Big Data Analytics Framework for Natural Disaster Management in Malaysia

Mohammad Fikry Abdullah[1], Mardhiah Ibrahim[1] and Harlisa Zulkifli[2]
[1]Water Resources and Climate Change Research Centre,
National Hydraulic Research Institute of Malaysia (NAHRIM), Malaysia
[2]Information Management Division, National Hydraulic Research Institute of Malaysia (NAHRIM), Malaysia

Keywords: Big Data Analytics, Disaster Management, Decision Making, Hydroclimate, Government, Public Sector, Big Data Framework.

Abstract: Decision making in natural disaster management has its own challenge that needs to be tackled. In times of disaster, government as a response organisation must conduct timely and accurate decisions to ensure rapid assistance and effective recovery for the victim involved can be conducted. The aim of this paper is to embark strategic decision making in government concerning to disaster management through Big Data Analytics (BDA) approach. BDA technology is integrated as a solution to manage, utilise, maximise, and expose insight of climate change data for dealing water related natural disaster. NAHRIM as a government agency responsible in conducting research on water and its environment proposed a BDA framework for natural disaster management using NAHRIM historical and simulated projected hydroclimate datasets. The objective of developing this framework is to assist the government in making decisions concerning disaster management by fully utilised NAHRIM datasets. The BDA framework that consists of three stages; Data Acquisition, Data Computation, and Data Interpretation and seven layers; Data Source, Data Management, Analysis, Data Visualisation, Disaster Management, and Decision is hoped to give impact in prevention, mitigation, preparation, adaptation, response and recovery of water related natural disasters.

1 INTRODUCTION

Information Technology (IT) plays a pivotal role as integrator in the disaster management system, particularly in tasks of managing the disaster data. However, the data acquired during disaster events such as floods, landslides, mud, soil erosion and so forth often presented in a large volume from heterogeneous sources, thus the data management process for natural disasters is a challenge to be tackled. Issues of heterogeneity of reliable data sources such as from sensors, social media, and others during the period of crises and disasters requires advanced and systematic analysis approach to execute disaster management plan. In times of disasters, government and authorities who act as decision makers are responsible to take action that demands immediate and fast relief activities in the devastated area. But the quality of the decision depends on the quality of data and information obtained (Emmanouil and Nikolaos, 2015). Hence, the data received in times of disaster need to be analysed thoroughly as an input to decision makers to made precise decisions in a limited and ad-hoc time manner.

Disaster management can be planned and organised intelligently if the data related to disaster are being managed efficiently and effectively before the disaster happened. This plan can be achieved by analysing existing historical and projected data that can be accessed daily such as rainfall, temperature, drought, and streamflow data to predict future disaster events. From the projected and prediction analysis, a holistic and comprehensive mitigation, control, and prevention can be drafted in advance, hence the risk of injuries, health impacts, property damages, loss of lives and services, social and economic disruptions and environmental damages can be reduced, minimised and avoided.

National Hydraulic Research Institute of Malaysia (NAHRIM) as a government agency responsible in conducting research on water and its environment are called to formulate a framework for natural disaster management using hydroclimate data acquired by NAHRIM. Considering the available data are numerous, the Big Data Analytics (BDA) technology

Abdullah, M., Ibrahim, M. and Zulkifli, H.
Big Data Analytics Framework for Natural Disaster Management in Malaysia.
DOI: 10.5220/0006367204060411
In *Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security (IoTBDS 2017)*, pages 406-411
ISBN: 978-989-758-245-5

is used to analyse existing data to gain meaningful insights.

In this paper, we reflect the relationship between big data and disaster management in the government sector and proposed our big data framework in managing natural disaster. The rest of this paper is organised as follows. Section 2 presents the background of this study. The following section discusses the proposed framework of NAHRIM's big data framework in handling disaster management and in Section 4, the conclusion is presented.

2 BACKGROUND STUDY

In Malaysia, there were 76 disasters has been recorded in the period of 1965 to 2016. The type of disasters including wildfire, storm, landslide, mudflows, epidemic, tsunami, drought and more than half of the disasters were flood related hazard (Amin, 2016). In this context, government agencies as the response organisation which by law are obligated to prepare for and manage such crises. Disaster management is the management of the risks and consequences of a disaster in order to reduce or to avoid potential losses from hazards (Othman and Beydoun, 2013). Disaster management is crucial in regard to provide rapid assistance and effective recovery for the victim involved.

