# Big Data Implementation in Malaysian Public Sector: A Review

Mohd Amiruddin Hamzah, Saiful Farik Mat Yatin, Maisahara Yusof, Tunku Sofiah Larasih T. Zainol Rashid, Hasnah Shuhaimi, Abu Bakar Suleiman, Ahmad Nazri Mansor, Khairul Mizan Taib

# Big Data Implementation in Malaysian Public Sector: A Review

[1]Mohd Amiruddin Hamzah, [2,3]Saiful Farik Mat Yatin, [4]Maisahara Yusof, [5]Tunku Sofiah Larasih T. Zainol Rashid, [6]Hasnah Shuhaimi, [2]Abu Bakar Suleiman, [2]Ahmad Nazri Mansor, [2]Khairul Mizan Taib

[1] Syariah Judiciary Department, Prime Minister's Department, Malaysia, [2]Faculty of Information Management, Universiti Teknologi MARA (UiTM) Selangor, Malaysia, [3]Members of Advanced Analytics Engineering Center (AAEC), UiTM Malaysia, [4]Teraju Binaan Sdn Bhd, [5]Public Service Department, Prime Minister's Department, Malaysia, [6]Educational Planning and Research Division (EPRD), Ministry of Education, Malaysia
Email: m_amiruddin@esyariah.gov.my, farik@uitm.edu.my

**Abstract**
Big Data is a new world phenomenon for information and knowledge management where the huge chunk of data set been collected and analyzed for further use in many sectors including security, business, investment, advertising, health and etc. The explosion of information sparked from mobile and internet technologies such as via social media and government agencies data gives the Big Data management an ultimate challenge for its characteristics including the aspect of volume for storage, variety in nature, velocity in access speed and veracity in data quality. In Malaysia, there are government agencies that play its vital roles in upholding the initiatives by introducing Big Data in public service (*Data Raya Sektor Awam*) as the key step in implementing Big Data at the national level. Malaysia's 11th Plan (2016-2020) has outlined the Big Data as a strategy to transform public sector for the better service delivery and to reduce the cost to the government. Despite the initiatives effort by the government, there are several challenges faced by the public sector in implementing Big Data approach in the government agencies.
**Keywords:** Big Data Analytics, Big Data Application, Big Data Challenges, Public Sector, Government of Malaysia

## Introduction
Big Data is a prominent topic in the latest development of world economy. The approach by the Malaysian Government to adopt the work towards 5G technology indicates the strong approached made by the government to ensure the digital technology approach becomes the greatest pillar of the government algorithm towards the nation building. Malaysian Government has proofed their initiatives in Big Data Implementation by formulating Big Data as one of the strategies in Malaysia 11th Plan (2016-2020) where they committed to leveraging

data to enhance outcomes and lower the costs in government machinery. The government has proliferating open data among agencies, encouraging cross-agency data sharing, and leveraging Big Data Analytics (BDA). Implementation of the Big Data requires effort from the central agencies with the full understanding from agencies in the government to culturize the data-based activity such as data sharing, digital documentation of the agencies records and to establish open data environment in public sector. The government also need to allocate huge amount of budget in order to realize the digital government concept to the agency. The implementation of Big Data approach in the public sector is a manifestation to the Malaysian Super Corridor (MSC) policy initiated by the government in early of 90's where the government agency such as Road Transport Department, Registration Department, courts, health institution and public universities have move to the information technology as their strategic plan under the e-government flagship. This article will discuss the following objectives which is to describe the diverse concepts within the literature on 'Big Data'; to review the concept of the *Data Raya Sektor Awam* (Public Sector Big Data) implementation; to identify the operational and strategic impacts of 'Big Data' in public sector; and to review the challenges in handling of Big Data in public sector.

## The Concept of Big Data and its Explosion
### Big Data
"Big Data refers to the rising flood of digital data from many sources, including the web, biological and industrial sensors, video, email and social network communication" (Lohr, 2012). Others descried Big Data in its characteristics as high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information process for enhanced insight and decision making (Gartner, 2019). Determine the terms of Big Data can be defined as sets of data, whereby the characteristics of Big Data (volume, velocity and variety of data) need to be to store, manage and analyse using special database software.

