Predictive and advanced analytics will be made possible due to the availability of history data patterns. Due to the nature of the complex file formats, there is a need to transform, enrich and prepare data. Real time plotting will enable map visualizations.

2.6 Conceptual Framework

Sensing devices that collect and transmit data about the water system on a real-time basis form the foundation of a smart water grid. Currently, the water delivery system is monitored by manually reading flow and pressure meters. The physical monthly visit by meter readers to consumer end meters is a common occurrence in formalized water utilities. In addition to that water quality monitoring for contamination is usually monitored by collecting water samples, which are analyzed in a laboratory environment. The utilities have designated sampling points for which manual samples are taken and turbidity, chlorine and pH tests, among others, are done in the laboratory.

In the proposed smart water system, these parameters are collected, stored, and transmitted to a cloud database in real-time. Sensors consist of the flow sensor for flow volumes and water quality sensors for contaminants. The smart sensors communicate via Wireless Sensor Networks.

The data collected is then stored in an online database that links to a web application that provides a visualization of the data, depending on the regulator’s indicators. It would be important to note that the consumers of the information will also benefit from the real-time monitoring of their water consumption, and will also be aware of the water quality.

The Figure 2.11 represents how the smart water network works. The data derived from the sensors is the water quantity through the water flow sensor, and water quality from the turbidity sensor.

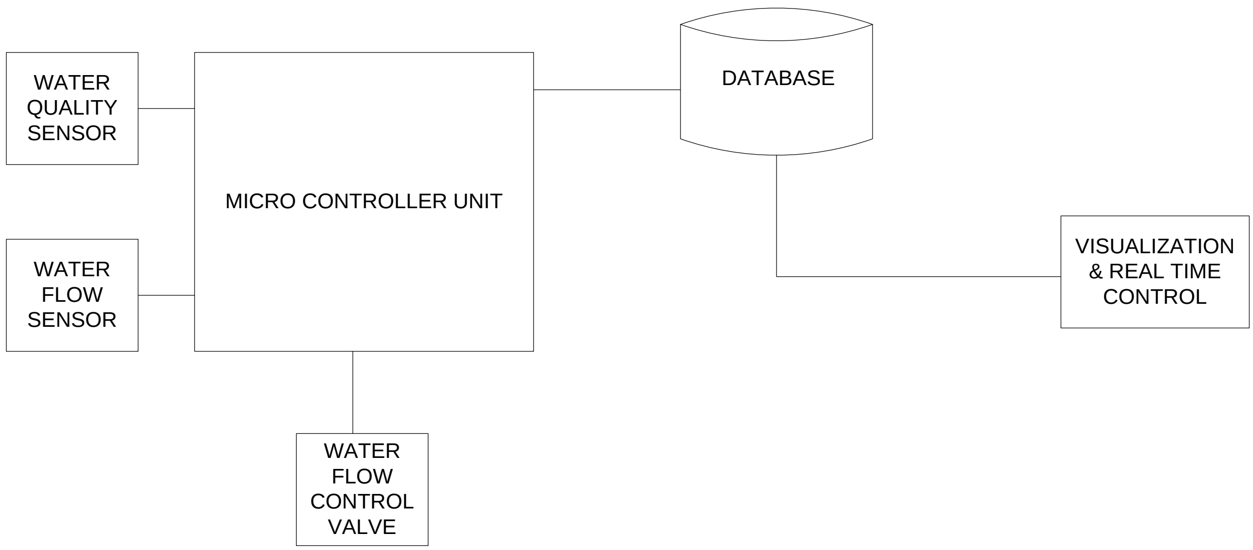

Figure 2:11: Conceptual Framework for the Proposed Smart Water Grid Model

The smart water infrastructure demonstrates the principles of Internet of Things at work. The prototype demonstrates the data capture by the devices and the transfer process to the database. The data is captured with precision as it is the actual data.

The prototype demonstrates the communication between the devices and the adjustment to the situation depending on the events, for instance, if the water is contaminated, the water quality sensor detects that, and triggers a message to the flow control valve to automatically stop the transmission of the contaminated water by shutting the valve.

Chapter 3: Research Methodology

3.1 Introduction

This chapter discusses the research methodology that was used to achieve the research objectives from gathering analyzing, reporting and system development. The type and source of data, the population and the sampling methods and techniques used to select the sample size are discussed. The method by which the data was collected is also explained. Primary and Secondary sources of data supported the research process. This chapter also describe in details, the system development methodology. The chapter also discusses how the smart water prototype was developed and evaluated.

3.2 Research Design

A smart water prototype that demonstrated quality data capture, using wireless sensor networks was developed. The prototype demonstrates water quality data, and its relation to the water volumes that are transmitted through the infrastructure. The output relates the data captured to the performance indicators monitored by the water services regulator. The prototype was executed using classic Software Development Life Cycle, SDLC. The waterfall methodology was used for the development of the prototype as the variables of concern, which are, water volumes and quality were already known, and each stage had to be terminal.

The research design was descriptive as it helped to provide the answers to the question of how water utilities can be monitored using wireless sensor networks. Descriptive research is used to obtain information concerning the status of the phenomena and to describe what exists with respect to variables or conditions in a situation (Sekaran, 2004).

The research was divided into two distinct stages. The first stage of the research involved intelligence gathering and data collection. The second stage involved the development of the prototype. Primary data was collected through first hand investigation, and from interviews. The secondary data was obtained from organizational records such as the Impact Report and the review of the Non Revenue Water quarterly reports.

3.3 Population and sampling

3.3.1 Research Site

The research was conducted from Strathmore University and at the participant's site. The participant in this context refers to the Water Services Regulatory Board, where the information gathering was done.

3.3.2 Population

The population refers to the entire group of people, events, or things of interest that the researcher wishes to investigate (Sekaran, 2004). The population that was used was of the personnel from the Directorate of Technical Services within the Water Services Regulatory Board which currently stands at six.

3.3.3 Sampling Design

A sample is a subset of the population. It comprises some members selected from it (Sekaran, 2004).The sample selected was the personnel who are involved directly in the data collection process from the Directorate of Technical Services. This is selection was based on the person's knowledge of the subject matter. The staff directly involved in the data collection exercise are three in number.

3.4 Data Collection Methods

Preliminary data collection was done through:

i. Carrying out in-depth interviews with three personnel from the Directorate of Technical Services within the Water Services Regulatory Board. The interviews were important as they determined the emotions, feelings, and opinions regarding the data collection challenges and the process.

ii. A semi-structured questionnaire (Appendix B) was used as an interview guide for the researcher. Open ended questions to guide the interview towards the satisfaction of the research objectives were prepared in advance. However, additional questions which enriched the non functional user requirements were also encouraged, depending on the respondents' answer.

iii. Literature review and analysis of business documents and reports of the current data collection process and challenges was done. The Impact report was reviewed, as well as operational data sample reports, such as the NRW Quarterly Report (Appendix C).

iv. Observation and participation during the scheduled data collection and validation workshops.

3.5 Review of Interview Responses

This review was important as it enabled to capture the information acquired from the respondents. Qualitative method was used to analyze the information. The results for the three respondents were summarized hence identifying the challenges that the regulator experiences when capturing monitoring data for Kenyan water utilities, as required in the first research objective.

3.6 Research Quality

The methodology chosen ensured data quality throughout the research process. The sample size for interviewees was selected from all the subject matter holders, thus enriching the quality. The interview guide contained specific and direct questions to ensure the research objectives are met. The interview guide was structured with reference to the data quality elements, hence enriching the quality of the interview. The software development methodology chosen also ensured the systematic development of the prototype- from analysis through to implementation and testing. Evaluation of the prototype was also done against the challenges identified by the end users.

3.7 System Development and Methodology

3.7.1 Overview

This research focused on capturing utilities operational data at source. The quality data assessment was done with reference to Li Cai's Data Quality Framework (Li Cai, 2015). The capturing of data at source and in real time enables availability, usability, reliability and relevance of data. Using visualization enhanced the presentation quality aspect.

The visualization aspect was aimed to demonstrate the use of data to make timely decisions. The system methodology that was used was the classic Software Development Life Cycle, SDLC, which consists of six distinct phases: Requirements Gathering and Analysis, Design, Testing, Deployment and Maintenance.

3.7.2 Requirements Determination, Analysis and Design

The system requirements and conditions were identified from two main sources. Firstly, from the feedback obtained from the interviews, and secondly from literature review of the publications which illustrate the current data collection process followed by the regulator. Process observation was also done from the data collection and validation workshops, which had been scheduled at the time of the research.

The requirements were analysed and mapped into design using UML. UML was used to define use cases, interaction diagrams, and data flow diagrams. These diagrams were then translated into to code, which was done in the Arduino microprocessor, and the web application.

3.7.3 Testing of the Prototype

The prototype was then evaluated and tested for validity using test cases. The interaction and communication between the sensors was verified. The prototype's adequacy in ensuring accurate data capture at the source is important. The visualization charts and further information derived from the data capture process were also verified.

3.8 Ethical Considerations

Unethical actions were avoided during the research (Sekaran, 2004). Such actions included; putting pressure on individuals to participate in the interviews through coercion, or applying social pressure; giving menial tasks and asking demeaning questions that diminish interviewees' self- respect; deceiving subjects by deliberately misleading them regarding the true purpose of the research; exposing participants to physical or mental stress; not allowing subjects to withdraw from the research when they want to; using the research results to disadvantage the participants, or for purposes not to their liking; not explaining the procedures to be followed; not debriefing participants fully and accurately after the interview; and not preserving the privacy and confidentiality of the information given by the participants.

Chapter 4: System Design and Architecture

4.1 Introduction

This chapter describes the system analysis and design process. The system needs analysis was done to identify the functional and non-functional requirements. Identification of the challenges experienced by the water regulator when capturing monitoring data was done by the use of guided interviews. The results were summarized and presented. The initial prototype architecture has been described. The system design is illustrated using UML- with the use of a use case diagram, a sequence diagram and data flow diagrams (Context Diagram and Level 0 DFD). Security considerations for the prototype are also discussed.

4.2 Requirements Analysis

The purpose of system requirements analysis was to specify the system requirements that enabled the development and the implementation of the system. The requirements analysis followed a two phase approach. Firstly, by analyzing the current system. The investigation of the current system emphasized on the process, effectiveness, quality of data and the suggested areas of improvements by the users. Secondly, by analyzing the current process of data collection. Literature review of the water regulator’s publications and process observation was done from data collection and validation workshops.

4.2.1 Evaluation of Current Process

The water regulator’s annual performance report, Impact publication ranks utilities performance describes the data collection and validation process. Process observation was done by participating in four of the annual regional data validation workshops. Interaction with the data validation teams during the validation sessions was done during the workshops.

As per the research objective of determining the challenges that the regulator has when capturing monitoring data for Kenyan water utilities, an interview guide (Appendix B) was developed and the results are summarized in the following sections. The population as described in Chapter Three focused on interviewees within the Directorate of Technical Services within the Water Services Regulatory Board. All the identified three personnel were interviewed. The interviewees also highlighted the areas requiring improvement.