In disaster management, there are many activities that involve decision making under the time pressure. However, decision making in government usually takes much longer and is conducted through consultation and mutual consent of a large number of diverse actors, including officials, interest groups, and ordinary citizens (Kim et al., 2014). This may be due to standard operating procedure, top management discussion and so forth. How those decisions are made is important as the result of the decision-making must compromise the citizens and country accepted standards. Timely decision-making to direct and coordinate the activities of other people is important to achieve disasters management goals (Othman and Beydoun, 2013).

Natural disaster is categorised as external risk, and it cannot be typically reduced or avoided through conventional approaches as it lies largely outside human control. It requires a different analytic approach either because their probability of occurrence is very low or it is difficult to foresee by normal strategy processes (Kaplan and Mikes, 2012). In response to great pressure for the government to provide service delivery within time and budget constraint, Big Data Analytics (BDA) may be one of the possible solutions to consider as government holds a great amount of earth-related data owned by the public agencies and departments. This is based on the assumptions that BDA exploitation can help local government to allocate resources where they will have the biggest impact and restructure services in such a way that early prevention is prioritised to avoid the need for more expensive interventions (Malomo and Sena, 2016).

BDA provides solution to support the management and analysis of multidimensionality, volume, complexity, and variety earth-related datasets and support scientific analysis process through parallel solutions (Kaplan and Mikes, 2012) as the focus of climate science is more about understanding than predicting (Faghmous and Kumar, 2014). In short, BDA deals with collection, management, and transformation of a large collection of digital data, which come in diverse forms, in order to reduce uncertainty in decision making (Ali et al., 2016).

However, there are challenges that must be overcome in order to integrate BDA in the government sector. The wide technology gap between industrial applications and decision makers is one of them (Tekiner and Keane, 2013). Decision makers need to understand the data and technologies better in order to extract information to aid strategic decision making. Besides, the role of Subject Matter Expert (SME) is also crucial as they understood the domain well, to support high level decision making process, with ICT as enabler. Data sharing among different public agencies and departments also remains a challenge (Kim et al., 2014). The data would have to be obtained not only from heterogeneous channels, but also involve data transfer across public agencies and departments. Without proper integration, data across public agencies and departments are maintained in silo (Ali et al., 2016). As suggested by Molomo and Sena (2016), the general legal framework has to be developed to facilitates data sharing among local authorities. Without coordination and structuring framework, there is likely to be much overlap amongst applications, duplication in stored information and confusion around the responsibilities of each business unit and application (Tekiner and Keane, 2013).

Malaysia government has acknowledged the big data's potential by specifying BDA project as one of the national agenda. The strategic collaboration between Malaysian Administrative Modernisation and Management Planning Unit (MAMPU), Malaysia Digital Economy Corporation (MDEC) and MIMOS Berhad has been agreed through BDA-Digital Government Open Innovation Network (BDA-DGOIN) in 2015. Four public agencies with five pilot projects were selected to develop Malaysia BDA Proof of Concept (POC), and the 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) and NAHRIM with the titled "Visualizing 90 Years of Projected Rainfall corresponding runoff after-effects based on river basin Malaysian Map". Figure 1 are some examples of NAHRIM BDA POC interface which have been developed.