### Big Data Explosion
The terms Big Data are commercialising in 1998, introduced by John Mashey whereby the discussion is focused on the amount of data generated by the use of information technology as well as the capacity of storage where the data is to be kept. The discussion about the growth rate of data starts way back in 1990's which known as the era of 'Information Explosion' where people starts to quantify the growth rate of the data generated within the industry. The history of the identifying the growth of data starts in 1944, whereby F. Rider estimated that the US Library will have about 2 petabytes of books in 2040. The estimation growth of the data is not rest on that year as in 2008, the growth of the data is expanding up to Zettabytes (Gu Jifaa, 2014).

### Big Data Framework
The characteristics of 3v's introduced by Doug Laney from Meta Group mentioned that new techniques, analytics and new architecture have to be taken in place in order to generate new information that are useful to the users and support others decision making. Technology advances over time, the size of datasets that qualify as Big Data will also increase and the definition can vary by sector, depending on what kinds of software tools are commonly available and what sizes of datasets are common in a particular industry (Manyika et al., 2011). The term 'big' in the Big Data shows the capacity and size of the data whereby the main

characteristics to describe Big Data are focused on the amount, capacity and size of the data that made available in the organization or people. Apart from 3V's framework introduce by Laney (2001), the characteristics of Big Data are expanding through its times and emergence of new technology taken place. Some of data scientist such as from IBM as well as Big Data Framework agreed to add Veracity as one of fundamental characteristics for Big Data. In addition to veracity, value is another trait that qualify the Big Data. The additional of veracity and value lead to the development of 5V's characteristics of Big Data (Ishwarappa, 2015). Moreover, there are more V's to be added by the specialist such as versatility, volatility, virtuosity, vitality, visionary, vigour, viability, vibrancy, and even virility as well as valueless, vampire-like, venomous, vulgar, violating and very violent (Discover Society, 2013). As there are too many V's that may characterized the Big Data and has reach up to 42V's (Shafer, 2017), the focused will be concern on the Volume, Variety, Velocity, Veracity, and Value.

*Volume*
Big Data size characteristics is referring as Volume of Big Data that results from the expansion of information creation tools such as social media, web page, as well as web apps. The volumes of data generated each second are huge from various sources. The volume of Big Data always discussed in the amount from Megabytes to Petabytes and currently moves to Zettabytes. Google as the largest 'big data' company in the world, processing 3.5 billion requests per day, storing 10 Exabytes of data. It is estimated that 40 Zettabytes (40,000 Exabytes) of data will be created by 2020 (Deep Web Technologies, 2016).

*Variety*
Variety of data in the other hand, described the types of data emerged; those data can be any physical or nonphysical types and in structured or unstructured types generated either by people, technology or information communication equipment. The ability to classify the data in different types that suits the data environment. Refers to the structural heterogeneity in a dataset. Technological advances allow firms to use various types of structured, semi-structured, and unstructured data (Amir Gandomi & Murtaza Haider, 2015). The data can be recorded in variety types and categories then presented in table, sound, video, and photo as well as in database format. Some authors agreed that Big Data involves a great variety of data forms: text, images, videos, sounds, and whatever that may come into the play, and their arbitrary combinations (the type system shall remain constantly open) as well Variety, as one of the essential characteristics of Big Data, is resulted from the phenomenon that there exists nearly unlimited different sources that generate or contribute to Big Data. (Duren Chen, 2013). Data variety often stored at various times for security and safety reasons. Data variety is considered a characteristic of Big Data that follows the increasing number of different data sources, and these unlimited sources have produced much Big Data, both varied and heterogeneous.

*Velocity*
Refers to the speed at the tremendous amounts of data are being generated, complied, composed and analysed. According to Amir Gandomi & Murtaza Haider (2015), velocity can be referred to the rate at which data are generated and the speed at which it should be analysed and acted upon. As example, the emergence of IT tools such as smartphones which enable people to create their post in every single of times via internet apps such as Instagram,

Facebook, Twitter, emails, online shopping apps etc. The trend currently required Big Data technology to analyse the data in order to respond to the speed of the data transmission.

*Veracity*
Refers to the messiness or trustworthiness of the data. This characteristic define that the data can be messy and lead to the error in the process of defining and capturing the data. The quality of data generated can be interrogated while the accuracy of the data is uncontrollable. Most of the data are inaccurate or lack of quality. Thus, the Big Data analytics technology will help organization or people to control the accurateness and quality of the data as well measure the appropriate data to be used as new information and knowledge. The veracity characteristics is result from the volumes of the data generated that make up for the lack of quality or accuracy (Marr, 2014).

