

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/319987216>

BIG DATA TECHNOLOGY IMPLEMENTATION IN MANAGING WATER RELATED DISASTER: NAHRIM'S EXPERIENCE

Conference Paper · August 2017

CITATIONS

3

READS

238

3 authors, including:



[Mohammad Fikry Abdullah](#)

National Hydraulic Research Institute of Malaysia

15 PUBLICATIONS 73 CITATIONS

[SEE PROFILE](#)



[Harlisa Zulkifli](#)

National Water Research Institute of Malaysia (NAHRIM)

9 PUBLICATIONS 31 CITATIONS

[SEE PROFILE](#)

BIG DATA TECHNOLOGY IMPLEMENTATION IN MANAGING WATER RELATED DISASTER: NAHRIM'S EXPERIENCE

MOHAMMAD FIKRY ABDULLAH⁽¹⁾, MARDHIAH IBRAHIM⁽²⁾ & HARLISA ZULKIFLI⁽³⁾

^(1,2) Water Resources and Climate Change Research Centre, National Hydraulic Research Institute of Malaysia (NAHRIM), Seri Kembangan, Malaysia,
fikry@nahrim.gov.my; mardhiah.iat@gmail.com

⁽³⁾ Information Management Division, National Hydraulic Research Institute of Malaysia (NAHRIM), Seri Kembangan, Malaysia,
harlisa@nahrim.gov.my

ABSTRACT

Data explosion related to water and environment have evolved in parallel with the development of data processing technologies that is done either automatically or manually. Water and environment data received should be used comprehensively and collectively through detailed analytic processes by group of experts specialised in water and environment domain. A good data with a good analysis technique can support decision making activities with an aid from experts' insight, knowledge and experience. The implementation of Big Data Analytics (BDA) to the data received, collected and analysed is not solely an Information Technology (IT) role, but it is a collaboration of the stakeholders such as Top Management team, Subject Matter Expert (SME) team and IT Technical team in determining the input, process and desired output of a BDA project. National Hydraulic Research Institute of Malaysia (NAHRIM) has been involved in the implementation of BDA technology by deliberating Climate Change domain as a case study in the project that consists of hydrologist, civil engineers, and researchers in the field of water resources, climate change and Information Technology. This paper is intended as an information sharing on NAHRIM experience in developing BDA project with the aim to encourage water related agencies or departments to implement BDA technology in water and environmental management. In this paper, we briefly share NAHRIM's first involvement and participation in BDA project and explain the mechanism and process of implementing our BDA project. Then, we discuss the common issues that arose during the execution of BDA projects and finally, we conclude this paper by presenting several suggestions to carry out BDA projects.

Keywords: Big data analytics; water management; disaster; climate change.

1 INTRODUCTION

"Data as an asset" is no longer a myth nowadays when the global evolution of data either in the quantitative or qualitative form carries not just monetary value, but also non-monetary value in almost every domain in the society. By understanding the value offered due to the explosion of data from the water domain perspective, those data could assist in preventing man-made disasters like overflowing rivers containing toxic waste and flooding, thus raising public awareness on water conservation and minimising the impacts of drought in arid regions (Cheung and Nuijten, 2014). Therefore, integration of exact tools with an accurate algorithm to find potential values from heterogeneous water datasets in a timely manner to support decision on water resilient is a real challenge. As we enter the age of Big Data, it is clear that we can take advantage of this phenomenon by engaging BDA in water domain. BDA projects are usually rigid and specific to the selected topic, but the results or outcomes generated and produced can be exploited and used by various parties depending on their level of understanding and critical thinking that are beyond the scope.

Like many terms used to refer to the rapidly evolving use of technologies and practices, there is no agreed definition of Big Data (Kitchin, 2013). However, researchers in this domain could have conceptualised Big Data by looking at the perspectives of product-oriented, process oriented or cognition-oriented (Ekbja et al., 2015). The product-oriented perspective highlights the novelty of Big Data largely in terms of the attributes of the data themselves, the process-oriented perspective seeks to push the frontiers of computing technologies in handling Big Data structures and relations and the cognition-oriented perspective conceptualises Big Data as something that exceeds human ability to comprehend and therefore required mediation through transdisciplinary work, technological infrastructures, statistical analyses and visualisation techniques to enhance interpretability. As defined by Gartner, Big Data is a high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation (Gartner, 2016). This definition covers all the perspectives; product-oriented, process-oriented and cognition-oriented, mentioned before. Nonetheless the definition, BDA is predominantly associated with two ideas: data storage and data analysis (Ward and Barker, 2013). The aim is to minimise hardware and processing costs and to verify the value of Big Data before committing significant resources (Khan et al., 2014).