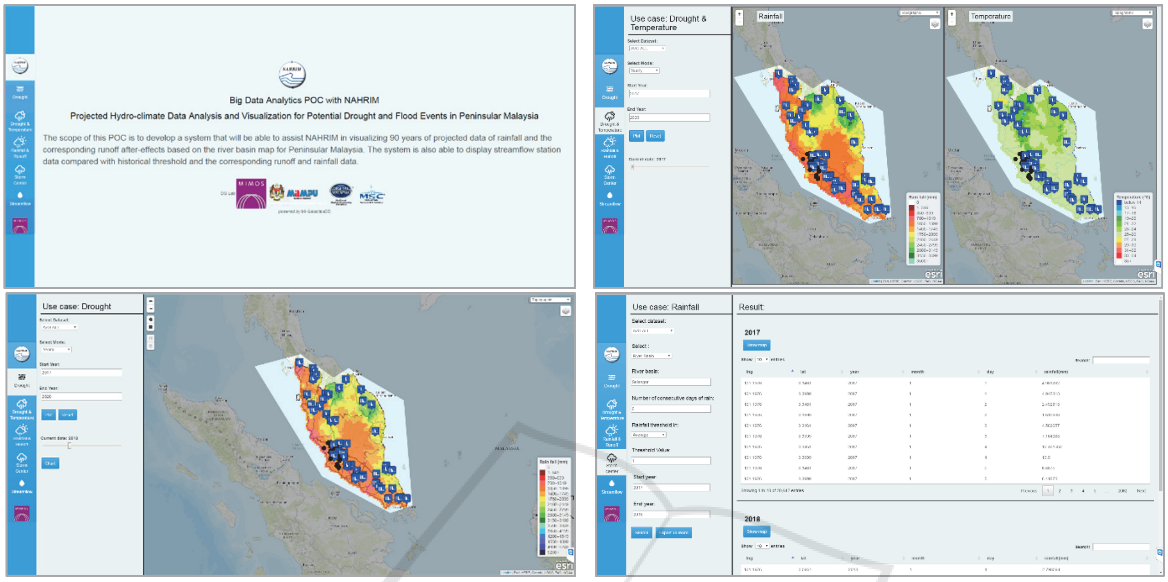

Figure 1: Examples of NAHRIM BDA POC interface.

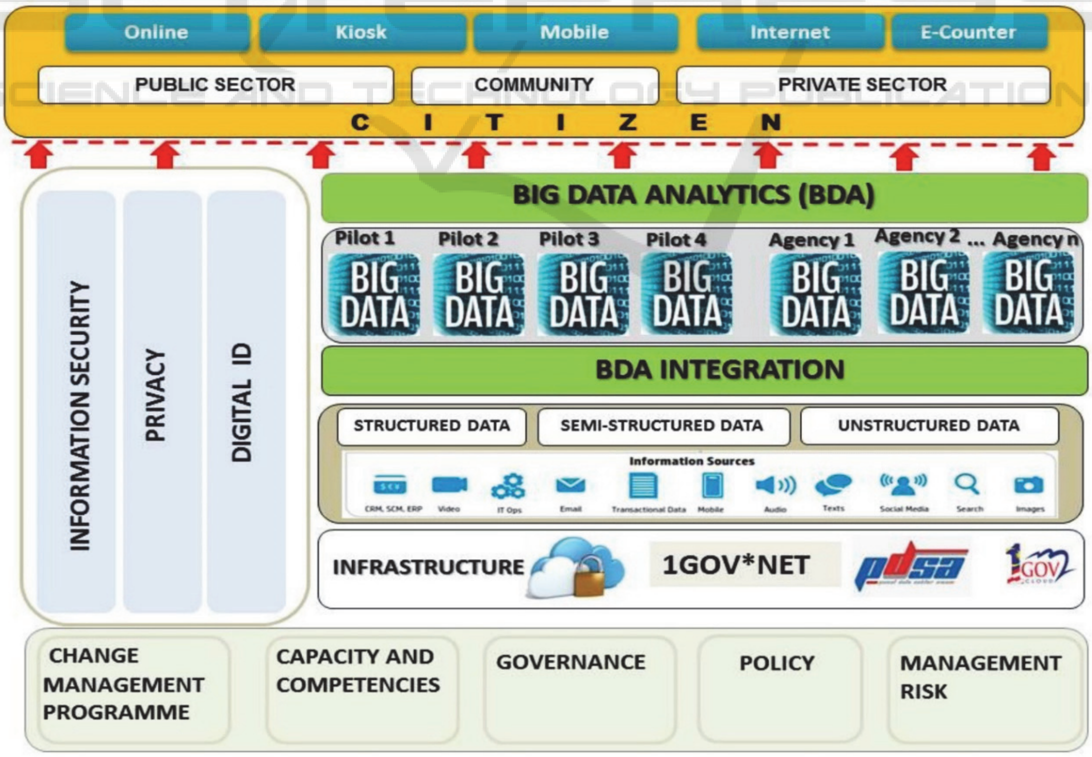

Figure 2: Malaysia BDA Framework formulated by MAMPU.

