BIG DATA FRAMEWORK TO EVALUATE AND ANALYSE NATIONAL COVID-19 IMMUNIZATION PROGRAMMED (NCIP) IN MALAYSIA: A COMPARATIVE STUDY

NAJHAN M. IBRAHIM[1,*], AMELIA R. ISMAIL[2]

[1]Department of Information System, International Islamic
University Malaysia, Jalan Gombak, 53100, Selangor, Malaysia
[2]Department of Computer Sciences, International Islamic
University Malaysia, Jalan Gombak, 53100, Selangor, Malaysia
*Corresponding Author: najhan_ibrahim@iium.edu.my

Abstract

Many governments around the world have launched their open government data (OGD) portal to improve the government's transparency by sharing their data with the public such as National Covid-19 Immunization Programmed (NCIP), which has been published at https://github.com/CITF-Malaysia/citf-public. However, increasing the number of datasets, data types, volume and complexity will be raised the integration issues. There-fore, it is essential to evaluate and analyses those huge amounts of these datasets. NCIP provides multiple data sources and datasets. These may raise the Big Data (BD) issues and pose various evaluation and analysis problems to produce valuable information. To generate meaningful linked data to support the purposes of this research study, the relationship between these disparate datasets needs to be identified and construct a comprehensive framework. In order to understand the causes of OGD development of big data, this study involves a detailed examination and comparison of existing theories and actual approaches to handle public sector open data concerns. According to the review, the framework was dominantly adopted over architecture, infrastructures, theoretical and conceptual framework in previous research to examine the revolution of government public accessible data. According to the findings, most existing frameworks do not con-sider the demand for public open data in health such as NCPI. Previous re-search on OGD for health has a lesser number of advanced BD frameworks. In the public sector, there is still a lack of investment and use of Big Data. The findings will aid academics in doing empirical research on the revealed need, as well as offer decision-makers with a better understanding of how to leverage OGD adoption in health by taking relevant actions.

Keywords: Big data framework, National Covid-19 immunization programmed, Open government data.

1. Introduction

Big data was first linked to three characteristics: quantity, variety, and speed. Big data was defined as data that was organised, semi-structured, or unstructured [1]. This research paper will look at the fundamental principles of big data and how they apply to different forms of data. In 2017, Graham, Milligan, and Weingart defined Big Data (DB) as "data that normal technology can't handle" or "data that requires specific computational intervention to operate" [2-5] are a couple such examples. Data has become one of the world's most valuable resources. Data can be used for various type of solution such as to generate a set of useful of information for the future, all of which are driven by creativity and innovation. To obtain information and provide services, data is also a critical component.

In recent years, in the global economy, Big Data (BD) has become a contentious issue. Big Data is a piece of new evidence and comprehension managing phenomena where enormous data amounts are composed and managed to be adopted in different of domains such as research, enterprise, industry, marketing, e-learning, health and media social. John Mashey invented the term "Big Data" in 1998 to describe the amount of data generated by information technology and the storage capacity required to keep it. Big Data also can be defined as an ever-increasing number of resources in a digital format with various from, comprising social media, public open data, and network sensors, image, spreadsheet, video, and email [3, 6-8].

As a result of significant advances in database technology, the problem of overflow in and for computer information systems has developed. Industry, academia, and government have entered into a number of comprehensive research and development cooperative agreements. There is also much discussion about whether Big Data can replace traditional data recordings. However, the ever-increasing amount of data such as NCIP has also increased the complexity of evaluation, analysis issues and the existing Big Data framework is not focus on Covid-19 Immunization Programmed [9].

In addition, technical developments appear to make it easier to collect more data in order to find more useful information. Additional data collection, on the other hand, may result in the development of more perplexing or complex data. As a result, consolidating and developing an analytical technique for selecting crucial data specifics for the research study's unique purpose is critical. The BD framework is presented to accommodate the analyses and evaluate Covid-19 vaccine distribution using deep learning approach to address the difficulties of the existing BD. While there are challenges with Big Data, it also has the potential to bring innovative solutions with broader applicability. As a result, this study recommended the Big Data Framework for OGD to analyse Malaysia's Covid-19 Immunization Program [10, 11].