*Value*
Refers to the ability to recognise the potential of data to improve decision making and deliver positives outcomes when extracting the data. Most of the organization require data that may help themselves in understanding the benefits of the data to ensure that the acquired data within the organization can be monetized as a return. By understanding their Big Data, the organization may offer products and services that meet the people's requirement at the real-time and at the right place. As mentioned by Firican (2017), "Substantial value can be found in Big Data, including understanding your customers better, targeting them accordingly, optimizing processes, and improving machine or business performance. You need to understand the potential, along with the more challenging characteristics, before embarking on a Big Data strategy".

**Implementation of Big Data in Public Sector in Malaysia**
Due to the nature of Big Data, it requires big scale of organization to assemble data in many aspects of organization. The public sector is an ideal sector to emphasize initiative of Big Data due to its nature of business where it delivers services to people from policy making up until the operational levels. It is well known for the government in the developed nation to adopted Big Data initiatives to indicate the important of data inclusivity and inter-agency approach to ensure the policy formulation and implementation fits the current situation in the country.

Hajirahimova (2017) has review on Big Data implementation in developed countries such as USA, Great Britain, France, Australia, China, Japan and South Korea. Among the characteristics of the implementation of the Big Data initiatives in the countries has concentrated after the year 2010. The implementation was started with the initiative on open data concept imposed to the government agencies. However, United States of America has experienced the need of the associated data early on. The results of the final 9-11 Commission Report identified important fact where the U.S. Government was unable to connect the dots between available pieces of data that could have led to the discovery of the potential of a terrorist attack before it happened. In order to fulfil this requirement, to help ensure this never happens again, the U.S. Government has deployed new innovative technologies and improved inter-and intra-agency information sharing (Vinsik, 2011). After 10 years, Big Data in USA has moved from research to implementation in 2012 where the US Government has indicated USD200 million in budget to various agency to initiate Big Data initiatives in multi-agencies platform (Hajirahimova, 2017).

The initiatives by other developed countries such as Great Britain, France and Japan has moved to the Big Data approach from the open data initiatives. Great Britain and France for example, has started its data initiatives by introducing open data initiatives in 2010 and Japan was in 2012. Great Britain also started with data service where its support data requirements by academician and researchers on governmental data from various agencies. Australia on the other hand, has implemented national level data framework namely Australian Public Service Big Data strategy in 2013. The framework helps the country to improve the public service and make them ready for Big Data approach in policy making and protected citizen security (AIIA, 2014).

**Implementation in Malaysia**
Malaysia is a developing country which has adopted e-government approach since the implementation of MSC in early 2000. However, like the other developed country, Malaysia has embraces Big Data analytics as part of its strategy to boost up the efficiency of the country government in serving its citizens. Public Service in Malaysia play the dominant aspect of development for a developing nation, where government dictates all policies that govern public sectors. Even though the privatisation has taken placed in early 1980s, the government still play its roles in placing appropriate policy to be implemented by government or private agencies.

**Open Data Initiatives in Malaysia**
After successfully implemented e-government initiatives in various government agencies, Malaysia has move forward by introducing open data in 2014. The government has developed a portal for public sector open data which was officiated during the CIO ASEAN Conference in 2014. The portal serves as one-service-centre to access and download the dataset available at the portal. The dataset mostly available in Microsoft Excel format are available openly to the public. The government agency will submit any new open data to the portal administrator and the portal will evaluate the data before setting it available for public access. The open data shared by the agencies must be approved by the Head of Agency to ensure the validity of the data. The portal has been upgraded in 2016 with the international open data requirement portal concept to ease the access to the dataset (Nor Aliah Mohd Zahri (2014). The open data aims to achieve the objectives of open data implementation such as to increase the government service transparency, to help the community to increase the creativity and innovation for a new product, as a platform for public to access government data and to reduce cost to the government (MAMPU, 2017)

Even though the open data concept has been implemented by the government, but the approach of using compiled dataset in Excel format may lead to the outdated data and not a real-time data from the government. For instances, a list of the halal certification entities under Department of Islamic Development Malaysia, the raw set is long outdated, which some was expired in 2015. However, for the researchers, this type of data could be very useful in for studying the trends in the past, grouping or statistical review, but for the open data concept for instant and timely referable could not be satisfied and be achieved by using this portal.