4.2.1.1 Assessment of Challenges in Data Availability

This assessment was done by capturing the interviewees experience on the accessibility and timeliness of the data that is submitted by the utilities. All the three respondents pointed out that the system for data capture system was always available online, and hence accessible by all water utilities and the water services regulator. The data is updated annually. One interviewee suggested increased frequency in reporting, to quarterly or monthly schedules. However, the process experienced some delays. One of the interviewees cited the lack of proper tools for monitoring facilitated the delay. The tool here referred to were lack of proper metering and lack of proper billing systems. The delay herein was in two ways. The first delay was between the time utilities receive the data collection request and the time they spent preparing the data for input to the system. The second delay experienced was during analysis of the data; from the time the data is already fed into the system and processing to meet the release requirements. The data is already publicly available through the performance monitoring report.

4.2.1.2 Assessment of Challenges in Data Usability

This assessment was done by determining the credibility of the data collected. The data in question is usually submitted by the water utilities based on the information obtained from their internal systems. It was identified that the utilities' systems are regularly audited. One respondent specifically pointed out that some utilities lack proper billing systems, hence affecting the credibility of data submitted. Some utilities however had gone the step further of updating their systems to improve on their data. Most of the data provided, however, existed within the range of known and acceptable values as this validation feature is inbuilt in the current data collection tool.

4.2.1.3 Assessment of Challenges in Data Reliability

This assessment was done using the accuracy, consistency, integrity and completeness elements. According to the respondents, the data provided was reliable for some utilities, given the basis of the consistency. However for some, there was a lot of inconsistency, and hence did not reflect the true state of the reality in the ground. In such scenarios, other reports (for instance WSP inspection reports, licensee reports and NRW quarterly reports) within the regulators' realm were corroborated to determine the accuracy of the data. There were no stated cases of data ambiguity.

After processing data, the data that is published, at times differed with the data submitted due to inconsistency of some data submitted by the utilities. The published data is the one that is concurred with during the validation sessions, after the rigorous and lengthy process. The system data is usually exported for further analysis outside the data collection tool. For some utilities though, the data remains constant and verifiable.

The data entry forms on the online system are easy to navigate through. With the current data collection tool, data structure and integrity is maintained. The completeness of that data collected is affected by the various multi components, as the indicators are related and clustered. Therefore this means that a deficiency in one component may affect another both in use and integrity.

There was a suggestion to automate the technical data to improve the credibility of the data. It was noted that some of the staff that were involved in the data entry exercise could not make sense of the data, hence the risk in error introduction. The integrity of the captured data in the system relied heavily on the integrity of the personality doing the data entry.

4.2.1.4 Assessment of Challenges in Data Relevance

This assessment was done by reviewing the fitness for use of data. The datasets that are retrieved from the system are within the regulator's needs. This is due to the fact that the data requirements for the current data collection tool were developed by the regulator; therefore, they aligned with the regulator's needs. In order to enrich the reports, uploading of reports such as water quality reports and financial statements to be done.

4.2.1.5 Assessment of Challenges in Data Presentation Quality

This assessment was done by reviewing the readability of the data. All the respondents found the data content and format clear and understandable. The respondents stated that the data provided the required needs and were in line with the performance assessment indicators. However, one respondent emphasized the fact that the report generation based on the current data collection tool needs improvements. According to the respondent, the current tool was rigid in report generation, as the predefined reports were not adequate, despite the fact that the data for report generation was already within the system. Generally, the current online tool experiences slow speeds, especially during peak times, therefore loading of data sets was slow. A recommendation to make the system faster in loading the variables was made.

4.2.2 Summary of Interview Findings

Table 4.1 summarizes these findings of the interviews that were conducted to identify the challenges faced by the regulator during the data collection and validation process. It should be noted that the feedback was later be used to validate the prototype's effectiveness.

Table 4:1 Identification of Challenges: Availability, Usability

Data Quality Element	**Identified Challenges**	**Observation and Recommendations**
Data Availability	Delay in correlating data by WSPs	General delay in deliverables leads to delay in the delivery of the final publication.
	Delay in analysis of data	
	Delay in formatting data for publication	
Data Usability	Lack of proper monitoring tools by some WSPs	A tool championed by the regulator will ensure consistency among the WSPs.

Table 4:2 Identification of Challenges: Reliability, Relevance, Presentation Quality

Data Reliability	Inconsistent data given by some utilities	Lengthy validation process as corroboration with other reports is required. The data capture and analysis time to be shortened.
	Instances of different published data compared to data in the online validation tool due to lack of timeliness in updating the data	The issue of timeliness of data needs to be addressed.
	'Bad data' will cascade to the related indicators as the indicators are related and clustered for efficient monitoring	Consistent and accurate data should be captured.
Data Relevance	No challenges identified, however a major area of improvement would be the provision of reports	Rich and real time reports are critical.
Data Presentation Quality	Exporting of data is done for further analysis. This may introduce human errors and is prone to tampering. It should be noted that the data still undergoes rigorous cross checks	Tamper proof and reliable process of data capture should be adopted.
	Rigidity in the quality of reports generated by the current tool which also has a slow loading process of the datasets	Rich and real time reports are critical.

4.3 General System Requirements

The general system requirements- both functional and non-functional were determined using the FURPS+ model. This model describes the Functional, Usability, Reliability, Performance, Supportability and Other sub factors such as implementation, interface, operations, packaging and legal requirements.

4.3.1 Major System Capabilities

These major system capabilities are the functional requirements and were the following:

i. The application must enable real-time transmission of data for water quality and water flow and the information which should be available online.
ii. The application must be able to issue alerts in the event the data captured has exceeded the acceptable threshold.
iii. The system should be self adaptive for instance, in the event contaminants are detected, the valve should immediately stop transmission of the water.
iv. The application must be accessible with any device connected to the internet.
v. The application should enable the visualization of the captured parameters.

4.3.2 Major System Conditions

The major system conditions specify the systems' constraints and were the following:

i. The system must use the wireless sensor network architecture
ii. The system must be reliable, that is, ensure 100% availability. The selection of a cloud database informed this decision.

4.3.3 System Interfaces

The system will in the future, interface with the regulator's online data collection tool, WARIS. The system will also interface with the utilities internal systems, such as the billing systems. This however is not be included in the prototype. The system is easy to interact with and follows the basic HCI principles.

4.3.4 System User Characteristics

The system will be accessed by the regulator for the purposes of report generation. The system can also provide data for the water service providers, in terms of the water quality alerts, which enables them to take immediate remedial action. The system was made using an open source platform, and hence will ensure supportability, and ease of configuration.

4.3.5 Initial Capacity Requirements

As the system is a real-time monitoring tool, the memory requirements are scalable. The prototype is currently hosted on the Google cloud, and the requirements will be determined as the usage increases. The initial capacity requirements are summarized in the Table 4.3. Consequent capacity requirements will be estimated using the current data amounts, planned number of users, and estimated number of transactions. The usage statistics will be generated from the hosting platform.

Table 4:3 Initial System Capacity Requirements

Real-time Database	Simultaneous connections	100
	GB stored	1 GB
	GB downloaded	10 GB/month
Storage	GB stored	5 GB
	GB downloaded	1 GB/day
	Upload operations	20,000/day
	Download operations	50,000/day
Functions	Invocations	125,000/month
	GB-seconds	40,000/month
	CPU-seconds	40,000/month
Hosting	GB stored	1 GB
	GB transferred	10 GB/month

4.4 System Architecture

As an online system which monitors water flow and quality and allows for adaption of the behavior based on the values captured by the sensors, the prototype was based on the principles of a wireless sensor network. The hardware requirements are sensors (water quality and water flow), a programmable microcontroller unit, a water flow valve and data storage mechanism, which is based on the cloud. The sensor used for water quality is the turbidity sensor and for water volumes, the Hall Effect flow sensor. The microcontroller is Arduino UNO. The software requirements for the microcontroller run on Arduino 1.8.1. Once the data is captured by the sensors, it is transferred to the cloud NoSQL database. The data is then sent to the database using the JSON format. For efficient storage and flexibility, the Google Firebase was used. Google Firebase provides storage and hosting of real-time databases. For network connectivity, GSM was used.

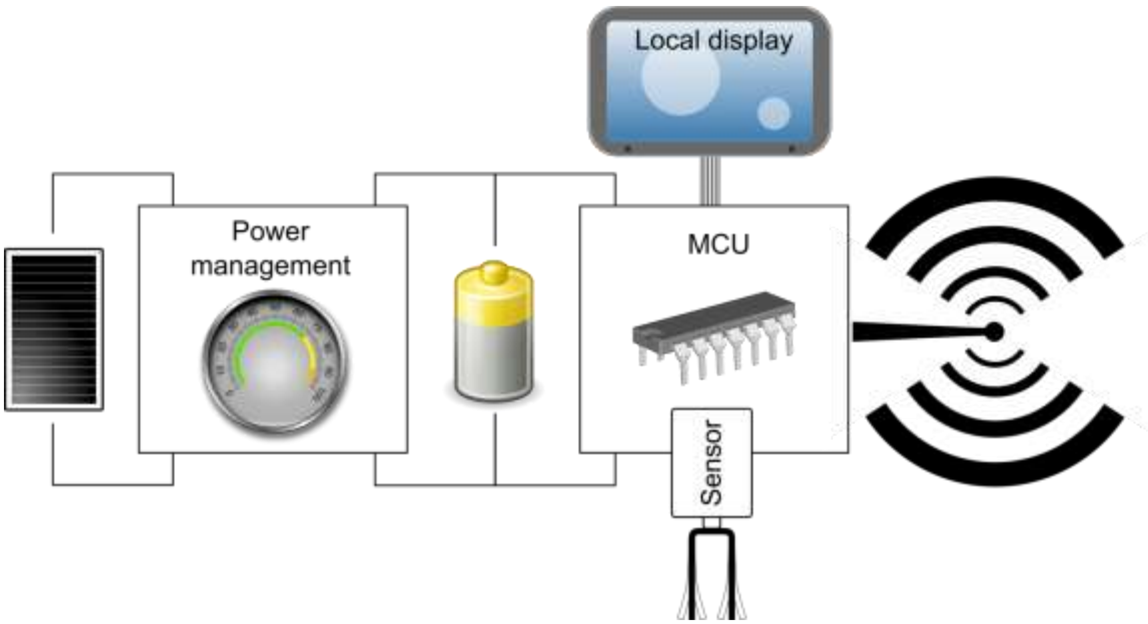

Figure 4:1: Wireless Sensor Network Architecture

4.5 System Design

The system design describes the system from architecture level including the services, hardware mapping, data management and access control. The purpose of the prototype is to demonstrate the use of wireless sensor networks for monitoring water data. The motivation behind the development is the challenges that are experienced during data collection by the water services regulator, as described in the problem statement. The system composition has been illustrated in the UML diagrams.

4.5.1 Use Case Diagram

4.5.1.1 Use Case Requirement in Context

The sensors record values, for instance, the water flow sensor records the flow volumes, while the water quality sensor records the turbidity of the water. In the case of contamination, the water quality sensor will detect the threshold pre calibrated has been violated and communication will be done to the valve to stop flow. An alert will be issued on the application. The water flow sensor will record a value of 0 as the flow of water will be stopped. The recorded data is stored in the cloud database in real-time.