Due to forces like population growth and climate change, the water cycle and water availability are in time of flux. Some of the ways they are changing are predictable, enabling regions to plan for the changes and take action but some of these changes are more difficult to predict, requiring regions to be flexible and responsive (The Aspen Institute, 2015). From this, it is shown that BDA in water related project requires involvement of IT Technical team to manage data storage component and Subject Matter Experts (SME) team to manage data analysis component.

Through effective use of data that is often available and already in place, BDA can provide various ways of achieving better water management, more adequate crisis management and even encouraging lower overall water consumption (Cheung and Nuijten, 2014). BDA technology manages water related disaster by monitoring and detecting hazards, mitigate their effects, and assist in relief efforts where the goal is ultimately to build resilience so that vulnerable communities and countries as a complex human ecosystem not just only 'bounce back' but also learn to adapt to maintain equilibrium in the face of natural hazards (Data-Pop Alliance, 2015).

Comprehending the value and potential offered through BDA technology, in 2015, National Hydraulic Research Institute of Malaysia (NAHRIM) started to implement BDA project focusing on climate change case study which has an impact on managing water related disaster in Malaysia. With the intention to encourage other agencies and departments to optimise the usage of data, NAHRIM would like to share the processes involved in N-HyDAA project through BDA Matrix Table and Gartner Analytics Ascending model which consist of descriptive, diagnostic, predictive, and prescriptive analytics. According to Lifescale Analytics (2015), descriptive analytics is the process of describing quantitatively what can be measured about a related domain. Diagnostic analytics look deeper into what has happened and seeks to understand why a problem or event of interest occurs. In predictive analytics, the analyst or SME will combine current observations into predictions of what will happen in the related domain by using predictive modelling and statistical techniques. The last analytic approach, prescriptive analytics will address decision making and efficiency as soon as a good measure of accuracy on the predictive algorithm is achieved, thus justify the prescriptive interventions.

The rest of this paper is organised as follows. Section 2 presents NAHRIM involvement in BDA project. Section 3 discusses the implementation of BDA project and explains the process of implementing BDA in water management. Section 3 discusses some common issues that arose during the execution of Big Data projects and finally, we conclude the paper by presenting several suggestions to carry out BDA projects.

2 NAHRIM INVOLVEMENT

NAHRIM as a research institute focusing on research and development (R&D) for water and environment, holds numerous water related and climate change data for Malaysia, either primary or secondary data, collected through sampling activities, modelling, simulation, and other R&D activities. Those data are being used for water and environment planning, to support decision making and to identify new potential R&D areas that can be diversified into various domain such as data projection analysis, climate change impact, sea level rise projection, hydro-climate and water resources related issues (Zulkifli et al., 2015).

Malaysia government has acknowledged the Big Data's potential by specifying BDA project as one of the national agendas. In 2013, the Prime Minister of Malaysia has officially announced the Malaysia BDA initiatives in which Malaysian Administrative Modernisation and Management Planning Unit (MAMPU) has been mandated to implement the BDA pilot projects in government agencies in 2015. This initiative is joined by Malaysia Digital Economy Corporation (MDEC) and MIMOS Berhad as a technology provider for this BDA projects. Four public agencies with five pilot projects were selected to develop Malaysia BDA Proof-of-Concept (POC) and NAHRIM was one of them. NAHRIM's BDA POC project titled "Projected Hydro-climate Data Analysis and Visualisation for Potential Drought and Flood Events in Peninsular Malaysia" was developed to assist NAHRIM in visualising and analysing almost 1450 simulation-years of projected hydro-climate data for Peninsular Malaysia based on 3888 grids. Other projects were "Islamist Extremist Amongst Malaysians" by Department of Islamic Development Malaysia (JAKIM), "Flood Knowledge Base from a Combination Sensor Data and Social Media" by Department of Irrigation and Drainage (DID), "Data Analytics to Analyse and Build Fiscal Economic Models" and "Sentiment Analysis on Cost of Living gathering from Social Media" by Ministry of Finance (MOF) (Dzazali, 2014).