**Big Data in Public Sector**

Government of Malaysia has started its initiatives to harness data and to develop data as part of government asset that should be share with public to increase the effectiveness of its agencies. The concept of Big Data Analytics (BDA) has been tabled during the 25th MSC Implementation Council Meeting held on 13th November 2013 where the Ministry of Communication and Multimedia (KKMM) with Malaysian Administrative Modernisation and Management Planning Unit (MAMPU) and Malaysia Development Corporation (MDeC) has been collaborate in the implementation of Big Data in Malaysia. During the Government IT and Internet Committee meeting (*Jawatankuasa IT dan Internet Kerajaan*/JITIK) 1/2014 on 28th March 2014, which governs all the internet and ICT activities in public sectors has instructed all government agencies to identify BDA initiatives in the agency core functions.

Ministerial Cabinet Meeting on 20th August 2014 has agreed for the Big Data implementation in public service where the MAMPU will be the play maker in the Big Data initiatives in government sector and MDEC will be the key player in the private sector. The government later on included the BDA in Malaysia 11th Plan (2016-2020) and Mid-Term Review of 11th Malaysia Plan in 2018. The BDA is the essential strategy by the government to ensure the effectiveness of government agencies and to uphold the Government initiatives to transform public sector to meet the developed countries standard.
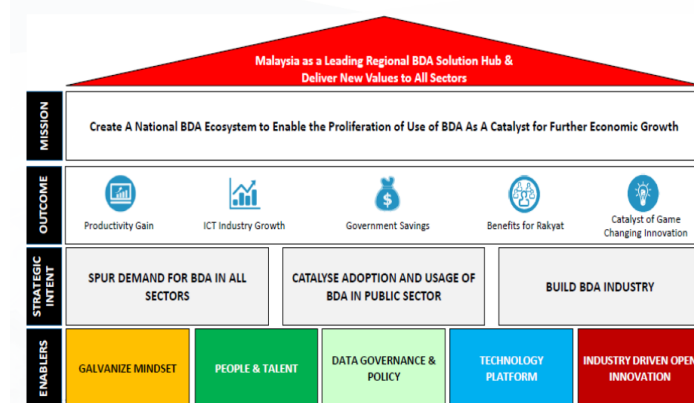
**National Framework of Big Data**



**Figure 1 : National Big Data Framework (MAMPU, 2014)**

In order to ensure the effective, structured implementation of Big Data, the government of Malaysia has developed a national framework for implementation Big Data to the national level. The main target for BDA implementation is to make Malaysia as a leading regional BDA solution Hub and deliver new values to all sectors. The framework consists all important elements of enabling BDA including people centric, government data and policy upholstery, industry backed operation and technology for the Big Data processing and support. With the focus to both sectors, public and private, the Big Data implementation aims to help the country to achieve productivity gain, ICT growth, expenditure saving and innovation to benefitting the people. The government has identified 7 BDA clusters, namely socio-economy, rural infrastructure, criminal, anti-corruption, education, transportation and health (Chandrasekaran, 2014)

**Implementation Big Data in Public Sector**
MAMPU has started Big Data initiative by introducing Data Raya Sektor Awam (DRSA), or Public Sector Big Data, an implementation across all agencies in public sectors. The Implementation started with the framework initiation by government and the Big Data initiative has been started with proof of concept projects and pilot project by the government in 2015. The proof of concept stage includes four (4) government agencies namely Department of Islamic Development to seek Islamic Extremism among Malaysians, Ministry of Finance with two (2) project which is to analyse and build fiscal economic models and to look into sentiment analysis on Cost of Living obtained through Social Media, while Department of Irrigation and Drainage is to develop a knowledge base flood based on the combination of sensor data and social media data and National Hydraulic Research Institute of Malaysia (NAHRIM) is to get the projected of 90-years rainfall in line with the spill over effects on the river bank in Malaysian map. While in 2016, the Government has started the pilot project in four (4) area which are the crime prevention, price watch, Hand, Foot and Mouth Disease (HFMD) and sentiment analysis. The pilot project involving five (5) stage of implementation: framework, platform at Public Sector Data Centre (PDSA), methodology, guideline and the government.

Methodology used in DRSA lists seven (7) steps to be undergo in Big Data implementation in a government agency. First, to understand the business and function of the analysis. Secondly, requirement definition to create a focus scope of analysis, and third stage is to perform data acquisition and exploration. The agency will need to develop analytical models before moving to the data product development before transitioning to the production and monitoring. For example, in HFMD, the data sources will be processed by the government including the data from Ministry of Health internal system and from others related environment department to create a forecasting data. The analytics also will overview the unstructured data from social media to see the current possibly outbreak area of the disease. The agency also needs to develop business questionnaire to see the area of the Big Data analytics, for example correlation between events and weather situation and on the high-risk area of disease outbreak. The visualization to the questions posted in this analytic will help the government to create a mapping solution to the forecasted area that may hit by the HFMD outbreak in the future based on the data collected from the department and social media earlier on. This could help the Ministry of health to raise flags to the agencies to plan the action to prevent from the outbreak and to prepare for the worst-case scenario if the outbreak happens (Yuslina Yunus, 2017).