4.5.1.2 Diagrammatic representation

A use case diagram uses actors and use cases to represent what the system will do. In the prototype, the actors are represented by the sensors and the subsystems, while the use cases will represent the behavior of the system.

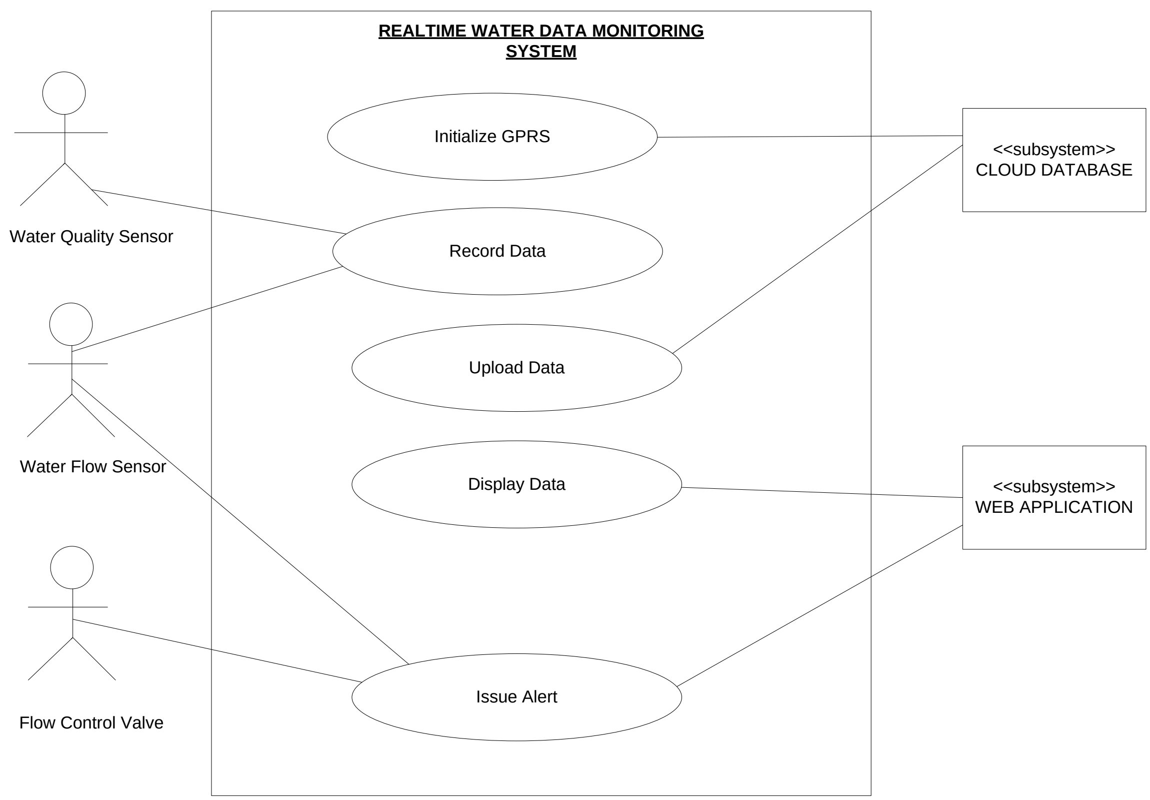

Figure 4:2: Use Case Diagram

4.5.1.3 Description of the use case scenarios

The Table 4.4 presents the Initialize GPRS use case description. A brief description of the case is given. The major inputs and outputs are stated. The preconditions, describing what must be fulfilled, and the post conditions indicating what will be achieved is captured. The additional use case descriptions for Record Data, Upload Data, Display Data and Issue Alert are in Appendix D.

Table 4:4 Initialize GPRS Use Case Description

Use Case Name	Initialize GPRS	Identifier: 1
Brief Description	Ensures that there is internet connectivity	
Type	Internal	
Major Inputs	**Source** Power	**Major Outputs** Internet connectivity is established
Pre Conditions	There has to be a source of power	
Post Conditions	Internet connectivity is established	
Flow of Events	1. The PCB Board is powered 2. GSM connection is established 3. GPRS is initialized	

4.5.2 Sequence Diagram

A sequence diagram is an interaction diagram that shows how objects interact via messages. In the Figure 4.3, the objects represent the sensors, microcontroller the subsystems which illustrates the sequence of actions performed by the prototype.

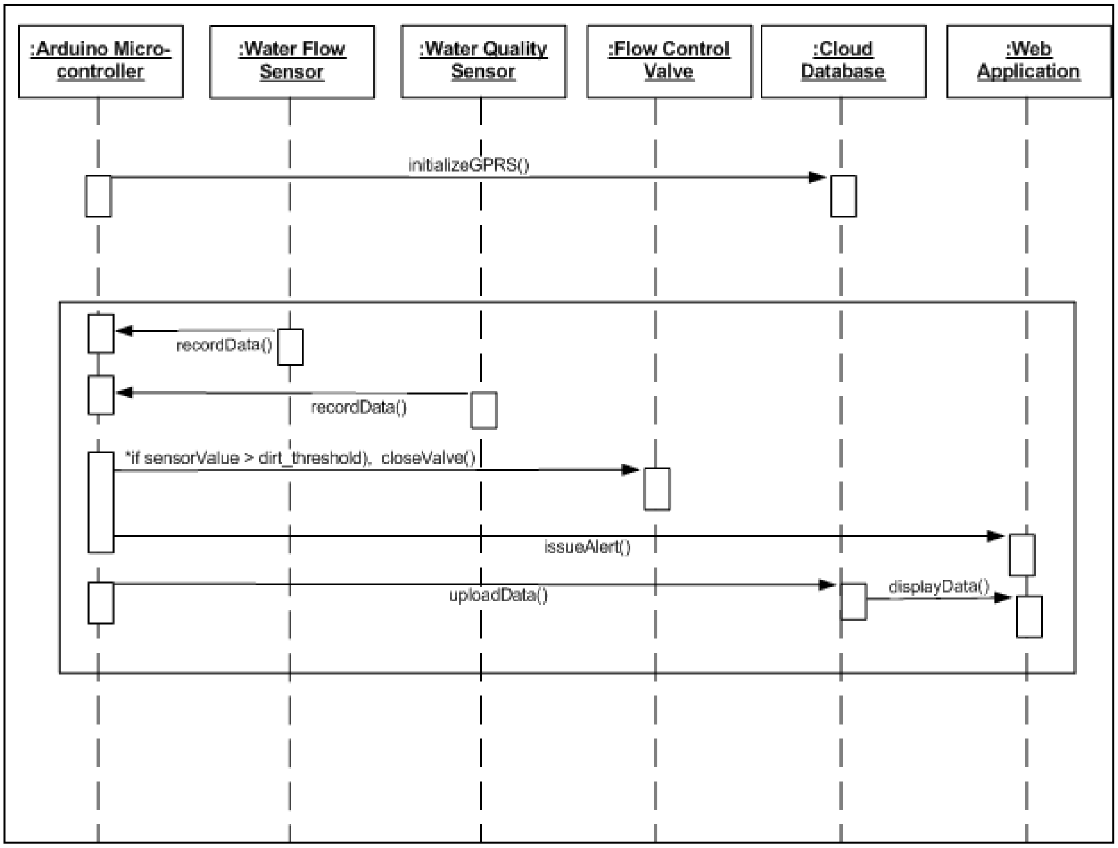

Figure 4:3: Sequence Diagram

4.5.3 Data Flow Diagrams (DFDs)

A data flow diagram shows how the data flows in the system, showing both the inputs and the outputs, The Figure 4.4 illustrates the Context Diagram for the prototype, while Figure 4.5 shows the Level 0 DFD. The data contained in the dataflow diagrams is limited to the scope of this research.

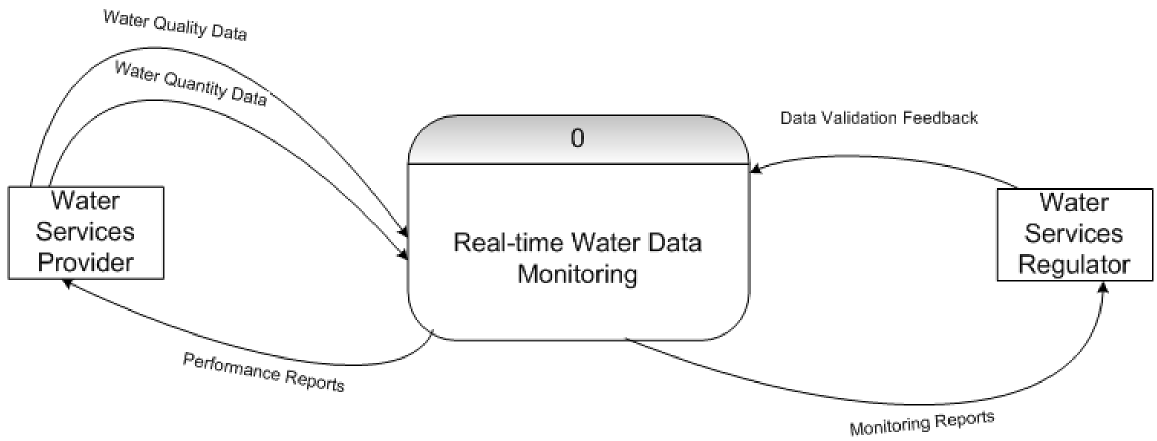

Figure 4:4: Context Diagram

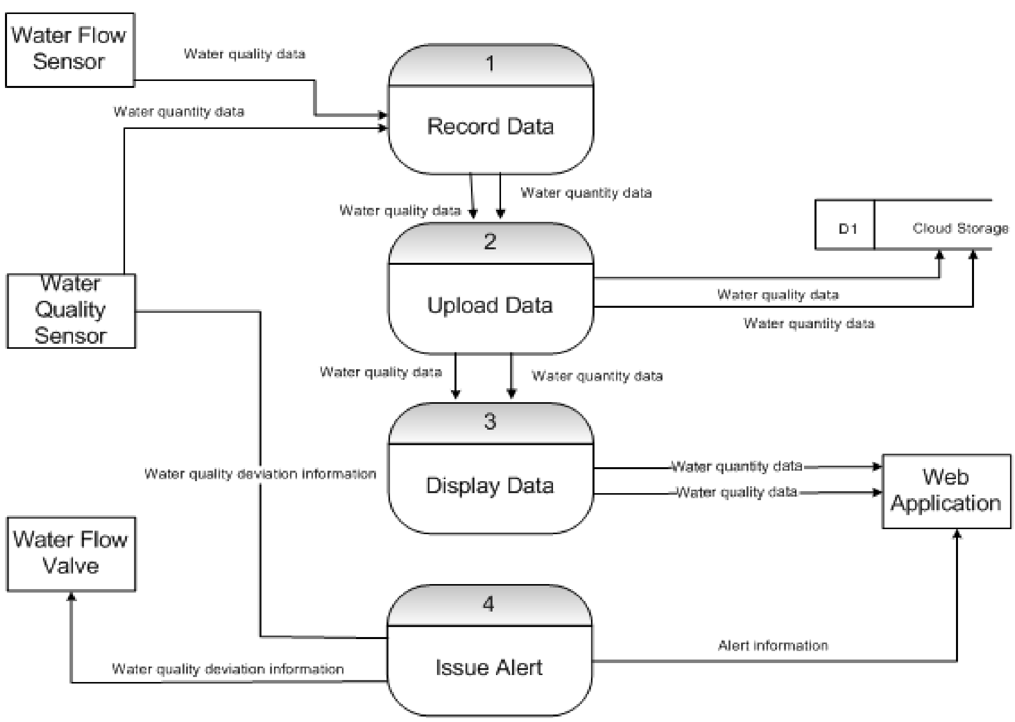

Figure 4:5: Level 0 Data Flow Diagram

4.5.4 Persistent Data Management

The data that is stored in the system is the water volumes recorded by the water flow sensor, the water quality data that is recorded by the turbidity sensor and the various alerts depending on the contaminants detected by the sensors. A No-SQL database was used, as it is able to handle large amounts of data, hence appropriate for a constantly evolving system. The data is accessible online from any device and is recorded real-time with a refresh rate of 10 seconds.