Modules involved in this BDA POC were Drought, Drought & Temperature, Rainfall & Runoff, Storm Centre and Streamflow. We successfully proved the concept of implementing BDA using NAHRIM hydro-climate datasets, comprises of time-series historical, current and projected data, acquired through the modelling of historical data. We were able to visualise 3,888 grids for Peninsular Malaysia, detected extreme rainfall and runoff projection data for 90 years, identified flood flow for 11 river basins and 12 states in Peninsular Malaysia, and traced drought episodes from weekly to annual rainfall data for 90 years.

NAHRIM BDA POC project was completed in September 2015, and later in August 2016, NAHRIM BDA project has turned to full project that catered three more modules; Climate Change Factor, Water Stress Index, and Water Stress Index Simulation. With the automated and systematic BDA project, later called NAHRIM Hydro-Climate Data Analysis Accelerator (N-HyDAA) system, reduces the current manual process

by humans and improves the quality and visual of data hence saves more time and cost. It will also have benefited in discovering the vast potential data, sharing information, and producing more effective decision making in timely manner.

To ensure the successful project completion, there are two teams involved in this BDA project and they are IT Technical Team and SME Team. IT Technical Team is responsible to provide technology consists of hardware, software and customisation services in developing the system. Meanwhile, SME Team for this project is the backbone or the brain that responsible for the solution, methodology, algorithm regarding of the chosen domain of the business case. SME Team for the project is a mixture of various background of educations and experiences, composed of hydrologist, engineers and IT researchers of NAHRIM's Water Resources and Climate Change Research Centre and Information Management Division.

3 NAHRIM BDA PROCESS

Our BDA project implementation in managing water related disaster has been through a systematic and thoroughly process to ensure the result and outcome produced from the project will provide a great impact on managing water issues. Viewing from ICT standpoint, BDA implementation is focusing on resolving dedicated business issues by integrating main key players in the proposed business project. The key players in this context referring to SME who are well-versed and expert about the domain choose for the project. Opinions from SME will be the core point on how the project should be understand and developed to cater the issues arise from the business.

SME of NAHRIM's BDA project play their part in every phase of the development process. Hydrologists and engineers focusing on the arithmetic calculation, parameters selection, analysis and data source while IT researchers deal with the system development, data management and visualisation. This collaboration ensures the project developed met the purpose and objective with results and outcomes that should benefit the users.

The complete process of implementing NAHRIM's BDA project comprises of six phases; BDA Matrix Process, Requirement Analysis, System Architecture, Data Flow Diagram, System Development and Deployment. Details on every phase of the process are explained in Table 1. In the next section, we will go in depth on three crucial phases of this process; BDA Matric Process, System Architecture and Data Flow Diagram.

Table 1. Detailed process of NAHRIM's BDA project.

No	Process	Explanation
1	BDA Matric Process	a. Identifying Business Direction and Business Problem Definition for proposed domain and project.
2	Requirement Analysis	a. Develop Requirement Analysis Book to highlight scope of work for the proposed project. b. Itemised scope of work based on importance of features (Must Have, Should Have, Could Have and Won't Have).
3	System Architecture	a. Designing System Architecture of the project based on Requirement Analysis Book.
4	Data Flow Diagram	a. Designing Data Flow Diagram of the system to determine the input data, processing and output data.
5	System Development	a. Development of BDA System including system testing.
6	Deployment	a. Deploy the System for user practice.

3.1 BDA matric process

NAHRIM's BDA project started by defining Business Direction and Business Problem Definition which are the main activities that guide the whole development of the project based on MAMPU assessment. In the Business Direction activity, the tasks were to identify Project Objectives of the proposed project based on identified domain or business case. The Key Measures as indicator that reflect the performance of the project also has been identified. In the Business Problem Definition activity, the tasks were to identify Business Function/Problem Area, Business Challenges, Problem Statements, Impact of The Problem, Business Questions and Data Sets Usage. The purpose of Business Function/Problem Area task is to identify the specific domain or business area for the project. The Business Challenges task is to explain the current challenges faced that must be tackled. Statements that will guide us through the implementation of the project is created in Problem Statements task as to ensure the project is on-track with the business direction. Impacts or issues based on business challenges were listed in Impact of the Problem task. In Business Questions task, potential question that could be asked from the project was prognosed and finally datasets required in the project were identified in Data Sets Usage task.