**Big Data Analytics Implementation Challenges and Solution**
The Big Data research discipline is still evolving and not yet fully established and there is a lack of comprehensive research studies to address the key challenges of Big Data. However, there is a need to relate the Big Data challenges, to have a better understanding of the Big Data phenomenon. Big Data refers to a very large volume of data and it is estimated about 2.5 quintillion bytes of data is produced on daily basis and by 2020, every person will produce about 1.7MB of data every second (Marr, 2018). There were three main categories of the Big Data challenges i.e. data challenges, process challenges and management challenges (Sivarajah et al., 2017).
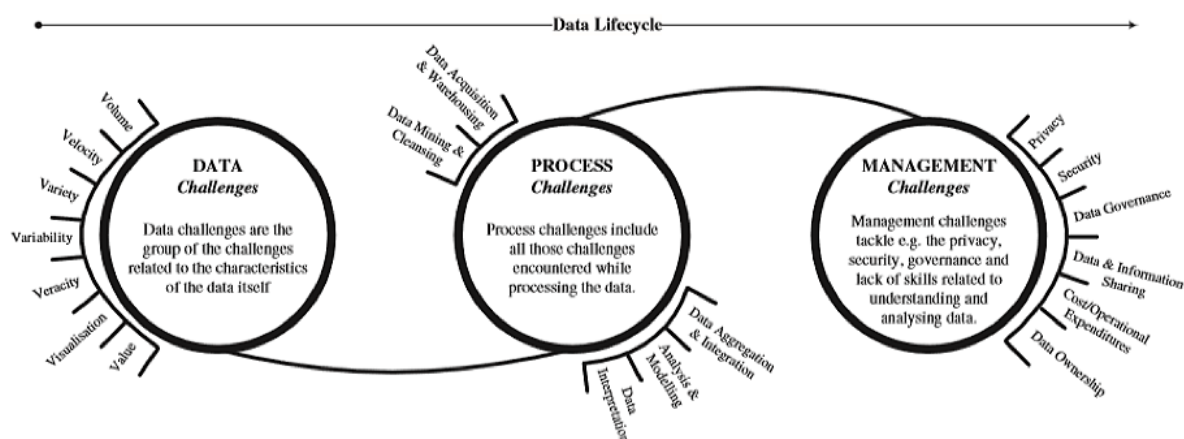
**Figure 2 - Conceptual Classification of Big Data Challenges (Source: Akerkar, 2013)**

**Management Challenges**
Many organizations have realised the potential of using Big Data to support their organization. However, many failed to effectively extract its valuable data. It is actually a moving target. New Vantage Partners in their study conducted in January 2019 reported that 95% of the Fortune 1000 entrepreneurs surveyed have undertaken a Big Data project in the past five years, but only 48.4% have managed to benefit from these projects (Davenport & Bean, 2019). The importance of Big Data is not about the data availability but how the organization makes use of the collected data and turns it into actionable insights. The insights will help in better decision making, cost saving, time reducing, better market condition understanding and gauge customer need. In order to get the insights, organizations need analytical tools to help them get results quickly so that they could respond to their marketplace.

While Big Data has many benefits to offer, it comes with its own sets of challenges. The key for solving these challenges is to properly analyse the organizations' need and choose correlate course of action. There are three elements involves in the Big Data concept i.e. people, process and technology.

The success of Big Data projects requires the collaboration from multi-disciplinary field and from various sources. To get the desired results, all parts in organizational structure need to change simultaneously such as tasks, technologies, people and structure. In the people element, organizational resistance from managers and employees contributes to the Big Data challenges.

Frequently, the organization's lack of understanding on what Big Data is and the technology or infrastructure will be best suited to them, leads to the failure of utilising the Big Data's benefits. The Big Data concept has to be accepted and acknowledged by the top management and then pass to the manager's level. Therefore, a series of workshop and training is needed in order to guarantee the Big Data comprehension and acceptance. With the acceptance, the management has to control and monitor the Big Data implementation and usage throughout the organization's operation.