4.5.5 Access Control and Security

In terms of security, the system has been viewed in three main layers. The hardware layer, which consists of the sensors, the microcontroller and its components, and the flow valve; the database layer, which stores the recorded values; and the front end application which is a web application. Security has been ensured at all the three levels. For this prototype, at the hardware level, for this prototype, physical protection of the devices will be ensured. At the database level, the data is stored in secured cloud server, powered by Google Inc. At the web application level, the data captured is read only and the site is hosted at a secured site (HTTPS).

4.6 Summary

The system requirements were obtained from the primary and secondary data collected from the interviews, observation and literature review. The designed diagrammatic representation depicts the actual system requirements, which were carried over to implementation.

Chapter 5: System Implementation and Testing

5.1 Introduction

This chapter focuses on the system implementation and testing. The implementation was done in three main phases- the implementation of the sensors and microcontroller, the implementation of the database, and the implementation of the web application. Testing was also done in the three phases.

5.2 System Implementation

5.2.1 Sensor Integration Implementation

The hardware implementation of the sensors was done by assembling the turbidity sensor, the flow sensor, the flow control valve and the microcontroller. The Arduino UNO microcontroller was used. The set up on the PCB is illustrated in the Figure 5.1 and the description of the components on Table 5.1.

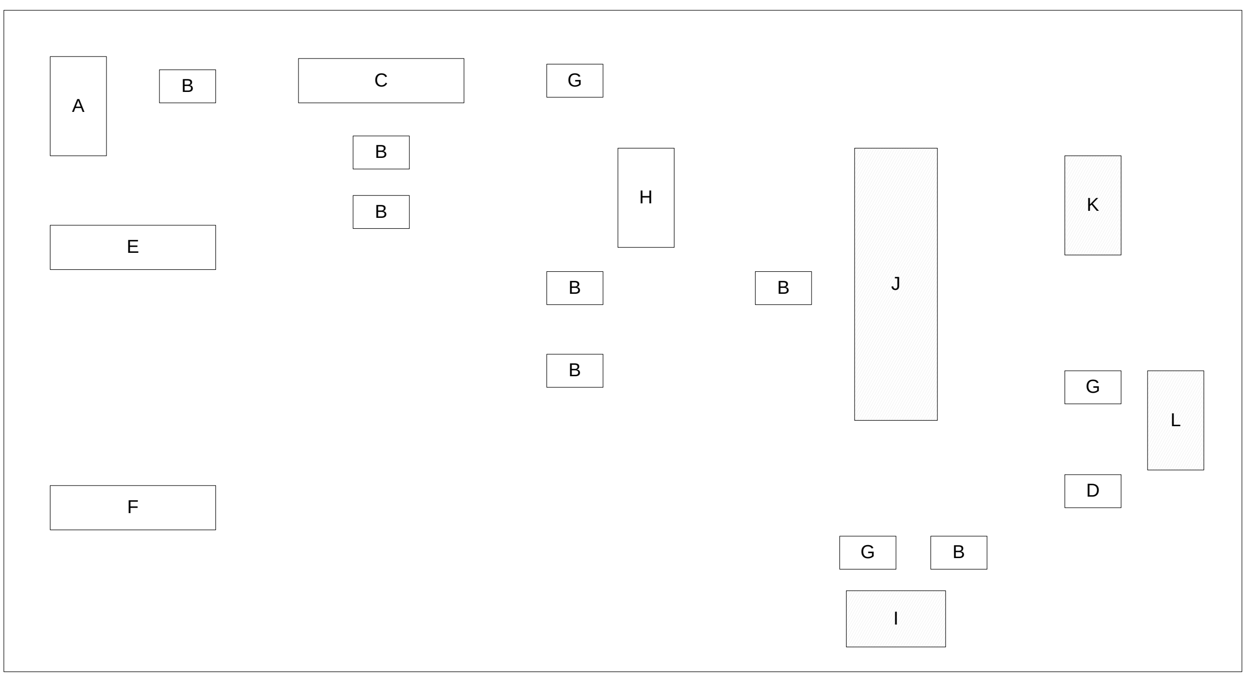

Figure 5:1: Components Layout on the Board

Table 5:1 Components Descriptions

COMPONENT	DESCRIPTION
A	12V Power Terminal
B	Ceramic Capacitor
C	Voltage Regulator
D	Low Dropout 5V Regulator
E	GSM_IN
F	GSM_OUT
G	Resistor
H	GSM Module
I	Water Flow Sensor
J	Arduino Microcontroller
K	Water Quality (Turbidity) Sensor
L	Flow Control Valve

The actual layout of the hardware components is illustrated in Figure 5.2.

Figure 5:2: Illustration of Assembled Hardware Components

The setup represents the assembled hardware components connected via a PCB. The valve is at the extreme end, while at the other end, there is a bucket which when filled with water builds up the pressure to allow water flow detection by the water flow sensor.

The Arduino microcontroller was programmed with the functions to enable communication between the sensors and the valve, and send the values to the storage database using GSM in JSON format. The microcontroller was initialized to connect to GSM. A delay of thirty seconds is allowed for the GSM to boot and a further two seconds to allow the GSM to connect to GPRS. Interrupts were also configured to allow for initializing and starting the flow sensor interrupts. The flow measured in milliliters per minute. Calibration of the turbidity sensor was set to 800 mg/l after determining the level recorded by clear water to be ranging between 650 mg/l-700 mg/l. The data is then uploaded to the cloud database in JSON format.

5.2.2 No SQL Database Implementation

The NoSQL Database was chosen due to its scalability and appropriateness for real-time databases. The Google Firebase platform enables storage of real-time data using the NoSQL format. Set up was done by designing the JSON API data format for the data that will be captured in the system, which is illustrated as in Figure 5.3.

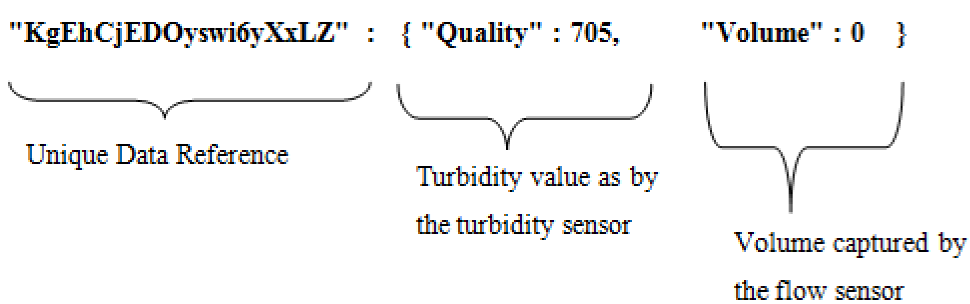

Figure 5:3: JSON Structure as Displayed in the Database

The database views are as shown in Figures 5.4 and 5.5.