2. Literature review

Academics and non-profit organizations with a close relationship to the open data movement have impacted the definition and popular understanding of the phrase open data in recent years. Open data is information that is freely available online and can be used, re-used, and disseminated by anybody for any purpose, with no technical limitations or prohibitions. The term OGD (Open Government Data) or

GOPD (Government Open Public Sector Data) refers to data that has been made publicly available in an open platform and is freely available for sharing, re-use, and dissemination on the Internet. [12-14]. OGD innovation can promote government accountability and transparency, empower citizen engagement in government service, and boost economic growth through the reuse of data. Openness and transparency are, of course, important objectives. Governments are more inclined to launch open data projects if they feel it would also drive economic growth, increase the efficiency of public service delivery, and lead to innovation, according to years of research and practice. It is also critical to keep emphasizing these value propositions, emphasizing the full range of benefits that open data might potentially be important to hold governments accountable [13, 15].

Even though such data has varying degrees of availability, it should be relatively easy to use. According to the broad consensus, Open access should always be freely accessible. Public open data, and from the other perspective of information, which is generated and maintained by government agencies for public access, therefore fits underneath the open access umbrella [9]. Even though data analytics and open data are related, they are not the same concept, according to Ijab et al. [9], large data can become more valuable and democratic if it is made open. Using open data principles to alleviate some of the problems that have occurred because of big data can be beneficial.

In a concept paper, Bugbee et al. [10] proposed a virtualized data system architecture that supports evolving cyber-infrastructures as presented in Fig. 1. The data pool acts as a center to store many types of datasets. Metadata generation, data input, reconfiguring, description, and regulatory compliances are among the basic requirement for the proposed framework to provide a basic service, as are services that help users search, visualize, and share information. Users can access data in a variety of ways thanks to this architecture. Furthermore, he stated that data programs should invest in cloud-native data systems that comply with this conceptual paradigm. It is significant because it shows how investing in collaborative, information-sharing stewardship solutions may help open research ideas.
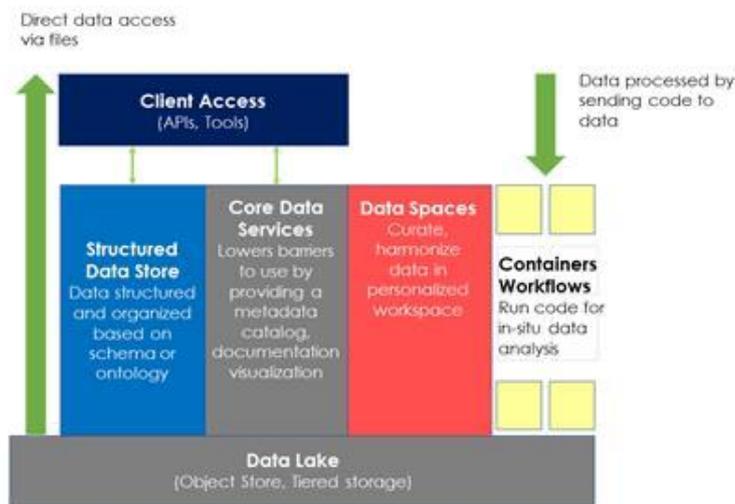

Fig. 1. Conceptual data architecture in the cloud [9].

Since the Multimedia Super Corridor (MSC) was established in the early 2000s, the Malaysian government has also been implementing an e-government agenda. Similar to other developing nation, Malaysia has begun to adopt BD technology into government strategy to develop future plan and increase the country's government's efficiency in providing public services to its citizens. According to Hamzah et al. [7], the government determines the regulations that regulate the public sector in Malaysia, making it the most crucial aspect of development for a growing country. Although privatization has been around since the 1980s, the authority still has an influence in what policies are acceptable for government and private firms to adopt. The Malaysian government has designed a national framework for Big Data implementation at the national level to assure its effective and systematic implementation. The fundamental goal of BDA implementation is to position Malaysia as a regional BDA solution hub, bringing new benefits to all industries [16].

