Zana Pekmez
University of Sarajevo
Faculty of Economics, Bosnia & Herzegovina
E-mail: zana_pekmez@hotmail.com

MINING BIG DATA FOR SUSTAINABLE WATER MANAGEMENT

Review
UDK: 626/628:004.6
JEL classification: Q55
Accepted for publishing: October 31, 2019

Abstract

The power of advanced analytics is substantial. Massive scales of big, structured and unstructured data relieve unthinkable patterns and help us redefine economic models, solve operational inefficiencies and optimize costs. The water utilities could substantially benefit from the data available from new digital assets and smart technologies. Many are facing damaged and failing infrastructure and lack of financial resources for makeovers. However, Industry 4.0 and Digitalization open new fronts and bring new assets such as real-time monitoring of critical systems via IoT and sensors, advanced metering and predictive analytics to improve customer billing, remote data collection systems at pumping stations and water storage facilities and many more. The power of "digital twin", as a virtual replica of a physical asset, and ways of enriching the traditional data sources with open source data increase considerably the available intelligence for more sophisticated correlation, linkages and insights. This study reviews the core values of big data, advanced analytics, smart technologies and its application in water resources management and it gives concrete recommendation how to accelerate the adoption of use of Big Data by leveraging on technology and innovation.

Keywords: Big Data Analytics, Water Resources, Smart technologies

1. INTRODUCTION

Industry 4.0 brings forward and at faster pace the adoption of smart city initiatives, smart water infrastructure and entrepreneurial startups to commercialize and promote the use of cloud-based big data solution, advanced analytics and visualization tools to achieve cost and operating efficiency and environmental sustainability.

The intent of this paper is to show how important the nexus between big data analytics, cloud technologies and use of big data for water systems management and operations in the context of sustainable development goals. The vast amounts of data which are available in water industry if used efficiently could be translated into useful information and actionable insights to significantly improve water quality, optimize distribution processes and reduce costs. The key to this process is how to adequately leverage smart technologies in order to enable efficient use of water and close the gap between demand and supply. The objective of this article is also to provide specific recommendations how to accelerate the adoption of Big Data technologies in the context of sustainable water management and for achievement of sustainable development goals overall.

From reputable consultants we often hear concepts such as citizen-centricity, future of work, human+worker (Accenture 2019) that all encompass the notion of rapidly changing conditions of how we work, collaborate and contribute. The innovation brought by artificial intelligence (AI) and predictive analytics will lead to smart digital solutions vitally improving the means how technology is utilized to accelerate plant´s performance, supply chains and distribution systems, infrastructure monitoring and disaster management.

It is important to emphasize that this process is an amalgam of many factors, namely data in the structured and unstructured form coming from multiple sources, robust data sourcing pipelines, infrastructure and technology stack that will enable processing & storing of such large volumes of data; cloud technologies that enable scale and deployment and finally the governance and organization fit and adaptive to support all this.

1.1. Nexus between Big Data & Sustainable Development Goals (SDGs)

In 2017 UN Global Puls published 17 ways on how Big Data and analytics could contribute to sustainable development goals (SDGs). Out of total 17 mechanisms which have been described in the UN document the following are the most relevant and interface directly with the topic of sustainable water management, namely:

Big Data & Sustainable Development Goals

How data science and analytics can contribute to sustainable development

6. Clean water and sanitation	*Sensors connected to water pumps can track access to clean water.*
7. Affordable and clean energy	*Smart metering allows utility companies to increase or restrict the flow of electricity, gas, or water to reduce waste and ensure adequate supply at peak periods.*
9. Industry innovation and infrastructure	*Data from GPS devices can be used for traffic control and to improve public transport.*
11. Sustainable cities and communities	*Satellite remote sensing can track encroachment on public land or spaces such as parks and forests.*
12. Responsible consumption and production	*Online search patterns or e-commerce transactions can reveal the pace of transition to energy efficient products.*
17. Partnership for the goals	*Partnerships to enable the combining of statistics, mobile and internet data can provide a better and realtime understanding of today's hyper-connected world.*

Notes: Non-exhaustive list

Source: https://www.un.org/en/pdfs/Bigdata_SDGs_single_spread_2017.pdf

The other principles in the Big Data SDGs Framework are also very relevant to the topic of this paper since all of them are supporting the idea of partnership to enable transparent statistics and patters recognition which may help many vital causes, such as poverty, climate, gender inequality. The advantages of Big Data in the context of sustainable development are often mentioned in the recent literature as a vital benefactor to *"governments and societies to make more effective local, regional, national and global progress toward truly sustainable societies."* (Song et al., 2012, 2015, p.2).