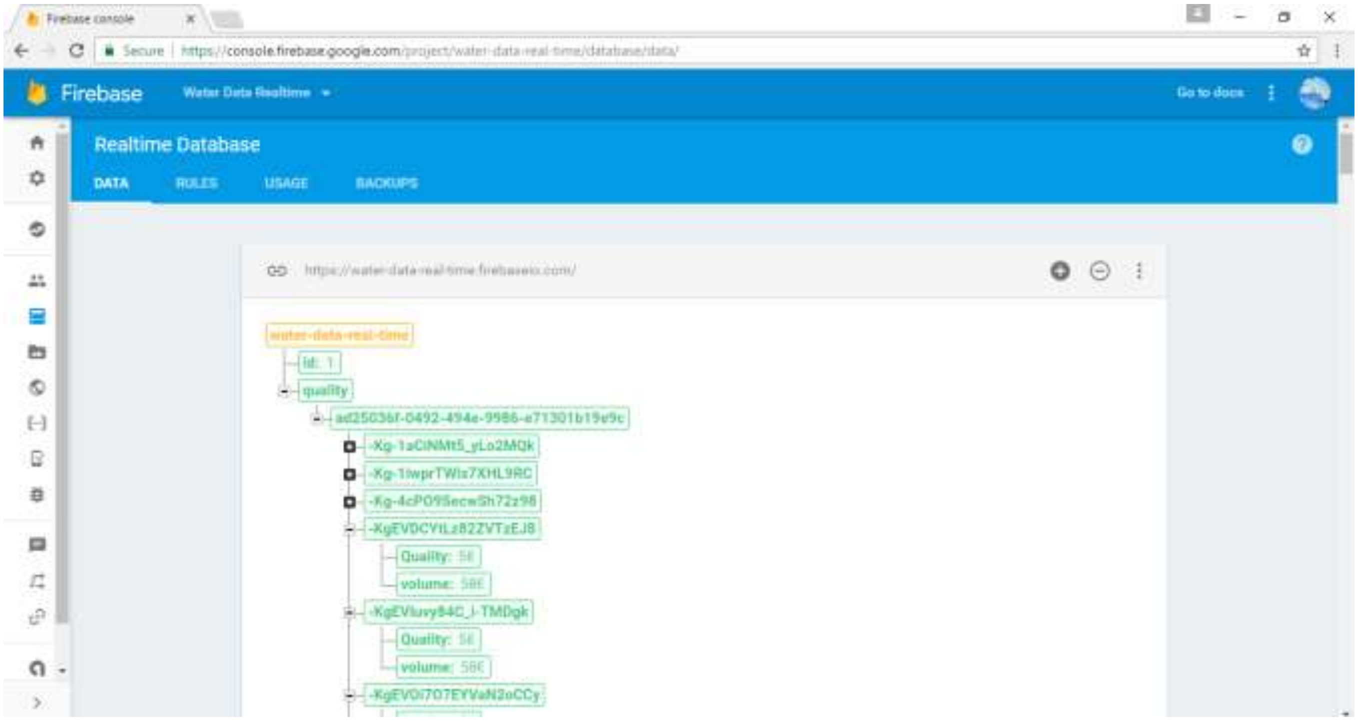

Figure 5:4: Real-time Database Data Capture Illustration 1

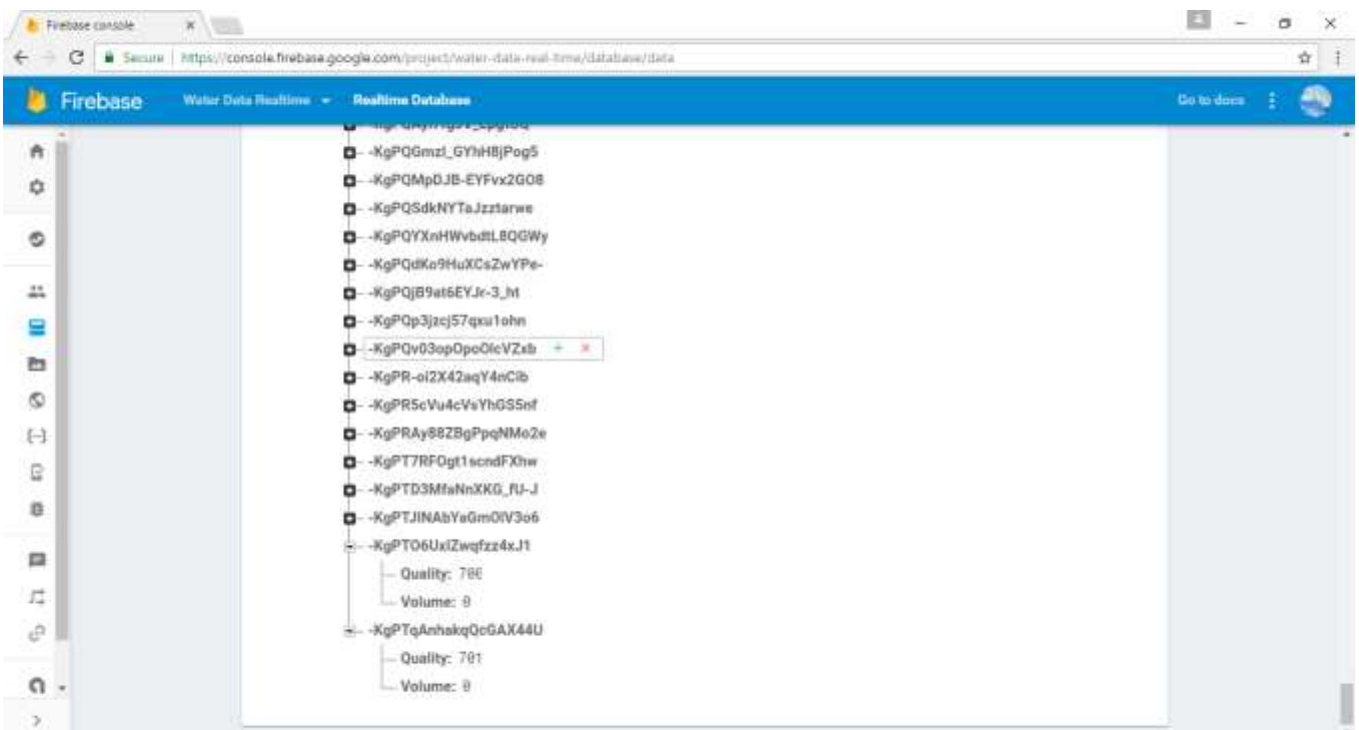

Figure 5:5: Real-time Database Data Capture Illustration 11

5.2.3 Web Application Implementation

The Web Application is meant to enable the end users view the real-time data captured by the sensors. The application was built using bootstrap which is a framework that has combined CSS, HTML and JavaScript. The features on the web application enable the visualization of the charts and calculation of the regulator's data for performance, related to the data captured. The webpage is illustrated in Figures 5.6 and 5.7.

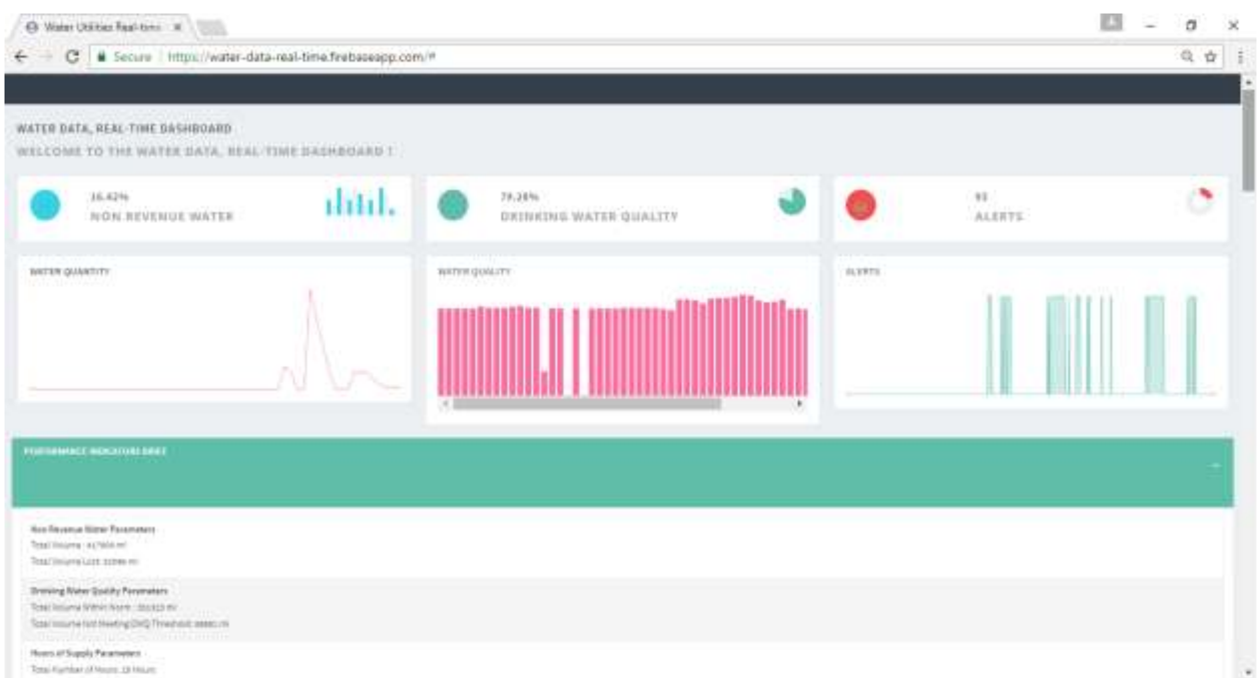

Figure 5:6: Web Application Visualization Page I

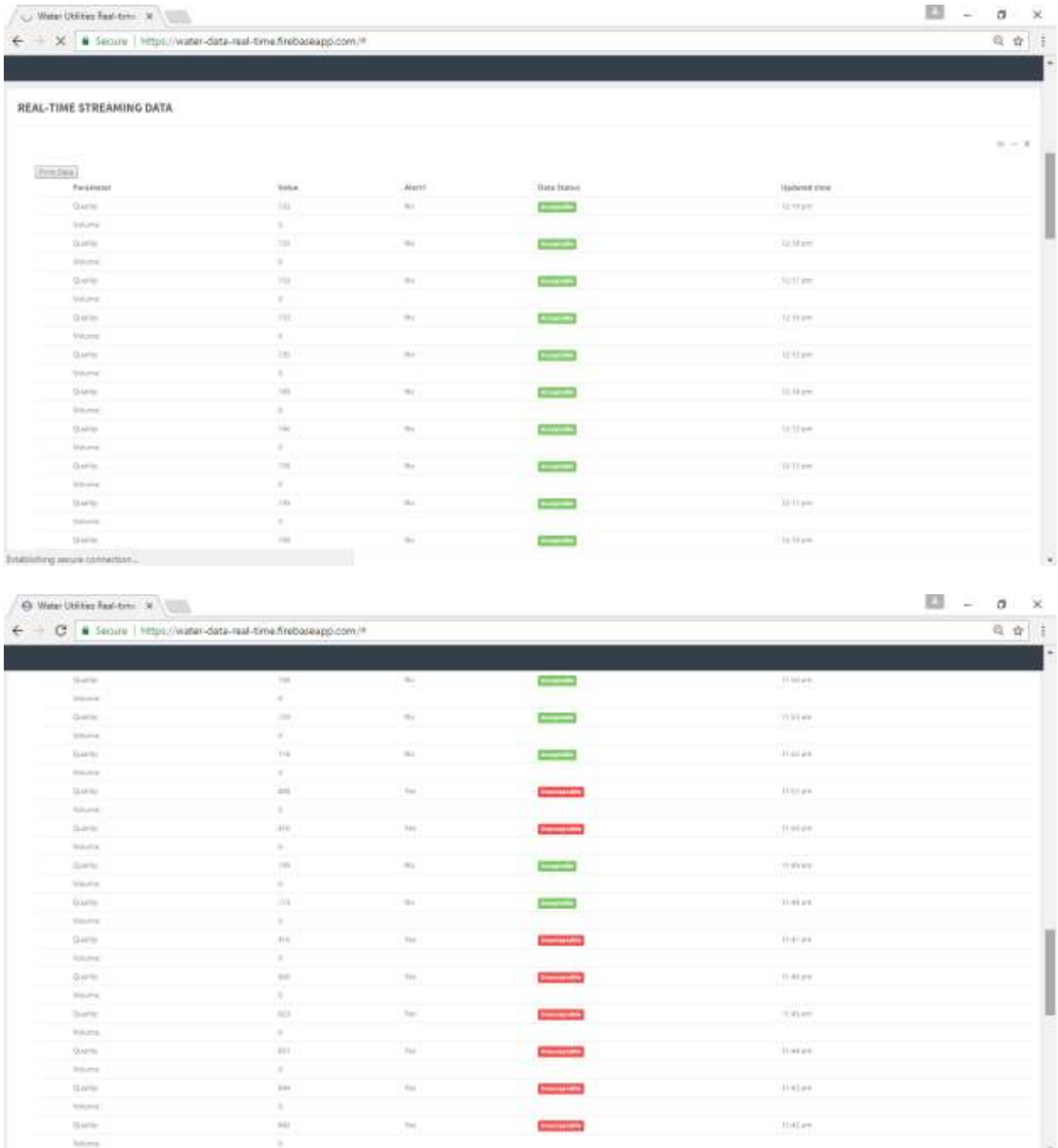

Figure 5:7: Web Application Visualization Page II

The web pages in Figure 5.6 show the different real-time charts that will be displayed from the real-time data. There is also a live feed showing the streaming water volume and quality data.

The screen shots of the visualization reports for water quality and volumes are as in Figures 5.8 and 5.9 respectively. Figure 5.10 shows a visualization of the aggregated alerts.