Figure 2 is the national fundamental of Big Data Analytics framework formulated by MAMPU (2014), designed to specifically enabled BDA in public sector. MAMPU's BDA framework has been the focal reference for all BDAs pilot projects as mentioned earlier. We also successfully proved the concept of implementing BDA using NAHRIM hydroclimate 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.

Seeing the great potential in this analysed data, we realised that there herein lies the opportunity for developing new big data framework to assist the government in making decisions concerning disaster management. By adapting and considering few attributes from MAMPU Malaysia BDA framework, we proposed a new big data framework in handling disaster management in Malaysia.

3 PROPOSED BDA FRAMEWORK

As depicted in the Figure 3, NAHRIM Big Data Framework for Disaster Management consists of three stages; Data Acquisition, Data Computation, and Data Interpretation.

Data Acquisition stage consists of a layer, that is Data Source. The aim of this stage is to aggregate information in a digital form for further storage and analysis because of the data may come from a diverse set of sources (Emmanouil and Nikalaos, 2015). At this stage, data are obtained from a historical-based modelling process, using high performance computing environment, consist of large volume of historical and projected hydroclimate data. The projected data were calculated, modelled, and simulated from raw datasets such as rainfall, runoff, temperature and streamflow and are mapped onto time-series format which are; yearly, monthly, weekly or daily projected data. The historical data on the contrary are the observed and simulated historical data that is also stored with respect to time series. These data can be presented disparately in spatial, non-spatial, structured, unstructured, and semi-structured data format.

Once the data were acquired, the significance computational technique has to be applied to the data sources. Second stage, the Data Computation Stage consists of three layers; Data Management, Analysis, and Data Visualisation. In Data Management layer, the projected and historical data that were collected will undergo the data cleaning process. Data cleaning is the process where incomplete and unreasonable data are identified (Hashem et al., 2015). These datasets will be filtered to the specific categorisation using a specific extraction method so that the semantics and correlations of data can be obtained.

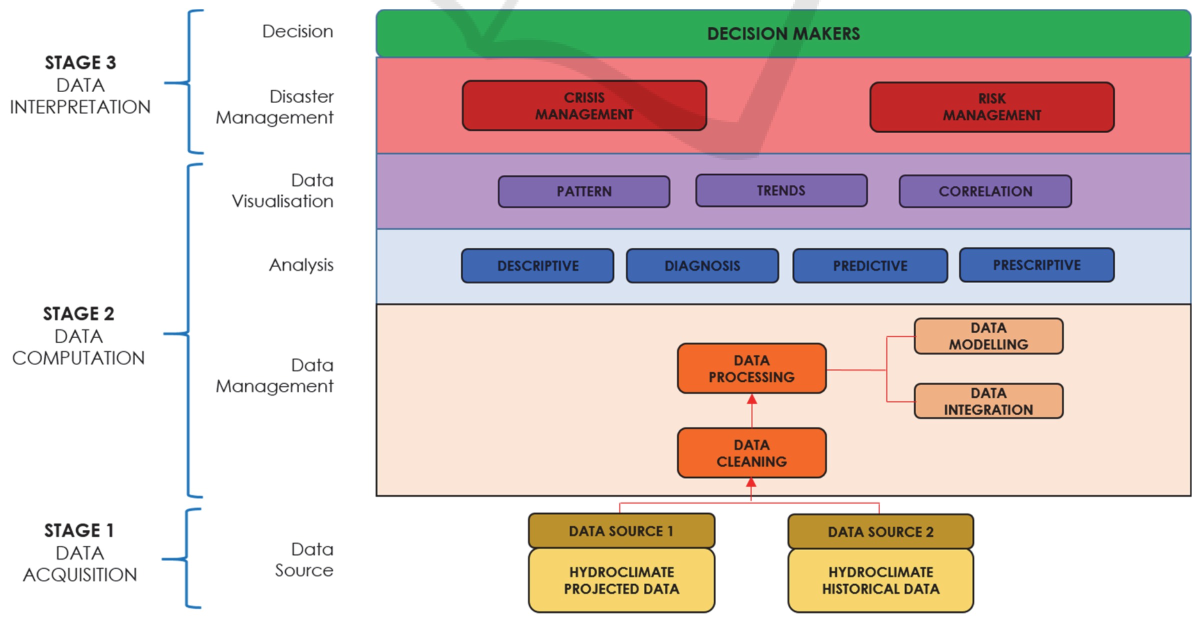

Figure 3: NAHRIM Big Data Framework for Natural Disaster Management.