However, in this article the focus will be mainly on six above mentioned factors set in the context of the following specific use cases and examples for sustainable water management though Big Data analytics, smart technologies and cloud infrastructure, as well as data science tools.

2. OVERVIEW OF APPLICATIONS OF MINING BIG DATA FOR SUSTAINABLE WATER RESOURCE MANAGEMENT

The pace of adoption of Big Data Analytics and cloud technologies is critical introducing the economic impacts of ineffective water resource management.

Water utility companies usually use less than 40% of the total generated data (Thompson & Kadiyala, 2014). Probably one of the key reasons is because the companies are lacking knowledge and technology to enable them to efficiently combine structured and unstructured data from multiple sources, process them to obtain understandable insights and stream the insights in real time to be able to make actions in day-to-day operations. Furthermore, the data streaming process should be properly monitored and visualized for mining the data for details and continuously optimized. However, it is hardly seen in practice to have all the data from multiple disparate systems holistically evaluated, followed by actionable insights and decisions.

Use-cases of how innovative data analytics tools could contribute to sustainable water management are numerous. Adamala (2017) mentioned a few starting from planning optimum water systems, detecting ecosystem changes through big remote sensing and geographical information systems, scheduling irrigations, mitigating pollution etc.

Furthermore, IBM has invested significant research funds into the application of Big Data for watershed management. The main objective was to capture variety of data from disparate sources, such as meteorological data, surface and sub-surface groundwater levels data, rain monitoring metrics, soil moisture metrics and many other sensor data which makes sense to be used in their use-cases. Kerala Water Authority in India uses IBM solution based on big data technologies for optimizing water distribution on the city of Thiruvananthapuram counting more than 3.3 million inhabitants (Adamala, 2017).

MERRA Analytical Services (MERRA/AS) is an example of cloud-enabled Climate Analytics as a service (CAaaS) that uses MapReduce analytics to integrates observational data and models the synthesis of 26 key climate variables based on global temporal and spatially consistent datapoints. MERRA is used by NASA Center for Climate Simulation and is an extraordinary example of state-of-the-art cloud data service architecture, tools and technologies, policies and protocols required to operationalize the cloud-enabled end-to-end enterprise platform. Major capabilities of MERRA are mainly: cloud computing, high performance, scalable data management, software appliance virtualization, adaptive analytics, domain-harmonized APIs (Schnase et al., 2014).

Cloud based software platforms will enable scalable modelling with large amounts of data. Shrivastava, Pandiaraj and Jagadeesen (2014) describe their experience in developing data driven prognostic ARIMA[1] models for forecasting the future reservoir water levels based on the 5 years of historical time series data

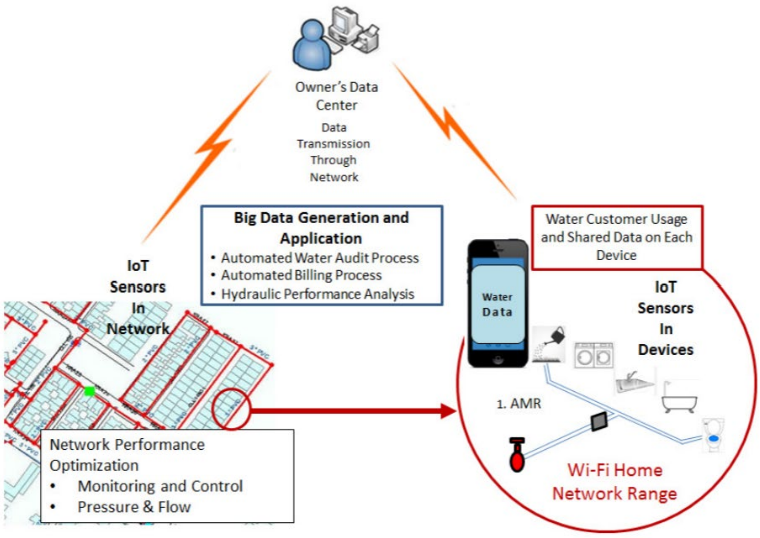
[1] ARIMA – Autoregressive Integrated Moving Average is an algorithmic rule for creating predictions (also called Box-Jenkins technique).