Table 2 shows the essence of NAHRIM's BDA project over NAHRIM requirement analysis. From Table 2, it shows hydrologist and engineers are the main pillar of this project that lead to the usage and direction of the project to cater issues on the Climate Change domain.

Table 2. NAHRIM's BDA matrix table.

Activity		Task	Role
Business Direction	1. Project Objectives	- Domain Climate Change has been identified in this project, by focusing on:- a. Water Security (Water Stress & Water Availability); b. Weather Extreme Events (Floods & Droughts); c. Disaster Risk Reduction; - To prepare early data & information of potential water related disaster to related Agency for:- a. Possible issues arise in identified domain; b. Risk management planning; - To ensure an effective & efficient asset and resource management in risk management plan and issues in related domain; - To avoid, reduce and safe guard a high-risk area due to water related disaster and climate change; - To identify and reduce lost (life, properties and ecosystem) due to disaster happened. - Number of lives that can be saved from hydrometeorology disasters; - Number of loss that can be saved from hydrometeorology disasters;	NAHRIM's SME
	2. Key Measures	- Number of risk management plan that has been implemented; - Size of potential area affected from hydrometeorology disaster; - Amount of area, infrastructure, life, properties and ecosystem that can be save from hydrometeorology disasters.	NAHRIM's SME
Business Problem Definition	1. Business Function/ Problem Area	- Water related disaster risk management & climate change impacts; - A systematic management and decision making for potential disaster related to hydrometeorology.	NAHRIM's SME
	2. Business Challenges	To highlight, use and disseminate disaster information related to climate change extensively in Malaysia such as flood, drought, sea level rise etc.	NAHRIM's SME
	3. Problem Statements	Citizens concern on hydrometeorological related disasters (pre & post), especially flood which has a high frequency and a large magnitude. Similarly, the occurrence of flash floods in major cities across the country in Malaysia where the cost to fix is very high.	NAHRIM's SME
	4. Impact of the Problem	- Loss of life; - Loss of properties (Government, Business and People); - Loss of incomes and jobs; - Loss of ecosystem. - How many loss of life in the events? - How much loss from the events (Government, Business, and People)?	NAHRIM's SME
	5. Business Questions	- What are the preparation taken by the Government? - What are the policies would be taken by the Government to face potential disaster? - Hydrometeorology and weather data; - Climate change projection data; - Landuse data; - Department of Stastical's data; - Catchment data;	NAHRIM's SME
	6. Data Sets Usage	- Waterbodies & Water Treatment Plan data; - Satellite data; - Water Intake data; - Infrastructure & Water Resource Facilities data (Groundwater & Surface); - Socio economy data.	NAHRIM's SME

3.2 System architecture

N-HyDAA is a web-based information system that uses internet web technologies to deliver information and services. In the N-HyDAA project, the design of System Architecture and Data Flow Diagram are prepared based on the Requirement Book gathered during Requirement Analysis phase. Figure 1 shows NAHRIM System Architecture that represents the IT and non-IT components identified and involved in the project. IT components deal with data management, data authorization, data authentication, data storage, operating system, web server and ICT infrastructure to run the system. The non-IT components in NAHRIM's BDA project involve modules that required Data Accelerator for accelerating data processing and which are:

- i. Drought/ Storm Center/ Streamflow/ Rainfall/ Runoff;
- ii. Climate Change Factor;
- iii. Water Stress Index;
- iv. WSI Simulation.