Then, the identified dataset is modelled and integrated in the data processing phase. The data modelling involves programming model that implements abstraction application logic and facilitates the data analysis applications (Hu et al., 2014). Data integration on the other hand helps the data analyst resolve heterogeneities in data structure and semantics as this heterogeneity resolution leads to integrated data that is uniformly interpretable within a community (Jagadish et al., 2014).

The next and most important stage is the Analysis layer. The aim of data analysis is to extract as much information as possible that is pertinent to the subject under consideration (Emmanouil and Nikolaos, 2015). According to Lifescale Analytics (2015) data analysis are classified into four different analytical approaches that can be used to solve a business cases or problem, or set of problems; descriptive, diagnostic, predictive and prescriptive analytics.

Descriptive analytics is the process of describing quantitatively what can be measured about a related domain. In our case, hydroclimate historical data will be fully utilised to quantify, track and report what might have previously been occurred and how things are going in the disaster management domain. Diagnostic analytics look deeper into what has happened and seeks to understand why a problem or event of interest occurs. Based on the processed hydroclimate data that has been obtained, the root cause of the problems will be uncovered as the set of data are converged with explanations. In predictive analytics, the analyst or SME will focus on answering the "What will happen next?" question. They 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, and thus justify the prescriptive interventions. This will not only give a credible explanation for how this disaster is more likely to return or reoccurred, but data analyst will also understand how predictable the disaster occurrence is.

The last phase in Data Computation Stage is the Visualisation phase. In every Big Data framework, visualisation phase is considered vital as it allows business users to mash up disparate data sources to create custom analytical views (Wang et al., 2015). In spite of the tremendous advances made in computational analysis, there remain many patterns that humans can easily detect, but computer algorithms have a difficult time finding (Jagadish et al., 2014). This is how visualisation plays a key role of the discovery process in big data framework. The more effective the data visualisation is, the higher the chances to recognise the potential patterns, trends and correlations between the analysed hydroclimate data.

The final stage is Data Interpretation stage which consists of Disaster Management and Decision Layers. At this stage, the SME and decision makers plays a critical action on understanding the data and information to make strategic, rational and relevant decisions based on the insights obtained from the presented analysis. Data Interpretation required knowledge and experience from domain experts such as hydrologist, climatologist, scientist, etc. to help further clarification on the analysis prior to make the decision. Disasters can be predicted, and wherever possible, can be avoided through mitigation, or adaptation if it really happens, if the decision is made based on quality and accurate data from the BDA. Crisis can be averted by early interventions from early warning that is gained from the insights. Assessment and risk management controls can be taken into action to reduce catastrophic impact of disasters based on the results of the analysis. Because of the fact that disaster management is characterised by complexity, urgency, and uncertainty, it is crucial for participating organisations to have a fast though smooth and effective decision-making process (Kapucu and Garayev, 2011).

4 CONCLUSIONS

There is a need to embark strategic decision making in government related to disaster management and this paper aims to fill the space. Effective decision-making in the government sector is possible when relevant participants receive timely and accurate information that is thoroughly analysed and filtered. Exploiting government data to their full potential by leveraging the benefits offered by Big Data Analytics will give impact in prevention, mitigation, preparation, adaptation, response and recovery of disasters. NAHRIM Big Data Framework for Natural Disaster Management is proposed to analysed hydroclimate data acquired by NAHRIM to support Malaysian government to effectively coordinate disaster and relief. A process of analysing hydroclimate data requires tools to speed up the process of accelerating data computation and this is the fundamental of BDA that support disaster management in Malaysia. The framework presented also attempts us to pave the way for future work by integrating method to mine, store, process and analyse and stream data gained from multiple sources such as social media and sensors. In a nutshell, this framework also hoped to act as guideline to help other government agencies and departments for creating their own data-driven decision making.