The breakthrough of Industry 4.0 is the phenomenon of applying Internet of Things (IoT) towards improving efficiency, costs, effective decision making and sustainability in general. Koo et al. (2015) describe three key conceptual components of IoT: sensing hardware, data transmission network and data process capability (Koo et al. 2015, pp. 490).

Implementing IoT and Big Data technologies together fully interface with all aspects of SDGs mentioned above and tackles precisely the following concrete benefits: mechanical system performance improvement (leaks and vulnerabilities detection); optimization of system performance; anticipatory and pre-emptive decision making (Koo et al., 2015).

In the current water resources management supervisory control and data acquisition (SCADA) systems are often used to make data driven decision, however these systems focus on the historical data and descriptive statistics. These current asset management systems rely on inventory databases without the information which is usually received from automated and smart meter reading applications (AMR) such as water loss, location of raptures and leaks and consumption (Koo et al., 2015).

Smart data solutions for sustainable water management are much more then what is currently seen used in practice. Smart data solutions are an amalgam of several key technologies and enablers: Big Data technologies, cloud services, connectivity and networking provided by IoT main functions; intelligence coming from prescriptive data analytics tools and machine learning algorithms – all this orchestrated in the overarching infrastructure system. This intelligent system will power utilities to anticipate future events and challenges ahead in the right time for needed action to operate at an optimal level.



Source: Koo et al. (2015). Towards Sustainable Water Supply: Schematic Development of Big Data Collection Using Internet of Things (IoT). Procedia Engineering, 118, pp. 495.

Figure 1 An Example of IoT and Big Data Applications

3. MAJOR CHALLANGES IN IMPLEMENTATION OF BIG DATA CAPITAL PROJECTS

Greatest challenges these companies are facing are unquestionably focused on the matter of holistic and robust infrastructure which will enable end-to-end management of big data. Lack of inhouse knowledge capabilities and the knowledge on how to use and put smart technologies in action remains an issue for most enterprises. In order to have the adequate insights required for fast decision making in the *"right time"*, online analysis of streamlining data is required. Sensors are collecting data from various sources, such as rain, marine life and soil and the goal is to merge and process this data together to obtain meaningful correlations from these datasets.

Processing big data is challenging and it requires large computing platforms, especially having in mind the ever-increasing connectivity of assets in the entire *Machine-to-Machine* and *Internet of Everything ecosystems*. Migrating the data to the cloud and scaling these operations in the cloud will become favored choice and is more cost effective, but this too remains a great challenge for most of the companies nowadays, namely the process of cloud migration and building cloud native services.

Following factors are identified as key challenges most utilities are facing in the process of planning a data strategy to leverage big data and smart technologies for sustainable water management. In the rest of the article some recommendations are given in how to tackle these roadblocks.

3.1. Architecting End-to-End (E2E) Data Logistics Pipeline

Data infrastructure strategy and orchestrating seamless data pipelines for smart solutions, digital twins, process visualizations in real-time (right-time) and leveraging technology disruptors seems to remain most complex difficulties for most of the water utilities.

3.2. Modelling approach and Data Solutions

Lack of innovation and talent to apply data science to predict outcomes, blockages, spills and therefore improve performance metrics and increase efficiency are very common challenge most utilities are facing nowadays.