The framework includes all necessary elements for supporting BDA, such as people-centricity, government data and policy upholstery, industry-backed operations, and Big Data processing and support technologies as presented in Fig. 2. Big Data implementation focuses on both public and commercial sectors to help Malaysia achieves productivity gains, ICT growth, cost savings, and innovation for the benefit of the people. Furthermore, the government service does not operate for revenue, and its primary goals are continue providing services to the public, create economic success, and protect residents' civil liberties. OGD innovation, According to Haini et al. [8], reusing data can increase government accountability and transparency, community empowerment in public sector, and economic growth. As a result, given the numerous concerns raised about the public's use of big data, its services are committed to strengthening operational efficiency and engagement for citizens while also addressing current problems [17].
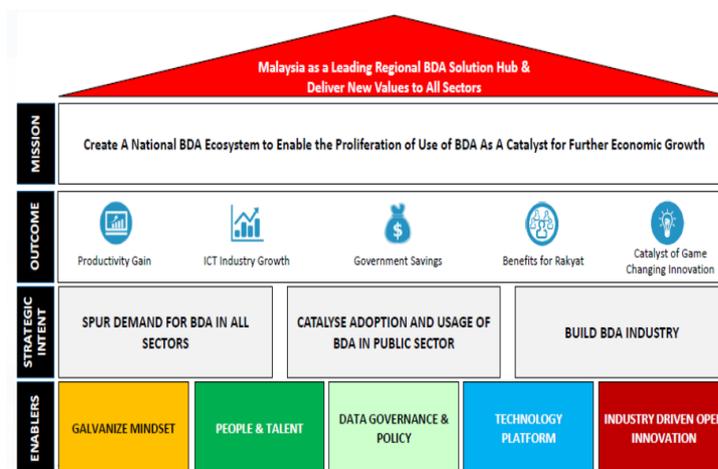

Fig. 2. Malaysia BD framework [14].

3. Research Methodology

This study's search methodology is based on the SLR (Systematic Literature Reviews) strategy for this project. The method is a methodical, explicit, thorough, and repeatable process for discovering, analysing, and summarising the current

corpus of finished and recorded research, academic, and practitioner work. The advantage of SLR is that it is a well-defined methodology. In a variety of instances, SLR may also give information about the ramifications of an occurrence utilising a variety of empirical approaches. If studies yield consistent results, systematic reviews demonstrate that a topic is robust and transferable. If the investigation's findings are inconclusive, the sources of variance may be investigated.

The purpose of the SLR is to assess and validate current research on the use of big data frameworks in OGD adoption for government organisations providing public services. The three stages of the SLR procedure are planning, conducting, and reporting. This method necessitates a comprehensive examination of multiple primary sources. It allows data integration by merging findings from studies using a variety of approaches and pulling data from studies with similar end measures. Between January 2017 and June 2021, the research process continues with a search of titles, keywords, and abstracts of papers published in electronic journal databases such as Scopus, Web of Science, Science Direct, ACM, Springer, and IEEE. At the executing stage, Boolean "AND" and "OR" were used to develop the search strategy. The goal of the search string was to learn more about Big Data framework for OGD study.

The study's primary research technique was a GOD case study in Malaysia (Covid-19 Immunization Programmed), which was released on Github at https://github.com/CITF-Malaysia/citf-public. Malaysian public open data was chosen for this study because the available data is extensive and has been publicly available for some years, with a range of data types. As indicated in Fig. 3, the research technique is separated into three phases.
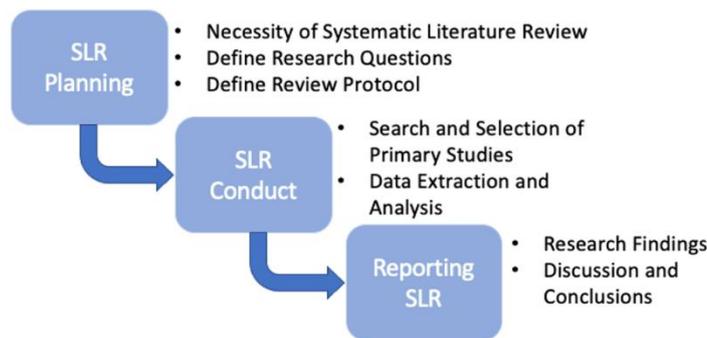
Fig. 3. The research methodology.

Phase 1: Analysis of existing big data framework

The proposed study will conduct a comprehensive analysis of the existing framework by conducting a literature review and identifying significant datasets. The inquiry step will entail a critical review of current and previous research using books, papers, conferences, research reports, and other academic materials. This investigation will involve knowledge of the current challenges in the big data framework. The purpose of this analysis is to identify research gaps, requirements, and the most appropriate approaches to apply in the suggested model.

- Conduct a literature review and comprehensive comparison study for the existing Framework.

- Identify the significant attribute for OGD to be included in comparation table.
- Formulate and analyse the selected attributes.