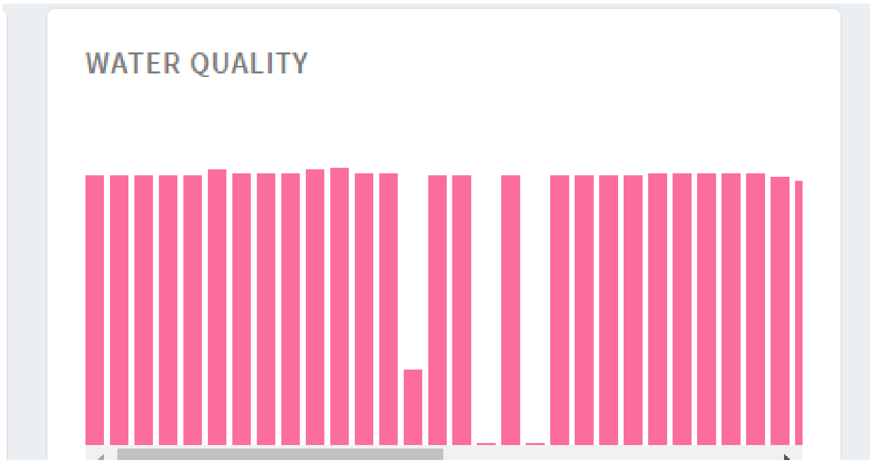

Figure 5:8: Web Application Real-time Water Quality Chart

The water quality chart will show the real-time recorded water quality data.

Figure 5:9 Web Application Real-time Water Quantity Chart

The water quantity chart will show the real-time recorded water volumes data.

The alerts chart shows the number of alerts triggered by recording the number of unacceptable water quality values as recorded by the turbidity sensor. In addition to the charts, the prototype will demonstrates calculation of three KPIs as defined in Annex A. The three KPIs are linked directly to the data captured by the prototype- which are Water Quality, Non-Revenue Water and Hours of Supply (Figure 5.11). This will enable timely decision making and real time monitoring.

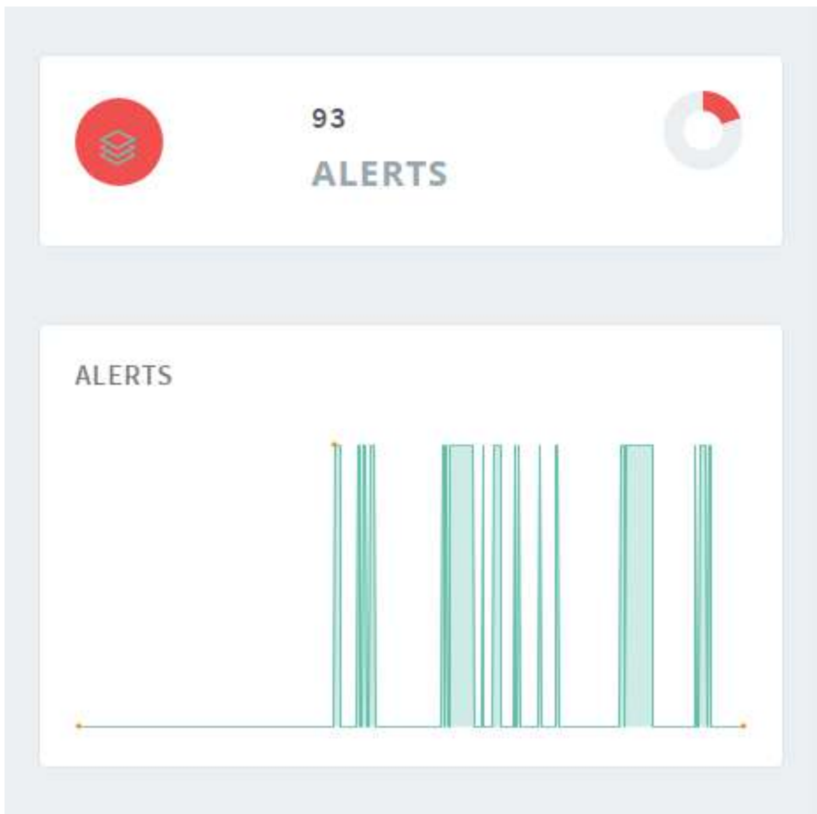

Figure 5:10 Web Application Alerts Chart

Figure 5:11 Calculation of Performance Indicators

5.3 System Testing

Testing of data capture was done to ensure that the required data is captured by the sensors, and sent to the cloud data base. Testing was also done to verify that the data is displayed in the visualization web page on the web app. The Table 5.2 presents the test results of the Record Data Use Case. The test results for Upload Data, Display Data and Issue Alerts, which demonstrate the successful integration of the hardware, successful communication and recording of data in database, are in Appendix E.

Table 5:2 Record Data Use Case Testing

Test Case Name	Record Data	Test Case Number: 1	
Brief Description	Data is captured by the sensors		
Pre Conditions	GPRS is initialized Sensors are working		
Steps	**Action**	**Expected Result**	**Pass/Fail**
Step 1	GPRS is initialized	Continuous blinking of the GPRS LED light	Pass
Step 2	Water quality and water volumes data is captured	Data displayed on the database	Pass

5.4 Summary

The developed prototype passed all the required tests and was able to send the operational data to the cloud database, and there was the successful visualization of the data from the web application. The real-time data enables timely decision making.

Chapter 6: Discussions

6.1 Introduction

This chapter focuses on the discussions following the successful design, implementation and testing of the real-time water data system. The research aimed to provide a prototype to demonstrate how water utilities can be monitored using wireless sensor networks. The chapter also discusses the validation of the prototype.

6.2 Interpretations and Opinions

The real-time data water system will aid the water services regulator with real-time information, hence enhance the decision making process. The prototype presented a scenario in which water quality and water quantity can be measured in real-time. The water quantity parameter that was measured by the system is the water volumes. The other parameters constituting water quantity are the water flow velocity, and water level. At the moment, more that 40% of water produced by water utilities is lost due to commercial losses in the water network (Water Services Regulatory Board, 2016). A smart monitoring system as the one proposed will enable early leakage detection, hence prevent water loss.

The World Health Organization guideline for Drinking Water Quality provides examples of operational monitoring parameters that can be used to monitor control measured. They are summarized in the Table 6.1.

In addition to the real-time water volume monitoring, the system detects real-time contamination of the water which should meet the prescribed drinking water quality standards. It should be noted that the drinking water quality parameters are diverse in number and include the conventional parameters, that is, PH Value, Temperature, Dissolved Oxygen (DO), Electrical Conductivity (EC), Hardness, Chloride, Fluoride and Total Suspended Solids. In additional to these parameters, the nutrient salt, inorganic ions, heavy metals and organic matter should also be measured. Water quality is very critical as it directly impacts on the health and early detection of anomalies is very important. The issuing of an alert by the system in the event an anomaly is detected was demonstrated in the prototype.

Table 6:1 Water Quality Monitoring Parameters

(World Health Organization, 2006)

Operational parameter	Raw water	Coagulation	Sedimentation	Filtration	Disinfection	Distribution system
pH		✓	✓		✓	✓
Turbidity (or particle count)	✓	✓	✓	✓	✓	✓
Dissolved oxygen	✓					
Stream/river flow	✓					
Rainfall	✓					
Colour	✓					
Conductivity (total dissolved solids, or TDS)	✓					
Organic carbon	✓		✓			
Algae, algal toxins and metabolites	✓					✓
Chemical dosage		✓			✓	
Flow rate		✓	✓	✓	✓	
Net charge		✓				
Streaming current value		✓				
Headloss				✓		
Ct[a]					✓	
Disinfectant residual					✓	✓
Oxidation–reduction potential (ORP)					✓	
DBPs					✓	✓
Hydraulic pressure						✓

[a] Ct = Disinfectant concentration × contact time.

6.3 Validation of the System

Validation of the system is evident from the test results obtained from the previous chapter. The expected results and the actual obtained results were similar. The Tables 6.2 to 6.4 provides the validation check list of the system. The validation is done against the stakeholder requirements, system requirements, the design and the process.

Table 6:2 Validation of System: Stakeholder and System Requirements

Item	Validation Check		
	Expectations	**Actual Results**	**Comments**
Stakeholder Requirements	From the interviews it was identified that the stakeholders desired: ▪ A more efficient way to ensure data is timely. ▪ A way to ensure that the data is consistent. ▪ Improvement in the nature of reports. ▪ Increase in reporting frequency.	The real-time water data system captures the live data from the sensors. In addition to this, the system through the web application offers visualization of the real-time data. This enables the tracking of performance indicators for the water utilities effectively.	The real-time water data system has achieved the requirements as that data is captured real-time, hence ensuring accuracy and timeliness of the data. This will in turn ensure usability, availability, reliability, and improve on presentation quality.
System Requirement	An effective real-time system that ensures real-time data monitoring	The real-time system is available online on the cloud.	The system has met the requirements

Table 6:3 Validation of System: Design

Item	Validation Check		
	Expectations	**Actual Results**	**Comments**
Design	▪ The system must use the wireless sensor network architecture. ▪ The system must be reliable, that is, ensure 100% availability. ▪ The system must enable real-time transmission of data for water quality and water flow and the information should be available online. ▪ The system must be able to issue alerts in the event the data captured is not the acceptable values. ▪ The system should be self adaptive for instance, in the event contaminants are detected, the valve should immediately stop transmission of the water. ▪ The system must be accessible with any device connected to the internet.	The real-time system is built on the wireless sensor network architecture. It is hosted on a reliable cloud database and ensures real-time transmission of water quality data through establishing the turbidity of the water, and water volumes, though the use of the hall effect sensor. The system also is adaptive in the sense that in the event of system contamination, no further flow of the water is permitted. On testing the application accessibility from any device with a browser, the web application is accessible, and the view is user friendly.	The system has met the requirements

Table 6:4 Validation of System: Process

Item	Validation Check		
	Expectations	Actual Results	Comments
Process	The system aims is to enhance the monitoring process by the regulator.	The real-time monitoring enables real-time decision making. The lengthy data validation process that the regulator goes through is eliminated.	The system has met the requirements.

6.4 Contribution of the Research to the Body of Knowledge

According to Mutchek and Williams (Mutchek & Williams, 2014), integrative, strategic and macro-level discussions for smart water grids are lacking in academic and other literatures. This research will contribute to this knowledge base.

The prototype has also proven that the water services regulator will enhance performance monitoring of utilities by having real-time data. Adoption of smart water infrastructures is inevitable in today's information driven society. Real-time monitoring of operational data will also ensure timely and effective decision making. The prototype will provide a basis for setting up such an infrastructure.

6.5 Achievement of Research Objectives

There were four main research objectives at the beginning of the research. The first objective was to investigate the challenges that the water services regulator has when capturing monitoring data. These challenges were identified from interviewing the staff who are directly involved in the data capture. The challenges were also identified from reviewing of publications. The main challenges were in the timeliness and consistency of data. The rigorous process of data validation is also lengthy, hence time consuming. A real-time system using wireless sensor networks as illustrated in the previous chapters will overcome this challenge.

The second objective was to review the wireless sensor network technology as a method to capture data for monitoring. This was done by reviewing the architecture in detail. The prototype was conceptualized using this architecture. The architecture enables wirelessly connected sensors to cooperatively sense and control their environment.