Smart data solutions need to integrate available and existing legacy datasets with new data sources for more integrated and overarching view to support timely decision making.

Increasing volume, type and categories of data, besides distributed and parallel processing mode, require sophisticated modeling and computational capabilities of data analysis and scientist which are predominantly scarce and difficult to retain. Besides the aggressive acquisition strategy for data science talent companies often neglect to strategize on how to retain this talent: "The capability of the analysis staff is vital for the competitive power of nation and enterprise." (Wang & Sun, 2013, pp. 322).

3.3. Sustainable Infrastructure & Deployment *(Cloud adoption)*

To meet the demand of processing large volumes of streaming data, translating data into actionable insights via machine learning algorithms, visualizing dynamic dashboards and finally storing the raw data and results, adoption of cloud service-oriented architecture and technology is necessary (Wang & Sun, 2013).

Massive volumes of data that needs to be streamed and analyzed on the fly demand extensive storage and computing power accompanied by high energy costs. It is almost impossible to integrate and deploy disparate solutions on premise using physical server infrastructure. The appearance of computational infrastructure on demand ("*elastic cloud*") powered the proliferation of use of Big Data solutions.

After the model has been developed and tested successfully, it must be deployed in production. Besides building the deployment pipeline, additional challenge remains in sustaining the model operations and supporting running data products and services on the continuous basis. Fusing big data with the IoT technologies creates a very co-dependent relationship based on multiple service levels and demands more sophisticated, intelligent and automated processes for optimal decision making. Often this becomes a question of organizational fitness, governance, security and data compliance to be ensued through overarching and effective operating models. However, this also demands an introduction of machine learning and AI roles that will enable automated decision making without human interaction.

4. RECOMMENDATIONS HOW TO ACCELARATE ADOPTION OF BIG DATA TECHOLOGIES

Accenture 2019 Technology vision announced on of the 5 most vital trends for 2019 as ***"Internet of Thinking"*** which pertains to businesses designing "intelligent" ecosystems/frameworks and bringing them into production via most recent technologies. Such deployments demand from companies to develop their infrastructure beyond their own legacy systems. [2]

4.1. Designing a Roadmap & Data Master Plan

The operating model for such large-scale capital projects should be focused on strategic and technical capabilities, as well as needed management process with agile and flexible roadmap making sure that vital milestones are gradually achieved.

Goals to be achieved by a transformational capital projects should be aligned with the organization's all-encompassing digital strategy e.g. cloud & data strategy. Liner and Kenel in their article *"Exciting future for big data solutions"* (2016), mention six key aspects of implementing a big data platform in the water utilities, namely: integration, analytics, visualization, development, workload optimization, and security and governance.

Before starting with implementation water utilities must plan how the data will be used. The plan should go beyond data management and data warehousing and include the integration approach between different data siloes.

Cloud strategy is a key pillar of the overarching capital project and should be considered first and it fits well with above mentioned aspect of workload optimization which needs to focus on achieving efficient processing and data storage.

It is also important to consider data modelling approach earlier in the planning process since it will define most of the data requirements and needed data-end points to be sourced-in.

There are many other vital components that are key to successful implementation, however focusing on each individually goes beyond the scope of this article. However, it is worth mentioning that during the implementation lifecycle for large capital project in utilities with complex network of stakeholders involved (engineering teams, construction contractors, other service supplies), digital collaboration enabled by digital platforms and cloud-based collaboration tools plays really a critical role for project success (Accenture, 2019).

4.2. Modular approach to Architecture Layers

Current modular and elastic approach to using smart technologies and infrastructure enables enterprises to build linked software solutions at any level of services and manipulate diverse and distributed data models and databases.

[2] "Are you ready for what´s next?", Accenture 2019. Available at: https://www.accenture.com/_acnmedia/PDF-94/Accenture-TechVision-2019-Tech-Trends-Report.pdf

For real-time data processing an automatic data ingestion system needs to be in place to support data acquisition at completely different rates and sizes. Such modular architecture is usually built on the principles of application specific microservices that make all together a so-called *"product suite"* (Schnase et al., 2014, pp.4).