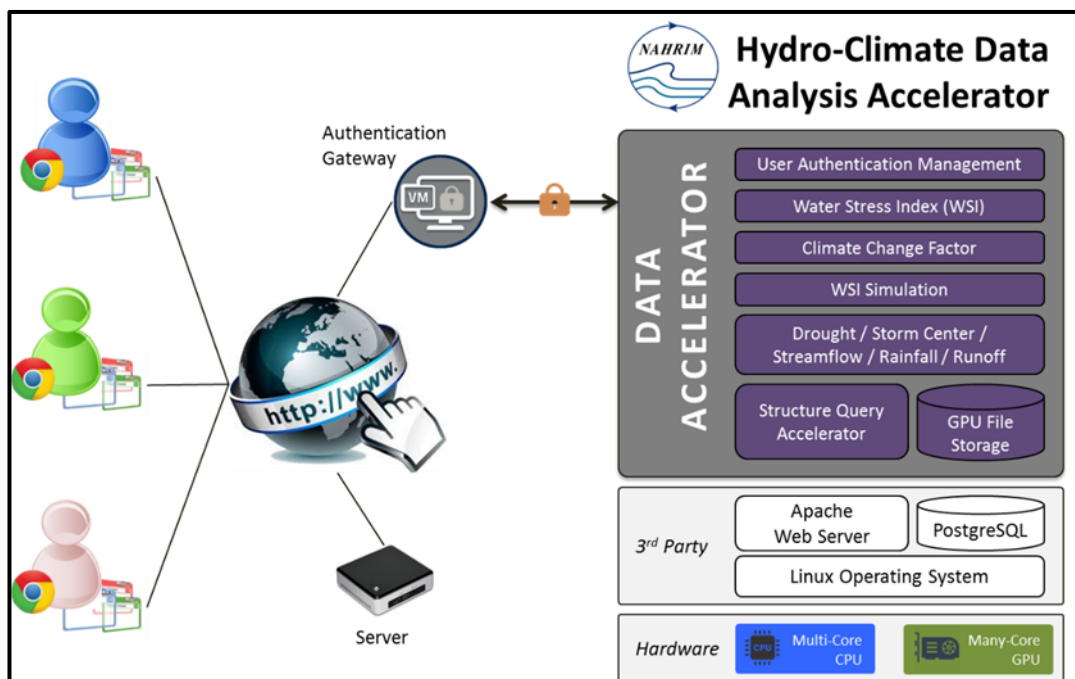


Figure 1. N-HyDAA system architecture.

3.3 Data Flow Diagram

Data Flow Diagram showed in Figure 2 explains the overall flow of data for BDA project that are required in the project. We divided the flow into 5 layers:

- i. Data Acquisition;
- ii. Data Cleaning & Integration;
- iii. Data Repository;
- iv. Analytics;
- v. Presentation.

Each data layer performs distinctive function. Data Acquisition layer consists of components to gather raw, pre-process or unclear input data from all sources, such as rainfall, runoff, streamflow, and so forth. Data Cleansing & Integration layer consists of integration and process components in amending or removing data flow from the sources to the data repository layer in the architecture. Data Repository layer stores data in a columnar storage format for accelerating data parallel processing and improving query performance and extensibility. In Analytics layer, the queried data will be extracted from the repository to make it easier for users to perform big query processing and what-if analysis. Presentation layer gives access to different set of users where it consumes the data through web pages that are defined in the reporting tool (National Hydraulic Research Institute of Malaysia, 2016).