ACKNOWLEDGEMENTS

This project was supported by MAMPU, Malaysia Digitial Economy Coporation (MDEC) and MIMOS and we are thankful to our team members from NAHRIM as well who provided expertise that greatly assisted the implementation of this project.

REFERENCES

Alhawari, S., Karadsheh, L., Talet, A.N. and Mansour, E., 2012. Knowledge-based risk management framework for information technology project. *International Journal of Information Management*, *32*(1), pp.50-65.

Ali, R.H.R.M., Mohamad, R. and Sudin, S., 2016, August. A proposed framework of big data readiness in public sectors. In F.A.A. Nifa, M.N.M. Nawi and A. Hussain eds.,, *AIP Conference Proceedings* (Vol. 1761, No. 1, p. 020089). AIP Publishing.

Amin, M.Z.M. (2016). Applying Big Data Analytics (BDA) to Diagnose Hydrometeorological Related Risk Due To Climate Change. *GeoSmart Asia*, [online] Available at: http://geosmartasia.org/presentation/applying-big-data-analytics-BDA-to-diagnose-hydro-meteorological-related-risk-due-to-climate-change.pdf [Accessed 1 November 2016]

Emmanouil, D. and Nikolaos, D., Big data analytics in prevention, preparedness, response and recovery in crisis and disaster management. In *The 18th International Conference on Circuits, Systems, Communications and Computers (CSCC 2015), Recent Advances in Computer Engineering Series* (Vol. 32, pp. 476-482).

Faghmous, J.H. and Kumar, V., 2014. A big data guide to understanding climate change: The case for theory-guided data science. *Big data*, *2*(3), pp.155-163.

Ford, J.D., Tilleard, S.E., Berrang-Ford, L., Araos, M., Biesbroek, R., Lesnikowski, A.C., MacDonald, G.K., Hsu, A., Chen, C. and Bizikova, L., 2016. Opinion: Big data has big potential for applications to climate change adaptation. Proceedings of the National Academy of Sciences, 113(39), pp.10729-10732.

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A. and Khan, S.U., 2015. The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, pp.98-115.

Hu, H., Wen, Y., Chua, T.S. and Li, X., 2014. Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, *2*, pp.652-687.

Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R. and Shahabi, C., 2014. Big data and its technical challenges. *Communications of the ACM*, *57*(7), pp.86-94.

Kaplan, Robert S., and Anette Mikes. 2012. Managing Risks: A New Framework. *Harvard Business Review 90, no. 6.*

Kapucu, N. and Garayev, V., 2011. Collaborative decision-making in emergency and disaster management. *International Journal of Public Administration*, *34*(6), pp.366-375.

Kim, G.H., Trimi, S. and Chung, J.H., 2014. Big-data applications in the government sector. *Communications of the ACM*, *57*(3), pp.78-85.

Lifescale Analytics. (2015). Descriptive to Prescriptive Analysis : Accelerating Business Insights with Data Analytics. *Lifescale Analytics*, [online] Available at: http://www.lifescaleanalytics.com/~lsahero9/application/files/7114/3187/3188/leadbrief_descripprescrip_web.pdf [Accessed 4 December 2016].

MAMPU (2014). Public Sector Big Data Analytics Initiative: Malaysia's Perspective. *MAMPU,* [online] Available at: http://www.mampu.gov.my/ms/penerbitan-mampu/send/100-forum-asean-cio-2014/275-1-keynote-mampu [Accessed 30 November 2016]

Wang, L., Wang, G. and Alexander, C.A., 2015. Big data and visualization: methods, challenges and technology progress. *Digital Technologies*, *1*(1), pp.33-38.

Malomo, F. and Sena, V., 2016. Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector. *Policy & Internet*.

Othman, S.H. and Beydoun, G., 2013. Model-driven disaster management. *Information & Management*, *50*(5), pp.218-228.

Tekiner, F. and Keane, J.A., 2013, October. Big data framework. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (pp. 1494-1499). IEEE.