Wang and Sun (2013) describe the framework for an ideal water resources and hydropower platform in 6 different layers; physical, virtual, data layer, big data platform, service layer and application layer.

Modularity in architecting the cloud services permits agile and fast time to market of new data-native services and products and provides the groundwork for more sophisticated offerings.

4.3. Selection of Key Technologies & Modelling Tools

"An integrated solution has to be bigger than one technology." **(**Liner & Kenel, 2016, pp. 21).

Significant value in terms of scale, knowledge, infrastructure and other associated capabilities stems from commercial cloud providers which offer different frameworks such as "Infrastructure as Service" (IaaS) and Platform as Service (PaaS) and finally "Software as Service" (SaaS). Combining the advantage of all three frameworks for sure ensures the time saved on creating data products and calibrating them from scratch to attain technically viable, scalable and sustainable services.

Overall, cloud computing empowers smaller teams to innovate, explore and experiment *("skunkworks experimentations")* and significant lowers costs and use open source technology which would not be reasonably priced (Schnase et al, 2014, pp 11).

4.4. Deployment & Adoption of aaS model

Adoption of cloud technologies means that platform and the systems will be simple to scale. Most of the cloud providers offer fully managed solutions which eliminates additional effort needed to maintain the infrastructure and services. This effort could be redirected in more meaningful tasks.

Furthermore, the adoption of cloud aaS model would ultimately enable utilities to jointly develop specific services in collaboration with operational public or private agencies. Development of such data native services in the cloud simplifies the interagency collaboration, service ownership and finally allocation of costs and profits (Schnase et al, 2014).

4.5. Organization & Governance

Operational focus in context of operational transparency and knowledge management is fundamental to identify knowledge gaps.

Evaluation process is vital to ensuring the actionable results. Aggregating data from multiple sources demands a dynamic KPIs dashboard with easy to understand visualizations. These dashboards are also deployable as cloud native solutions.

In the context of governance, security aspect is critical for working with sensitive data and such data needs to be properly secured and protected.

4.6. Public-private Partnerships

Public-private partnership is also seen as a critical enabler for successful implementation of capital-intensive projects in sustainable water resources management and sustainable development projects in general. An example of such successful private investment from venture capital firm is funding from XPV Water Partners in Canada to FATHOM based in Phoenix Arizona. FATHOM is a software-as-a-service cloud based geospatial data integration platform enabling water utilities faster adoption of smart water practices (Liner & Kenel, 2016).

Raising awareness about Industry 4.0 advantages and smart city initiatives, especially in the water utilities domain should be prioritized. Strategic policy studies should focus on further promoting the role of smart water practices in the context of implementation and viability of smart city initiatives and adoption of smart water practices should be granted adequate budget, resources and funding approvals.

Unfortunately, some of the largest cities in Europe are already facing major problems caused by delaying the capital projects in water utilities, such as aging infrastructure, old equipment, reductions in water consumption and droughts. These problems have direct impact on citizens and therefore citizens engagement is very important for raising awareness about the importance of adoption of smart water infrastructure and its immediate benefits to the community (Liner & Kenel, 2016).

To have sustainable water supply systems demands significant front-end investments which could also financially impact the end-users, citizens in various ways. Therefore, the benefits and advantages need to be fully transparent and understandable to all to ensure the willingness to pay potentially higher rates for provided and ensured long-term sustainability (Koo et. Al, 2015).

4.7. Continuous Innovation & Value-added Services

A famous article *"Will Data Analytics Change the Way We Deliver Water?"* appeared in 2013 in the American Water Works Association Journal (Neemann et al.) with a key message that analytics, data visualization and data driven utilities management are necessary pre-conditions for long-term systems optimization, operational efficiency and continuous innovation.

In order to continuously innovate and optimize the business processes in the water utilities we must set up first the fundamentals right.