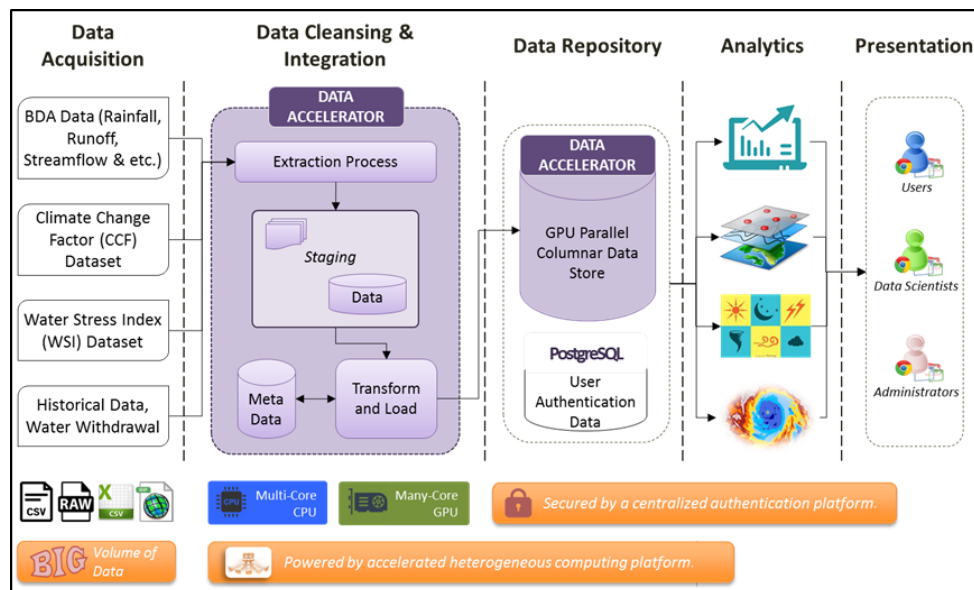


Figure 2. N-HyDAA Data Flow Diagram.

4 ISSUES AND SUGGESTIONS

Implementing BDA technology in water domain does imposed several issues and misconceptions in the initial stage. Below are the lists of misconceptions arose and our suggestions to carry out the BDA projects.

4.1 The thought of completing the V's

In the case of NAHRIM's BDA project, the first issue raised was the thought of complying all the Vs (Volume, Velocity, Variety, Veracity, Visualisation, and so forth) for our datasets and had to be fulfilled mandatorily. NAHRIM's assumption was that BDA project will not be successfully developed if our data did not fall in every criterion of the Vs for BDA. But throughout this project, NAHRIM decided to focus on optimising and exploring our 10 billion hydro-climate simulated projected data in structured format and by the result of the analysis, we were proven wrong as in Volume, Velocity and Visualisation are more than enough to give us the outcomes that we required. Understanding our own data is the key-successful factor of BDA implementation. Based on the data, organisation should know what are the expected result required and the processes involved including identifying the data to be used, type of data, technology, technique to store the data, method to process the data and how to integrate them all to gain insights and depth to solve real problems.

4.2 The role of IT in BDA

Most of the domain owners believe that BDA elements comprise of technical aspects only. The truth is, business owners should be involved at all events of BDA development to solve problems related to their business and field. In NAHRIM case study, we use BDA as a tool to develop an application that can centralised the information and analysis on web application. NAHRIM took advantages of the current BDA technology through outsourcing mechanism by technology provider while NAHRIM SME team focused on providing advices and contents that consist of data, methodology and algorithm in the analysis phase of the project. Our BDA project will not be successful without good cooperation among researchers and engineers in providing the data and it is not subject to technological problems alone.

4.3 The dispersion of data

Unstructured data is one of the data type that revolved around us most rapidly and increasingly. This data type mostly is random and not modelled. Therefore, there is an assumption that only unstructured data is used for BDA. As a data provider, NAHRIM is no exception into thinking that combination of unstructured and structured data is a must to be used for BDA. The initial form of data we collect such as rainfall, runoff and streamflow, are structured data used for hydro-climate projection for the years 2010-2099. These structured data are then analysed through algorithm accelerated by technology provided by IT Technical team. Through this data processing, we able to produce data output such as drought visualization by state, month and year, rainfall patterns, magnitude of storms and so on. In this case, NAHRIM do not intent to use unstructured data since NAHRIM would like to focus on optimising the current structured data. This experience can refute the notion that combination of structured and unstructured is a must to perform BDA.

4.4 The relevancy of BDA

Most BDA use cluttered data, and not all data is valuable for analysation. Because of this criterion, the process employed to analyse the data obtained are often time-consuming starting from the data collection,

data processing and data visualisation. Hence, there is a perception that the result of BDA is nothing else than the data visualisation on the dashboard. The right thought should be “what are the accurate action that business owners and organisations need to take with the analysed data?” Based on NAHRIM experience in applying BDA, we only provide data for analysis. The information that has been generated through this technology is hopefully can help the ministries, government departments and agencies such as Ministry of Natural Resources and Environment (NRE), Ministry of Energy, Green Technology and Water (KeTTHA), Ministry of Agriculture and Agro-based Industry (MOA), Department of Public Works (JKR), Department of Irrigation and Drainage (DID), state governments and private sector to make a strategic planning and immediate action in a holistic manner that lead to sustainable development and climate resilience such as water management issues, drought and flood.