Phase 2: Conduct comparative study

- Select the relevant existing framework.
- Conduct in depth comparative study.
- Discuss the finding

4. The findings

This section will be explored considering the existing BD framework and Public Open Data Architecture. This section also contributes to the research by offering a comprehensive review of current research studies. The investigation was carried out by filtering and evaluating relevant publications. To capture relationships between variables, we used a few conceptual and theoretical frameworks from past quality research. The importance of conducting a thorough analysis and proposing a feasible solution for the corresponding Big Data and GPOD issues, which have become the key resources for generating and investigating various sorts of information, cannot be overstated. Only a few research studies on the Big Data framework in GPOD for health are now available.

Table 1 highlights and analyses the current Big Data framework, which is centred on the use of Open Data in government sector. The table's row details the research's aims, focus, and gap. There are two sorts of existing research work on Big Data and Open Data: theoretical and practical approaches. The approaches were picked from a variety of research projects, some of which are theoretical and conceptual frameworks for public accessible data with important highlighted points of view. Others are concentrating on the Big Data framework's infrastructure and architecture. However, none of the existing frameworks for NCIP, such as the Covid-19 pandemic, are focused on Big Data.

Table 1. Current big data framework.

Year	References	Objectives	Focus	Research Gap
2019	Ijab et al. [9]	Development of data quality framework for generic view of big data	Focus on big data quality framework	Does not support big data, development for NCIP
2021	Ibrahim et al. [15]	Development of big data interoperability framework	Focus on integration between multiple type of dataset.	Does not support big data, development for NCIP
2017	Joaheer and Nagowah [16]	Development of big data architecture for diabetes	Focus on big data architecture	The proposed architecture only supports for Diabetes
2020	Dinesh et al. [18]	Development of open data health heatmap	Focus on Open Data Health Heatmap in general	Does not support big data, development for NCIP
2016	Jetzek et al. [19]	Development of Open data value driven framework	Only focus on private sector for open data framework	Does not support big data, development for NCIP
2020	Kurshid et al. [20]	Development of conceptual model and theoretical model of public open data.	Focus on theoretical concept of public open data	Does not support big data, development for NCIP

5. Conclusions

The goal of incorporating big data framework and open data is to enhance outcomes while lowering costs and better analysis of huge available data. One way to address this is to promote open data among government agencies. Encourage data exchange across agencies and take advantage of big data technology.

The government can use the Public Sector Open Data program to produce innovative data-driven goods with the purpose of increasing public service and availability while considering the needs to access quality information. DB technology is a new technology area that a country needs to plan for enormous scientific advances and digitalization.

To achieve a beneficial outcome for the boost their performance, the government and government agencies must use technology, particularly when it comes to accurately writing government policy. In perspective of a national initiative and trend is expected to continue to change public officials' and agencies' attitudes regarding data transparency.

Although the theoretical and conceptual potential of free public data has been established, there is still more work to be done. If publicly accessible data is to become a reality, many obstacles must be addressed, as well as significant gaps in our understanding. Indeed, to assist in the resolution of difficult traditional system challenges and the improvement of data analysis to provide relevant knowledge for the future.

Acknowledgments

References

1. Graham, S.; Milligan, I.; and Scott, W. (2015). *Exploring big historical data: The Historian's macroscope*. World Scientific Publishing Company.

2. Ijab, M.T.; Ahmad, A.; Kadir, R.A.; and Hamid, S. (2017). Towards big data quality framework for Malaysia's public sector open data initiative. *International Visual Informatics Conference*, *IVIC* 27: Advances in Visual Informatics, Bangi, Malaysia, 79-87.

3. Gyamfi, N.K.; Appiah, P.; Sarpong, K.A.-M.; Gah, S.K.; Katsriku, F. and Abdulai, J.-D. (2017). Big data analytics: Survey paper. *Conference Proceeding: Dialogue on Sustainability and Environmental Management*, Accra, Ghana, 101-112.

4. Singh, R.K. (2018). Taxonomy of big data analytics: Methodology, algorithms and tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4(12), 101-104.

5. Sun, A.Y.; and Scanlon, B.R. (2019). How can big data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letter*, 14 073001.

6. Rathore, M.M.; Son, H.; Ahmad, A.; Paul, A.; and Jeon, G. (2018). Real-time big data stream processing using GPU with spark over Hadoop ecosystem. *International Journal of Parallel Programming*, 46(3), 630-646.

7. Hamzah, M.A.; Yatin, S.F.M.; Yusof, M.; Rashid, T.S.L.T.Z.; Shuhaimi, H.; Suleiman, A.B.; Mansor, A.N.; and Taib, K.M. (2020). Big data implementation in Malaysian public sector: A review. *International Journal of Academic Research in Business & Social Sciences*, 10(11), 1461-1474.