The third objective was to propose a prototype using wireless sensor networks to capture monitoring data. The prototype presented options for capturing a water volume element- flow volumes, and water quality element- turbidity. There was also communication between the sensors hence making the prototype adaptive. The real-time data capture is done at source.

The fourth objective was to evaluate the protoype. The protoype was evaluated and the stakeholders' expectations were compared against the system. The system met all the prescribed requirements.

6.6 Shortfalls of the Research

The main shortfall of the research is the scope. The scope of the research was limited to only a water volume element- flow volumes, and water quality element- turbidity. As seen from the initial discussions, there is a wide array of elements which can be captured by sensors. These additional parameters are illustrated in table 6.1. The scope was limited due to the unavailability of the sensors, notwithstanding, confined within set budget and time.

The research only focused on the water services regulator's needs and therefore did not include an interactive customer portal, in line with the customer meter numbers. This inclusion would have further enriched the system by help the customers log in to the various accounts and monitor usage as well as the water quality.

Chapter 7: Conclusions and Recommendations

7.1 Conclusions

To be effective, smart systems need to be instrumented, interconnected, and intelligent. The main challenges have been how users interact with, manage and control an ensemble of devices in this connected environment (IEEE Computer Society, 2015). The benefits of smart water networks cut across various domains in the society. There is hence a need for research and development is to enable such work.

Smart water grids as integrated systems have just started to gain popularity among researchers and utilities. Focus of such has been in mostly in electricity as utility. In the literature- Moving Towards Sustainable and Resilient Smart Water Grids, there are efforts to develop and analyze the components of smart water grids (Mutchek & Williams, 2014). Integrative, strategic, and macro-level discussions of smart water grids are lacking in academic and other literatures.

There has been a problem of integration, as remote sensing systems in place are not typically well integrated with analytics and modeling software, leading to non-standard, hard-to-support solutions and inefficiencies in system operation, including management of overwhelming amounts of data. This can be addressed by the use of prototyping, such as with the scenario captured by this paper.

All utilities are unique and one model for an AMR or AMI system would not meet the needs of all water utilities. However, when examining the full scope of an AMI deployment, the architecture itself would look very similar to the electric system. Instrumentation enables the collection of timely, high-quality data through embedded sensors that communicate over wireless networks. Instrumented devices such as smart meters for gas, electricity, and water continually monitor the supply and demand for these utilities, and might actuate strategies devised by intelligent components (IEEE Computer Society, 2015).

Interconnection creates linkages among data, systems, and people. High degrees of interconnectivity enable smart systems to become reality. The interconnections among people, objects, and systems are enabling new ways to gather, share, and act on information. Intelligence in the form of new computing models, algorithms, and advanced analytics will enable better decisions and outcomes for government, organizations and individual users. Smart connected objects, whether embedded, mobile, or wearable, will generate tremendous amounts of useful data to enable the development, deployment, and use of smart products and services (IEEE Computer Society, 2015).

The implementation of a smart water grid system supports the Water Supply Network's (WSN, Singapore) mission to supply good water 24/7 to its customers. The WSN Department manages the 5490 km of potable mains, 573 km of high-grade reclaimed water mains and 42 km of industrial water mains that deliver the water to more than 1.4 million customers (Allen, Preis, Iqbal, & Whittle, 2011). With sensors and analytic tools deployed island-wide to provide a real-time monitoring and decision support system, the smart water grid system enables WSN to manage the water supply network efficiently, ensuring that all Singaporeans will continue to enjoy a reliable and sustainable water supply for generations to come (Khuan, 2016).This is the researcher's vision of the water sector.

To more efficiently operate and manage water distribution systems, it is important for utilities to intelligently increase the amount of measurements made within the system It is important to use intelligent analytics to extract information from the data. A smart water grid will allow water utilities to better manage their infrastructure from source to consumer through improved monitoring and data analysis (Allen, Preis, Iqbal, & Whittle, 2011). This in turn will improve the quality of data for the water regulator's use (Water Services Regulatory Board, 2016). The proposed solution in the research will enable smart operational data accessibility by the regulator directly from the utilities' network.

7.2 Recommendations

The recommendations made from the research are as follows:

i. Adoption of a smart water infrastructure will enhance the monitoring efforts of the regulator will regards to the operational data.
ii. There is a broad application in the extent to what type of data can be monitored by the regulator. A focus on the operational and quality of service indicators, from

the 9 Key Performance Indicators in Annex A, indicates that Drinking Water Quality, Hours of Supply and Metering Ratio can be directly monitored from a smart network.

iii. The linkage of the real-time data to the regulator's data collection tool, WARIS will consolidate the water regulator's data and is proposed.

iv. The real-time data visualization should be drilled down to specific utilities as well, hence introducing maps and the use of GPS on the sensors for location accuracy.

7.3 Suggestions for Further Research

Integration of the multiple sensors would provide enriched monitoring and would introduce the element of synchronization of data received, hence the use of algorithms. Integration of the web application with end user consumers can also form a further basis of research, whereby customer perceptions on smart water grids and their impact and effectiveness will be captured.

This opportunity for further research will draw an emphasis on the benefits of smart water grids, as Mutchek and Williams (Mutchek & Williams, 2014) anticipate, and hence progressively address the changes as developments occurs.

REFERENCES

Allen, M., Preis, A., Iqbal, M., & Whittle, A. J. (2011). Case study: A Smart Water Grid in Singapore. *Case Study: A Smart Water Grid in Singapore*, 2.

Audit Commission for Local Authorities and The National Health Service in England. (2007). *Improving Information to Support Decision Making: Standards for Better Quality Data*. Milibank, London: Audit Commission.

Cerf, V., & Senges, M. (2016, February 10). *Announcing the Google Internet of Things (IoT) Technology Research Award Pilot*. Retrieved May 31, 2016, from Google Research Blog: http://googleresearch.blogspot.co.ke/2016/02/announcing-google-internet-of-things.html

Cisco. (2016). *Cisco 2016 Annual Report.* San Jose, Carlifornia: Cisco Systems.

D'Angelo, G., Ferretti, S., & Ghini, V. (2016). Simulation of the Internet of Things. *Proceedings of the IEEE 2016 International Conference on High Performance Computing and Simulation (HPCS).* Bologna.

Datawatch. (2015). You vs the Sensors: Six Requirements for Visualizing the Internet of Things (IoT). *Internet of Things North America Conference* (p. 13). Chicago: Gartner.

Du, R. (2016). Wireless Sensor Networks in Smart Cities: The Monitoring of Water Distribution Networks Case. Stockholm, Sweden: KTH Electrical Engineering.

Faustine, A., Mvuma, A., Mongi, H., Gabriel, M., Tenge, A., & Kucel, S. (2014). Wireless Sensor Networks for Water Quality Monitoring and Control within Lake Victoria Basin: Prototype Development. *Scientific Research*, 281-290.

Greg Weeks, D. T. (2014). *The Case For Smart Water And Why It's A No Brainer.* West Monroe.

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A V, Architectural Elements, and Future Directions. *Future Generation Computer Systems*, 1645-1660.

Hertzog, C. (2014, July 28). *http://www.smartgridlibrary.com*. Retrieved May 31, 2016, from Reimagining Infrastructure: http://www.smartgridlibrary.com/tag/potable-water-grid/

Herzog, T., Scheuren, F., & Winkler, W. (2007). What Is Quality Data And Why Should We Care? In T. Herzog, F. Scheuren, & W. Winkler, *Data Quality and Record Linkage Techniques* (pp. 7-15). ISBN.

Hu, X. (2007). *A Simulation-Based Software Development Methodology for Distributed Real-Time Systems.* Retrieved from acims.asu.edu/wp-content/uploads/2012/02/Xiaolin_dissertation.pdf

IEEE Computer Society. (2015, November/December). Smart Systems. *IT Pro*, pp. 14-17.

Innovation. (2016, September 16). Retrieved from US Environmental Protection Agency, EPA: https://www.epa.gov/innovation/water-technology-innovation-ten-market-opportunities

International Electrotechnical Commission. (2014). *Internet of Things: Wireless Sensor Networks.* Geneva: International Electrotechnical Commission.

International WaterAssociation, IWA. (2016, October). Crisis Management at Water Utilities Creating Resilient Organisations. Brisbane, Australia.

Internet of Things Architecture. (2016, July). Retrieved July 30, 2016, from Internet of Things Architecture: http://www.iot-a.eu/public

Internet Of Things Overview. (2016). Retrieved August 15, 2016, from Cisco : http://www.cisco.com/c/dam/en/us/products/collateral/se/internet-of-things/at-a-glance-c45-731471.pdf

Kenya Gazette. (2016). Water Act 2016. Nairobi.

Khuan, P. J. (2016). Managing the water distribution with a smart water grid. *Public Utilities Board Singapore Smart Water (2016)*, 1-2.

Kolenc, M., & Zajc, M. (2007, June). Wireless Sensor Networks and Data Analysis in Smart Grids. Tuzla.

Li Cai, Y. Z. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 2.

Microsoft Azure. (n.d.). *NoSQL vs SQL*. Retrieved March 9, 2017, from https://docs.microsoft.com: https://docs.microsoft.com/en-us/azure/documentdb/documentdb-nosql-vs-sql

Mutchek, M., & Williams, E. (2014). Moving Towards Sustainable and Resilient Smart Water Grids. *Challenges*, 123-137.

Mwangi, C. (2016, June). Mobile Water Metering System Based on a Hall Effect Sensor. Nairobi, Kenya: Strathmore University.

Nairobi Water Targets Big Consumers With Smart Meters. (2014, March 20). Retrieved August 10, 2016, from Business Daily: http://www.businessdailyafrica.com/Corporate-News/Nairobi-Water-targets-big-consumers-with-smart-meters/539550-2263514-6jjpcvz/index.html

Ovidiu Vermesan, P. F.-C. (2013). *Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems.* Aalborg: River Publishers.

Sekaran, U. (2004). *Research Methods for Business: A Skill-Building Approach.* Massachusetts: John Wiley & Sons, Inc.

Springer. (2016). *Wireless Sensor Networks: Concepts and Components*. Retrieved February 2016, from http://www.springer.com

Team Ray Technologies. (2015). *Real-Time Data Collection.* Retrieved January 10, 2017, from Team Ray Technologies: https://www.rockwarecorp.com/

The Hammersmith Group. (2010). *The Internet of things: Networked Objects and Smart Devices.* New York: Constantine A. Valhouli and The Hammersmith Group.

Vermesan, O. (2014). Internet of Things beyond the Hype: Research, Innovation and Deployment. *Internet of Things beyond the Hype: Research, Innovation and Deployment*, 15-112.

Vermesan, O., & Freiss, P. (2013). *Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems.* Aalborg: River Publishers.

Water Services Regulatory Board. (2015). *Impact Report, A Performance review of Kenya's Water Services 2013-2014.* Nairobi: Tara Consultant Limited.

Water Services Regulatory Board. (2016). *Impact Report, A Performance review of Kenya's Water Services 2014-2015.* Nairobi: Tara Consultant Limited.