It is important for organizations to hire professionals who understand BDA. There is a critical shortage of highly experienced and certified subject matter experts such Data Scientists, Big

Data Analyst and Data Engineers. It is estimated by IBM that the annual demand for these professionals will lead to 700,000 new recruitments by 2020 (Redazione, 2018). In dealing with new technologies, training people from entry level can be expensive which makes organizations opt for automation solution such as artificial intelligence and machine learning to build insights but this also requires well-trained people or outsourcing people.

It is also important to consider the appointment of Chief Data Officer (CDO) in aligning the organizations' purpose to the Big Data's governance. The Big Data's governance includes the security aspects, business policies, data and data quality itself.

### Data Challenges

The data challenges are a group of challenges related to the characteristics of the data itself (Sivarajah et al., 2017). International Data Corporation (IDC)'s report, the amount of data store around the world about doubles every two years from 2012 to 2020 (Gantz & Reinsel, 2012). The definitions of Big Data also vary according the user's point of view or communities that are interested. The Big Data comes in many forms for example, structured data, semi structured data and unstructured data. The differences as follows:

- Structured data is a well-defined data definition and considered as the traditional form of data;
- Semi structured data is a form of structured data that does not reside in a rational database but easier to analyse due to its organizational properties;
- Unstructured data does not have a pre-defined data model and often in free text, graphics, multi-media content.

The general computational solutions have not been discovered for Big Data processing especially when using unstructured data (Kaisler et al., 2015). In addition, for data classifications purposes, IDC indicates that 90% unstructured data is never analysed (Xavier Pornain, 2014). While Gartner Research predicts that between 2012 and 2017, there will be 800% data growth of which 80% is unstructured (Egli, 2016).

### Process/ Technology Challenges

Process challenges are the group of challenges encountered while collecting, processing, analysing, synthesizing the data to interpreting and presenting the end results. Things related to process challenges is Data acquisition and warehousing; Data mining and cleansing; Data aggregation and integration; Data analysis and modelling and also Data interpretation.

Organizations need to invest on information technology infrastructure, tools and data warehouse architecture to analyse and synthesize available data. Currently, the conventional database methods and information management tool cannot process it efficiently due to the Big Data's own character. To keep up with the data needs, a wide variety of data sources, data collection strategies and format can also be a challenge in terms of data integration management. For solution, the organization can opt to invest on a new generation of analytical tools help to significantly reduce the time needed to get results so that they can respond to the marketplace as fast as possible.

Organization can control the output data quality by using data governance that requires a combination of policy change and technology including allocating a dedicated people to

monitor data and defining rules and procedures. Organizations have another option, which is to use data management solutions designed to simplify data governance towards Big Data storage, quality and accuracy. The Data Warehouse Institute reported that in 2003, data quality problems such as user input error, missing data, incorrect data linking, logic conflicts, and inconsistent data and duplicates data, costs US companies around $600 billion every year (Eckerson, 2003).

Many organizations should understand the Big Data implications on the data privacy and data security aspect. It is necessary to introduce best practices for secure data collection storage and retrieval. Only 39% of organizations use additional security measures for their data repositories were reported in a study by IDG (Brown, 2019). For example, even though it is more convenient to the organization in terms of cost, to opt for storing data on cloud-based system, organization should not tolerate with sensitive data that they work with. Some of the most popular additional measures include selecting trustworthy data sources, using data segregation, using data encryption, using identity control, restricting access.

In most countries, the current legislation is generally applicable to Big Data. The current law on data protection are often based on subjective individual right. The protection of privacy and personal data is at risk when adopting Big Data. Using data mining tools, large personal data collections is analysed to find patterns and predict preferences and interests. These of patterns and predictions are stored in organization databases and combined with new data. Some organizations' business models are built by using and selling user profiles generated from these data sources. Thus, the U.N. members adopted a resolution aimed at protecting the privacy right against unlawful surveillance in the digital age (Munir et al., 2015).

In other hand, the Big Data concept has a fundamental conflict with the data protection principle. The current law is focused on relatively static stages of data while in Big Data, data is continuously going through a circular process i.e. data is linked, aggregated and anonymize and then again de-anonymized and then again pseudonymised. This implies that data are not collected and process at the individual level but rather being process by aggregating the data to generate general patterns for statistical and group profiles. The data collection so widespread that it is impossible for individual to access each data process to determine whether it includes their personal data (van der Sloot & van Schendel, 2016). So, it is unclear to how individual interest is directly affected.