5 CONCLUSIONS

As the aim of Big Data Analytics is to “turn data into insights” for better decision making (Dargam et al., 2015), NAHRIM has responded to Malaysia government initiative to implement BDA technology in managing our datasets that related to water and its environment. We shared briefly NAHRIM involvement in Big Data projects and explained the process of implementing Big Data Analytics which we have completed by identifying our own BDA Matrix Process, Requirement Analysis, System Architecture, Data Flow Diagram, System Development and Deployment. We also have discussed some of the common issues that we have encountered during the execution of Big Data projects and finally, we conclude the paper by presenting several suggestions to carry out big data projects. Throughout this NAHRIM's BDA project, it shows and indicates that with a correct data, people and technology, BDA concept can be implemented especially in Government sector despite that there are challenges and confusions towards understanding the BDA concept itself.

ACKNOWLEDGEMENTS

This project was supported by Malaysian Administrative Modernisation and Management Planning Unit (MAMPU), Malaysia Digital Economy Corporation (MDEC) and MIMOS Berhad and we are thankful to our team members from NAHRIM as well who provided expertise that greatly assisted the implementation of this project.

REFERENCES

- Cheung, C. & Nuijten, M. (2014). Big Data and The Future of Water Management [PDF Document]. Available: <https://www.rvo.nl/sites/default/files/2014/05/Big%20Data%20and%20the%20Future%20of%20Water%20Management.pdf> [Accessed 01/02/2017].
- Dargam, F.C.C., Zaraté, P., Ribeiro, R. & Liu, S. (2015). The Role of Decision Making in the Big Data Era. In *1st EWG-DSS International Conference on Decision Support System Technology on Big Data Analytics for Decision Making (ICDSST 2015)*. Retrieve from http://oatao.univ-toulouse.fr/15327/1/dargam_15327.pdf.
- Data-Pop Alliance. (2015). Big data for climate change and disaster resilience: Realising the benefits for developing countries [PDF Document]. Available: <http://datapopalliance.org/wp-content/uploads/2015/11/Big-Data-for-Resilience-2015-Report.pdf> [Accessed 01/02/2017].
- Dzazali, S. (2014). Public Sector Big Data Analytics Initiative: Malaysia's Perspective. [PowerPoint slides]. Available: <http://www.mampu.gov.my/ms/penerbitan-mampu/send/100-forum-asean-cio-2014/275-1-keynote-mampu> [Accessed 01/02/2017].
- Ekbja, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V.R., Tsou, A., Weingart, S. & Sugimoto, C.R. (2015). Big Data, Bigger Dilemmas: A Critical Review. *Journal of the Association for Information Science and Technology*, 66 (8), 1523-1545.
- Kitchin, R. (2013). Big Data and Human Geography: Opportunities, Challenges and Risks. *Dialogues in human geography*, 3(3), 262-267.
- Gartner. (2017). Big Data. Available: <http://www.gartner.com/it-glossary/big-data> [Accessed 01/02/2017].
- Lifescale Analytics. (2015). Descriptive to Prescriptive Analysis: Accelerating Business Insights with Data Analytics [PDF Document]. Available: http://www.lifescaleanalytics.com/~Isahero9/application/files/7114/3187/3188/leadbrief_descripprescrip_web.pdf [Accessed 01/02/2017].
- National Hydraulic Research Institute of Malaysia (NAHRIM). (2016). *NAHRIM Hydro-Climate Data Analysis Accelerator (N-HyDAA)* (PJM-16366_N-HyDAA_Final_Report). Seri Kembangan, Selangor: Malaysia.
- The Aspen Institute. (2015). Data intelligence for 21st century water management [PDF Document]. Available: https://assets.aspeninstitute.org/content/uploads/files/content/docs/pubs/2015_Water_Forum_Report_FIN_AL.pdf [Accessed 01/02/2017].
- Ward, J.S. & Barker, A. (2013). Undefined by Data: A Survey of Big Data Definitions. *arXiv preprint arXiv:1309.5821*.
- Zulkifli, H., Kadir, R.A. & Nayan, N.M. (2015). Initial User Requirement Analysis for Waterbodies Data Visualization. *International Visual Informatics Conference*, 89-98.