What is SCADA?: Inductive Automation. (2016, October 26). Retrieved from Inductive Automation: https://inductiveautomation.com/what-is-scada

World Health Organization. (2006). Guidelines for Drinking Water Quality. Geneva, Switzerland.

Ye, Y., Liang, L., Zhao, H., & Jiang, Y. (2016). The System Architecture of Smart Water Grid for Water Security. Beijing, China.

APPENDICES

Appendix A: Definition of WASREB's 9 Key Performance Indicators

Appendix A1: Definition of WASREB's 9 Key Performance Indicators

KEY PERFORMANCE INDICATOR	DEFINITION
Water Coverage	Water Coverage refers to the number of people served with drinking water by a utility expressed as a percentage of the total population within the service area of the utility. It assesses performance in executing the core mandate of the utility of supplying potable water to consumers.
Drinking Water Quality	Drinking Water Quality (DWQ) measures the portability of the water supplied by a utility. It is a critical performance indicator since it has a direct impact on the health of consumers. This is a composite indicator measuring compliance with residual chlorine standards (40%) and bacteriological standards (60%).
Hours of Supply	Hours of Supply refers to the average number of hours per day that a utility provides water to its customers. It measures the continuity of services of a utility and thus the availability of water to the customer. It is an important indicator of service quality and shows the extent to which the utility is making progress towards the fulfillment of the human right to water and sanitation in terms of availability of water in sufficient quantities.
Non-Revenue Water	Non-Revenue Water (NRW) refers to the difference between the amount of water produced for distribution and the amount of water billed to customers. It measures the efficiency of the utility in delivering the water it produces to customer take-off points. It captures both technical losses (leakages) and commercial losses (illegal connections/water theft, metering errors and unbilled authorized consumption). High levels of NRW indicate that utilities are losing revenue and will not be able to render proper service in terms of water availability and price.

Appendix A2: Definition of WASREB's 9 Key Performance Indicators

KEY PERFORMANCE INDICATOR	DEFINITION
Metering Ratio	Metering Ratio refers to the number of connections with functional meters expressed as a percentage of the total number of active water connections. It measures the extent to which a utility has implemented metering as a tool to manage NRW so that consumers can only pay for what they consume.
Staff Productivity	Staff Productivity refers to the number of staff in employment for every 1,000 connections (total registered water and, where applicable, sewer connections). It measures the efficiency of utilities in utilizing its staff. Thus, a low figure is desirable.
Revenue Collection Efficiency	Revenue Collection Efficiency refers to the total amount collected by a utility expressed as a percentage of the total amount billed in a given period. It measures the effectiveness of the revenue management system of a utility.
O+M Cost Coverage	Operation and Maintenance (O+M) Cost Coverage is the extent to which internally generated funds cover the cost of running a utility.
Personnel Expenditure as a percentage of O+M Costs	Personnel expenditures as a percentage of O+M Costs measures whether personnel related expenses are proportionate to overall O+M costs

Appendix B: Interview Guide

I. Accessibility

- Is there a current data interface to enable accessibility of data?
- Is the data collected publicly available?

II. Timeliness

- Does the data arrive on time?
- Is the data regularly updated?
- Does the time interval from data collection and processing to release meet requirements?

III. Credibility

- Is specialized data acquired from other institutions in the industry?
- Do experts or specialists regularly audit and check the correctness of the data content?
- Does data exist within the range of known or acceptable values?

IV. Accuracy?

- How would you gauge the accuracy of data submitted?
- Does the data representation (or value) reflects the true state of the source information?
- Are there cases of ambiguity?

V. Consistency

- After data has been processed, do their concepts, value domains, and formats still match as before processing?
- During a certain time, does data remain consistent and verifiable?
- Is the data and the data from other data sources consistent or verifiable?

VI. Integrity

- Is the data format clear and does it meets the criteria?
- Is data consistent with structural and content integrity?

VII. Completeness

- Will the deficiency of a component impact the use of the data for data with multi-components?
- Will the deficiency of a component impact data accuracy and integrity?

VIII. Fitness

- Are most datasets retrieved within WASREB's need?

IX. Readability

- Is the data (content and format) clear and understandable?
- Is it easy to judge that the data provided meet needs?
- Is the data description, classification, and coding content satisfy specification and is easy to understand?

Appendix C: Sample Non Revenue Water Quality Report Extract

Water Service Provider

1st Quarter 2016/17 NRW Data

		Jul-16			Aug-16			Sep-16			1st Quarter 2016/2017		
No	Region/Zone/DMA	Water Produced M^3	Water Billed M^3	NRW Ave. %	Water Produced M^3	Water Billed M^3	NRW Ave. %	Water Produced M^3	Water Billed M^3	NRW Ave. %	Water Produced M^3	Water Billed M^3	NRW Ave. %
1	a	443,570	351,727	**20.71**	466,968	343,591	**26.42**	458,394	372,631	**18.71**	1,368,932	1,067,949	**21.99**
2	b	334,413	209,802	**37.26**	359,416	234,605	**34.73**	364,751	234,825	**35.62**	1,058,580	679,232	**35.84**
3	c	189,799	140,364	**26.05**	197,327	156,986	**20.44**	189,489	144,664	**23.66**	576,615	442,014	**23.34**
4	d	126,320	77,976	**38.27**	125,040	75,490	**39.63**	127,320	77,866	**38.84**	378,680	231,332	**38.91**
5	e	121,390	85,676	**29.42**	154,150	106,130	**31.15**	249,000	165,370	**33.59**	524,540	357,176	**31.91**
6	f	48,634	33,984	**30.12**	48,634	37,241	**23.43**	48,634	36,902	**24.12**	145,902	108,127	**25.89**
7	g	497,455	170,415	**65.74**	494,042	171,365	**65.31**	447,792	160,197	**64.23**	1,439,289	501,977	**65.12**
8	h	92,060	50,840	**44.78**	95,441	54,703	**42.68**	105,645	56,526	**46.49**	293,146	162,069	**44.71**
9	i	96,979	55,000	**43.29**	100,170	50,274	**49.81**	110,893	59,022	**46.78**	308,042	164,296	**46.66**
10	j	16,225	9,952	**38.66**	35,445	25,663	**27.60**	34203	25,444	**25.61**	85,873	61,059	**28.90**
11	k	30,000	19,298	**35.67**	30,000	19,352	**35.49**	30,000	22,066	**26.45**	90,000	60,716	**32.54**
	Total/Average	**1,996,845**	**1,205,034**	37.27	**2,106,633**	**1,275,400**	**36.06**	**2,166,121**	**1,355,513**	**34.92**	**6,269,599**	**3,835,947**	**35.98**

Appendix D: Use Case Table

Appendix D1: Record Data Use Case Description

Use Case Name	Record Data	Identifier: 2
Brief Description	Data is captured by the sensors	
Type	Internal	
Major Inputs	**Source** Water Flow Sensor, Water Quality Sensor	**Major Outputs** Captured data on water quality and water volumes
Pre Conditions	GPRS is initialized Sensors are working	
Post Conditions	Water quality and water volumes data is captured	
Flow of Events	1. GPRS is initialized 2. Water quality and water volumes data is captured by the microcontroller	

Appendix D2: Upload Data Use Case Description

Use Case Name	Upload Data	Identifier: 3
Brief Description	Captured data is uploaded to the cloud database	
Type	External	
Major Inputs	**Source** Water Flow Sensor data, Water Quality Sensor data	**Major Outputs** Data on water quality and water volumes is accessible online
Pre Conditions	Sensor data has been captured	
Post Conditions	Data on water quality and water volumes is available on the cloud database	
Flow of Events	1. Water quality and water volumes data is captured 2. Using the wireless network, data is uploaded to the cloud storage	

Appendix D3: Display Data Use Case Description

Use Case Name	Display Data	Identifier: 4
Brief Description	Captured data in the cloud database is displayed on the web application	
Type	External	
Major Inputs	**Source** Water Flow Sensor data, Water Quality Sensor data	**Major Outputs** Visualization of the captured data
Pre Conditions	Sensor data has been uploaded to the cloud database	
Post Conditions	Data on water quality and water volumes is accessible online in the form of charts and graphs	
Flow of Events	1. Water quality and water volumes data in cloud database is visually displayed 2. Charts and graphs can be generated based on the graphs	

Appendix D4: Issue Alerts Use Case Description

Use Case Name	Issue Alerts	Identifier: 5
Brief Description	Alerts are issued when contaminants are detected in the water	
Type	External	
Major Inputs	**Source** Water Quality Sensor data	**Major Outputs** Stop flow of water, valve is closed
Pre Conditions	Contaminants detected in the water	
Post Conditions	The flow control valve closes, and therefore no water is transmitted	
Flow of Events	1. Contaminants detected in the water 2. The flow control valve closes, and therefore no water is transmitted	

Appendix E: Test Cases

Appendix E1: Upload Data Use Case Testing

Use Case Name	Upload Data	Test Case Number: 2	
Brief Description	Captured data is uploaded to the cloud database		
Pre Conditions	Sensor data has been captured		
Steps	**Action**	**Expected Result**	**Pass/Fail**
Step 1	Water Flow Sensor data is uploaded to database	Reading of captured volumes reflected on the database	Pass
Step 2	Water Quality Sensor data is uploaded to the database	Reading of captured turbidity data reflected on the database	Pass

Appendix E2: Display Data Use Case Testing

Use Case Name	Display Data	Test Case Number: 3	
Brief Description	Captured data in the cloud database is displayed on the web application		
Pre Conditions	Sensor data has been uploaded to the cloud database		
Steps	**Action**	**Expected Result**	**Pass/Fail**
Step 1	Water volumes that have passed through the flow sensor are displayed on the web application interface	Visualization charts of the captured volumes data	Pass
Step 2	Water turbidity values that have been captured by the water quality turbidity sensor are displayed on the web application interface	Visualization charts of the captured turbidity values data	Pass

Appendix E3: Issue Alerts Use Case Testing

Use Case Name	Issue Alerts	Test Case Number: 4	
Brief Description	Alerts are issued when contaminants are detected by the turbidity sensor in the water		
Pre Conditions	Contaminants detected in the water		
Steps	**Action**	**Expected Result**	**Pass/Fail**
Step 1	Water Quality data is captured by the water quality sensor as having a dirt threshold of greater than 50	Stop flow of water, valve is closed	Pass

Appendix F: Originality Report

ENHANCED MONITORING OF KENYAN WATER UTILITIES USING WSNs

ORIGINALITY REPORT

27%	21%	13%	17%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Strathmore University Student Paper	2%
2	Allen, Michael. "Case study: a smart water grid in Singapore", Water Practice & Technology, 2012. Publication	2%
3	www.mdpi.com Internet Source	2%
4	www.wao.gov.uk Internet Source	1%
5	datascience.codata.org Internet Source	1%
6	www.internet-of-things-research.eu	1%

FILE	RESEARCH_THESIS_008093.DOCX (2.97M)		
TIME SUBMITTED	06-APR-2017 05:33PM	WORD COUNT	16725
SUBMISSION ID	738948786	CHARACTER COUNT	105590