8. Haini, S.I.; Rahim, N.Z.A.; . Zainuddin, N. M.M.; and Ibrahim, R. (2020). Factors influencing the adoption of open government data in the public sector: A systematic literature review. *International Journal on Advanced Science, Engineering and Information Technology*, 10(2), 611-617.

9. Ijab, M.T.; Surin, E.S.M.; and Nayan, N.M. (2019). Conceptualizing big data quality framework from a systematic literature review perspective. *Malaysian Journal of Computer Science*, Special Issue 1, 25-37.

10. Bugbee, K.; Ramachandran , R.; Maskey, M.; Barciauskas, A.; Kaulfus, A.; That, D.-H.T.; Virts, K.; Markert, K.; and Lynnes, C. (2020). Advancing open science through innovative data system solutions: The joint ESA-NASA multi-mission algorithm and analysis platform (MAAP)'s data ecosystem. *Proceedings of IGARSS* 2020 - 2020 *IEEE International Geoscience and Remote Sensing Symposium*, Waikoloa, HI, USA, 3097-3100.

11. Carrillo-Larco, R.M.; and Castillo-Cara, M. (2020). Using country-level variables to classify countries according to the number of confirmed COVI D-19 cases: An unsupervised machine learning approach. *Wellcome Open Research*, 5: 56.

12. Pung, R.; Chiew, C.J.; Young, B.E.; Chin, S.; Chen, M.I.-C.; Clapham, H.E.; Cook, A.R.; Maurer-Stroh, S.; Toh, M.P.H.S.; Poh, C.; Low, M.; Lum, J.; Koh, V.T.J.; Mak, T.M.; Cui, L.; Lin, R.V.T.P; Heng, D.; Leo, Y.-S.; Lye, D.C.; and Lee, V.J.M. (2020). Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *Lancet*, 395(10229),1039–1046.

13. Khurshid, M.M.; Zakaria, N.H.; Rashid, A.; Ahmad, M.N.; Arfeen, M.I.; and Shehzad, H.M.F. (2020). Modeling of open government data for public sector organizations using the potential theories and determinants - A Systematic review. *Informatics*, 7(3), 24.

14. Mustapa, M.N.; Hamid, S.; and Nasaruddin, F.H. (2019). Exploring the issues of open government data implementation in Malaysian public sectors. *International Journal on Advanced Science, Engineering and Information Technology*, 9(4), 1466.

15. Ibrahim, N.M.; Hussin, A.A.A.; Hassan, K.A.; Breathnach, C. (2021). *Big data interoperability framework for Malaysian public open data*. In: Saeed F.; Mohammed, F.; Al-Nahari, A. (eds). *Innovative systems for intelligent health informatics*. IRICT, KL, Malaysia, 2020. Lecture Notes on Data Engineering and Communications Technologies, vol 72. Springer, Cham, 421-429.

16. Joaheer, R.; and Nagowah, S.D. (2017). A big data framework for diabetes in Mauritius. *Proceedings of the 2017 International Conference on Infocom Technologies and Unmanned Systems* (*Trends and Future Directions*) (*ICTUS*), Dubai, United Arab Emirates, 126-132.

17. Ariffin, A.S.; and Aziz, M.A.S.B.A. (2022). Public servants' perception towards publishing quality and impactful open data to support open science initiatives in Malaysia. *Journal of Science, Technology and Innovation Policy*, 8(1), 44-57.

18. Dinesh, A.S.; Mathur, V.; Ansil, B.R.; Chandru, V.; Chellam, R.; Vanak, A.T.; Ramakrishnan, U.; and Rajagopal, P. (2020). Health heatmap of India: An open data platform. *Journal of the Indian Institute of Science*, 100(4), 701-716.

19. Jetzek, T.; Avital, M.; and Bjorn-Andersen, N. (2019). The sustainable value of open government data. *Journal of the Association for Information Systems*, 20(6), 702-734.

20. Khurshid, M.M.; Zakaria, N.H.; Arfeen, M.I.; Rashid, A.; Shehzad, H.M.F.; Ahmad, M.N. (2020). An intention-adoption behavioral model for open government data in Pakistan's public sector organizations - An exploratory study. In: Sharma, S.K.; Dwivedi, Y.K.; Metri, B.; and Rana, N.P. (eds). Re-imagining diffusion and adoption of information technology and systems: A Continuing Conversation. TDIT 2020. IFIP Advances in Information and Communication Technology, vol 617. Springer, Cham., 377-388.