Big Data and IoT technologies and applications are rapidly evolving leaving behind many options and a range of alternatives for its use as standalone solutions or integrated with numerous other services, free and paid-for versions. Most of these solutions provide support for open-source projects (e.g. *Apache Software Foundation*) with extensive support from the entire community of scientists and engineers. This type of support and collaboration is a powerful booster of innovation and continuous improvement in products and services for further value creation (Romero et al., 2017).

5. CONCLUSIONS

This article emphasized the importance of the nexus between Big Data and other smart technologies (IoT, cloud, ML&AI) to achieve sustainable water management. The objective of the article is to underline, *from a practitioner point of view,* key practical recommendations for water utilities in planning and implementing enterprise data strategy.

It was important to underline some current uses and existing application and use-cases where Big Data and IoT have been utilized for sustainable management of water supply and demand, however the true focus of the article is on the practical measures and success factors needed to achieve the data strategy set forth and successfully orchestrate the implementation of the overarching intelligent system for sustainable water management, including running the services in the maintainable and cost effective manner.

Smart data solutions are an amalgam of several key technologies and enablers, namely Big Data technologies, cloud services, connectivity and networking provided by IoT main functions; intelligence coming from prescriptive data analytics tools and machine learning algorithms. All these components are orchestrated or supported by the overarching infrastructure system and implemented via well-organized, cross-functional and agile operations teams.

REFERENCES

Accenture. (2019). *"Are you ready for what´s next?",* https://www.accenture.com/_acnmedia/PDF-94/Accenture-TechVision-2019-Tech-Trends-Report.pdf, [accessed 16.07.2019].

Adamala, S. (2017). An Overview of Big Data Applications in Water Resources Engineering. *Machine Learning Research, 2,* pp. 10-18.

Chalh et al. (2015). Big Data Open Platform for Water Resources Management. Ecole Mohammadia d'Ingenieurs, Marocco. IEEE.

Koo et al. (2015). Towards Sustainable Water Supply: Schematic Development of Big Data Collection Using Internet of Things (IoT). *Procedia Engineering, 118,* pp. 489-497.

Liner, B. & Kenel, P. (2016). Exciting future for big data solutions. *World Water, March/April 2016*, pp. 19-21.

Neeman et al. (2013). Will Data Analytics Change the Way We Deliver Water? *American Water Works Association, Nov. 2013*, pp. 25-27.

OSIsoft (2017). *Leveraging Big Data for Intelligent Water Utility Management,* https://www.osisoft.com/Presentations/Leveraging-Big-Data-for-Intelligent--Water-Utility-Management/, [accessed 7.7.2019].

Romero et al. (2017). Levering Big Data Tools and Technologies: Addressing the Challenges of the Water Quality Sector. *Sustainability, 9, 2160,* pp. 1-19.

Schnase et al. (2014). MERRA Analytical Services: Meeting the Big Data challenges of climate science through cloud-enabled Climate Analytics-as-a-Service. *Computers, Environment and Urban Systems*, pp. 1-14.

Shrivastava, P. & Jagadeesan, J. (2014). Cloud Based Software Platform for Big Data Analytics In Water Reservoir Level Forecasting. *IOSR Journal of Computer Engineering, 16 (2)*, pp. 17 – 20.

Song et al. (2012). Will environmental logistics be promoted by changing industrial structure? A quantitative analysis from 1978 to 2007 in China. *Supply Chain Management, 17 (1)*, pp. 5-14.

Song et al. (2015). Improving natural resource management and human health to ensure sustainable societal development based upon insights gained from working within "Big Data Environments". *J. Clean Prod., 94,* pp. 1-4.

Thompson, K. & Kadiyala, R. (2014). Leveraging Big Data to Improve Water System Operations. *Procedia Engineering, 89,* pp. 467 – 472.

United Nations. (2017). *Big Data and Sustainable development goals*, https://www.un.org/en/pdfs/Bigdata_SDGs_single_spread_2017.pdf, [accessed 23.05.2019].

Wang, X, & Sun, Z. (2013). The Design of Water Resources and Hydropower Cloud GIS Platform Based on Big Data. In F. Bian et al. (Eds.): *GRMSE 2013, Part II*, CCIS 399, pp. 313-322.