Consequently, instead of replacing the current regulation with the Big Data regulation, its best to formulate new rule in addition to the current regulatory framework. The number of countries that have enacted or enacting data protection laws around the world is growing. However, there is a concern that the current law hinders technological innovation over privacy protection. The government has to take consider these two issues:
- The current regulatory framework is regarded as restrictive and as a result, make it difficult for technological innovation and the use of new technologies;
- Many organizations are unsure on how Big Data processes can be applied and interpreted within the current regulation framework.

Penemon Institute reported that 66% of the respondents indicated that their insufficient knowledge on managing threat in their organizations' data governance. The study also shows

that even though 66% of the organizations had experienced data breach, only 9% indicate that they will spend budget on sensitive data management (Egli, 2016).

**Conclusion**

BDA is a new dimension to the world technology which is extremely important for a government to prepare for the vast technology evolution including IR4.0 and 5G technology. Government must take advantage of these technology to improve the effectiveness of government agencies especially to accurately formulate the national policy to ensure the outcome to the government. In Malaysia context, Big Data approach by the government is at par with other developed nation in term of the national initiation and implementation to change mind set of the government servant and agencies towards the idea of openness of data. However, there are still a big room to explore to the government to make the Big Data as an effective asset to the government by defining the Big Data approach. Not only providing previous old dataset to the researchers, but also need to consider real time access to the government data. The government must be bold enough to offer the latest access technology to the researcher by providing live data, such as using application programming interface API protocols. This is to help the researchers to develop a strong analytics system to quickly and accurately processing the data for the data quality and reliability. This also could help the government in harvesting data in real time and to effectively give the new dimension to the government in harnessing the data and to meet the expectation of government and private agencies to strive to the best for the nation.

**References**

Akerkar, R. (Ed.). (2014). *Big Data computing*. Florida, CRC Press.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

AIIA (The Australian Information Industry Association). (2014). The Australian Public Service Big Data Strategy. Retrieved June 28, 2019, from https://www.aiia.com.au/documents/policy-submissions/policies-and-submissions/2014/the_australian_public_service_big_data_strategy_25_03_2014.pdf

Kumar, T. N., Jindal, H., Subramanian, P., Sharma, A. K., Alok , & Krishna, S. (2019). *The Four V's of Big Data*. Big Data Framework©. https://www.bigdataframework.org/four-vs-of-big-data/.

Brown, C. (2019). *Big data challenges impacting data-driven business goals*. InsiderPro. https://www.cio.com/article/3396182/big-data-challenges-impacting-data-driven-business-goals.html.

Chandrasekaran, Arun. (2014). Seminar Big Data & Open Data Sektor Awam: Using Big Data Analytics to Improve Government Performance. Retrieved 1 June 2019 from https://www.mampu.gov.my/ms/penerbitan-mampu/send/102-seminar-big-data-open-data-sektor-awam/295-1-keynote-address-mr-arun-gartner-big-data-in-government

Che D., Safran M., Peng Z. (2013) From Big Data to Big Data Mining: Challenges, Issues, and Opportunities. In: Hong B., Meng X., Chen L., Winiwarter W., Song W. (eds) *Database Systems for Advanced Applications*. DASFAA 2013. Lecture Notes in Computer Science, vol 7827. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40270-8_1. https://doi.org/10.1007/978-3-642-40270-8_1

Davenport, T. H., & Bean, R. (2019, January). Retrieved from https://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf

*Deep Web Tech Blog*. (2016). Let's take a look at some really BIG "big data". Retrieved from https://www.deepwebtech.com/2016/06/lets-take-a-look-at-some-really-big-big-data/

Eckerson, W. W. (2002). *Data quality and the bottom line: Achieving business success through a commitment to high quality data*. Seattle: The Data Warehousing Institute. http://download.101com.com/pub/tdwi/Files/DQReport.pdf

Economic Planning Unit, Prime Minister Office. (2015). *Eleventh Malaysia Plan 2016-2020 Anchoring Growth on People*, Putrajaya, Percetakan Nasional Malaysia Berhad

Egli, M. (2016). Mitigating the Risks of Insider Threat on Unstructured Data through Data Governance. Retrieved May 22, 2019, from https://scholarsbank.uoregon.edu/xmlui/handle/1794/21963

Firican, G. (2017). *The 10 Vs of Big Data*. Transforming Data with Intelligence. https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx.

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big Data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, *2007*(2012), 1-16.

*Gartner.* (2019). Retrieved 15 May, 2019, from IT Glossary: https://www.gartner.com/it-glossary/big-data

Gu Jifaa, Z. L. (2014). Data, DIKW, Big Data and Data science. *2nd International Conference on Information Technology and Quantitative Management, ITQM* (pp. 814 – 821). Elsevier. doi:10.1016/j.procs.2014.05.332

Hajirahimova, M. S., Aliyeva, S. A., & Sh., M. (2021). *Big Data Initiatives of Developed Countries.: Problems of Information Society.* İTP Jurnalı. https://jpis.az/en/journals/146. 10.25045/jpis.v08.i1.02.

*IBM Big Data & Analytics.* (n.d.). Retrieved 1 June, 2019, from The Four V's of Big Data: https://www.ibmbigdatahub.com/infographic/four-vs-big-data

Ishwarappa, J. A. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technolog. *International Conference on Intelligent Computing, Communication & Convergence* (pp. 319-324). Odisha, India: Procedia Computer Science. doi:https://doi.org/10.1016/j.procs.2015.04.188

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and challenges moving forward. In *2013 46th*. *Hawaii International Conference on System Sciences* (995-1004). IEEE.

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety.* Stamford: Meta Group.

Lohr, S. (2012). *New U.S. Research Will Aim at Flood of Digital Data*. The New York Times. https://www.nytimes.com/2012/03/29/technology/new-us-research-will-aim-at-flood-of-digital-data.html.

MAMPU (Malaysian Administrative Modernisation and Management Planning Unit). (2017). Data Terbuka Sektor Awam (DTSA). Retrieved from https://www.mampu.gov.my/ms/penerbitan-mampu/send/157-tahun-2017/997-data-terbuka-sektor-awam-dtsa

Marr, B. (2014). *Big Data: The 5 Vs Everyone Must Know.* Retrieved from Linked: https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/

Marr, B. (2018). How much data do we create every day? The mind-blowing stats everyone should read. *Forbes, May 21st, www. forbes. com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/, accessed May 2nd*.

McArdle, R. K. (January–June, 2016). What makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets. *Big Data & Society*, 1–10. doi:https://doi.org/10.1177%2F2053951716631130

Ministry of Economic Affairs (2018). *Mid-Term Review of The Eleventh Malaysia Plan 2016-2020, New Priorities And Emphases*. Putrajaya. Percetakan Nasional Malaysia Berhad

Munir, A. B., Yasin, M., Hajar, S., & Muhammad-Sukki, F. (2015). Big Data: big challenges to privacy and data protection. *International Scholarly and Scientific Research & Innovation*, *9*(1).

Redazione. (2018). *Data Analyst, the most in-demand job of the coming years*. Morning Future. https://www.morningfuture.com/en/article/2018/02/21/data-analyst-data-scientist-big-data-work/235/.

Zahri, N. A. M. (2014) Seminar Big Data & Open Data Sektor Awam: Rangka Kerja Big Data Sektor Awam. Retrieved 1 June 2019 from https://www.mampu.gov.my/ms/penerbitan-mampu/send/102-seminar-big-data-open-data-sektor-awam/298-5-5dato-drrangka-kerja-big-data-sektor-awam-mitc-10-04-2014-v2-1

Pornain, X. S. (2015). *Rewiring to Tackle Unstructured Data*. Wired. https://www.wired.com/insights/2014/07/rewiring-tackle-unstructured-data/.

Discover Society. (2013). *Focus: Big Data, Little Questions?* Discover Society. https://discoversociety.org/2013/10/01/focus-big-data-little-questions/.

Shafer, T. (2017). The 42 V's of Big Data and Data Science. Retrieved June 28, 2019, from https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263-286.

Van der Sloot, B., & Van Schendel, S. (2016, September 19). *Ten Questions for Future Regulation of Big Data: A Comparative and Empirical Legal Study*. JIPITEC. https://www.jipitec.eu/issues/jipitec-7-2-2016/4438.

Vinsik, S. (2011). *9/11, A Decade Later -- Connecting the 'Big Data' dots: A decade of lessons learned*, retrieved June 2, 2019 from https://www.gsnmagazine.com/article/24545/911_decade_later_connecting_%E2%80%98big_data%E2%80%99_dots_decade

Yunus, Y. (2017). *Seminar Kerajaan Digital: Mentransformasi Penyampaian Perkhidmatan: Program Berpacukan Data - Analitis Data Raya*. Retrieved 1 June 2019 from https://www.mampu.gov.my/ms/penerbitan-mampu/send/157-tahun-2017/996-mentransformasi-penyampaian-perkhidmatan-program-berpacukan-data-analitis-data-raya