LINKED DATA APPROACH IN ACCESSING GEOSPATIAL BIG DATA

NOORAFIZA BINTI NORMAN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Philosophy

Faculty of Built Environment and Surveying
Universiti Teknologi Malaysia

OCTOBER 2019

DEDICATION

Alhamdulillah,

All praise and thank to Allah S.W.T for His greatness for giving me the strength and courage to complete this thesis.

This thesis is dedicated

To my beloved family; umi, abah, mamat and nabila,

Thank you for the non-stop *doa*, love, patience, support, understanding and motivation that have been given to me throughout the roller coaster ride of my student life.

To my respectful lecturers,

Thank you for your great patience and kindness of teaching and advising me during my study times here. You have taught me that even the largest task can be accomplished if it is done one step at a time.

To my last long friends,

Thank you for being friends with me for all the time we have spent during ups and downs of my university life, always stood up beside me and sometimes understanding me more than myself. I cannot list all the names here but you are always on my mind.

and

To myself,

Thank you for not giving up. You have done well and good work.

"If you can't fly, then run.
Today we will survive.
If you can't run, then walk.
Today we will survive.
If you can't walk, then crawl.
Even if you have to crawl, gear up"

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Dr. Nurul Hawani Binti Idris, for encouragement, guidance, critics and friendship. I am also very thankful to the other lectures who indirectly give their guidance, advices and motivation throughout my years of completing the research. Without their continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to Universiti Teknologi Malaysia (UTM) for accepting me to pursue my Master's study here. My fellow colleagues at UTM, Tropical Map and GNG research groups also deserve special thanks for their supports and advices. My sincere appreciation also extends to all my friends outside UTM campus who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. Also to the faculty staff, thank you for your help with all the procedure from the registration till the submission of my thesis. I am grateful to all those who have helped me in completing this thesis.

ABSTRACT

Today, linked data is frequently associated with Geographic Information Systems (GIS) as its technology stack is utilized in alleviating geospatial data integration issue. Geospatial data have become ubiquitous as they have emerged everywhere and these data can be geo-referenced. One of the types of georeferenced data that is lacking in Malaysia is the insufficient availability of Malaysian oceanographic data. It is a great relief to know that most of the earth observation agencies have granted access into their data obtained from satellite altimetry. Consequently, the exponential growth of geospatial data as well as its complexity and diversity has led to big data problem and caused information sharing and exchange on the web becoming more complicated. To resolve this issue, linked data should be used in handling geospatial big data. Linked data is one of the best practices for exposing, sharing, publishing and connecting the structured data on the web. This study explored linked data as an approach to provide access to the Malaysian physical oceanography datasets on the web, which would allow the data to be standardized in a machine-readable format. The research reviewed the existing software tools used in publishing linked data, identified an appropriate software tool to generate Resource Description Framework (RDF) presenting geographical data and built a physical oceanography data website based on linked data principles. Initially, document analysis was conducted to review the existing linked data tools that have been used for geospatial data. Various scholarly articles, journals, tutorials and web pages were used as references to investigate the use of linked data tools. Based on the review, five software tools, namely Geometry2RDF, TripleGeo, Datalift, OpenLink Virtuoso and KARMA were identified as the appropriate tools to generate the RDF. Each of this software tool has its own capabilities and functionalities. Next, the tools were compared with one another based on literature review to get the best possible tool that can manage georeferenced oceanographic data. After the comparison, this study identified the best software tool to transform the shapefile into the RDF format was Datalift. Finally, a web-based information system was built to publish the linked data to data interlinking and sharing by web users. In conclusion, this study has introduced an alternative way to publish and access geospatial data, particularly related to physical oceanography datasets using linked data principles. Using such an approach would facilitate stakeholders and unveil information within the big data, thus enriching the discovery of geospatial information on the web.

ABSTRAK

Pada masa kini, data terangkai sering dikaitkan dengan Sistem Maklumat Geografi (GIS) kerana kombinasi teknologinya digunakan dalam mengurangi isu integrase data geospatial. Kewujudan data geospatial ini tidak ada batasan memandangkannya muncul di mana-mana dan boleh digeoreferensikan. Salah satu jenis data georeferensi yang kurang di Malaysia adalah ketersediaan data oseanografi yang tidak mencukupi. Adalah sangat melegakan apabila kebanyakan agensi pemerhatian bumi telah memberikan akses kepada data mereka yang diperolehi daripada satelit. Kesannya, jumlah pertumbuhan eksponen data geospatial serta kerumitan dan kepelbagaiannya telah membawa kepada masalah data besar dan menyebabkan perkongsian dan pertukaran maklumat di laman sesawang menjadi lebih rumit. Untuk menyelesaikan masalah ini, data terangkai harus digunakan untuk mengendalikan data besar geospatial. Data terangkai adalah salah satu kaedah terbaik untuk mendedahkan, berkongsi, menerbitkan dan menghubungkan data berstruktur di laman sesawang. Kajian ini meneroka data terangkai sebagai satu pendekatan dalam menyediakan akses kepada dataset oseanografi Malaysia di laman sesawang yang membolehkan data diseragamkan dalam format yang boleh dibaca oleh mesin. Kajian ini mengkaji semula perisian sedia ada yang digunakan untuk menghasilkan data terangkai, mengenal pasti perisian yang sesuai untuk menghasilkan Rangka Keterangan Sumber (RDF) yang dapat mempersembahkan data geografi dan membangunkan laman sesawang bagi data oseanografi fizikal berdasarkan prinsip-prinsip data terangkai. Pada mulanya, analisis dokumentasi dijalankan untuk mengkaji semula perisian data terangkai yang sedia ada yang digunakan terhadap data geospatial. Pelbagai artikel, jurnal, tutorial dan laman sesawang digunakan sebagai rujukan untuk menyelidik kegunaan perisian data terangkai ini. Berdasarkan tinjauan tersebut, lima perisian, iaitu Geometry2RDF, TripleGeo, Datalift, Openlink Virtuoso dan KARMA telah dikenal pasti sebagai perisian yang sesuai untuk menghasilkan RDF. Setiap perisian ini mempunyai keupayaan dan fungsi yang tersendiri. Seterusnya perisian tersebut dibandingkan antara satu sama lain berdasarkan kajian literatur untuk mendapatkan perisian terbaik yang mampu menguruskan data oseanografi yang digeoreferensikan. Setelah perbandingan dilakukan, kajian ini mengenal pasti perisian yang terbaik untuk menukarkan fail bentuk kepada format RDF adalah Datalift. Akhirnya, sistem maklumat berasaskan laman sesawang telah dibina untuk menerbitkan data terangkai yang membolehkan perhubungan dan perkongsian data antara pengguna. Kesimpulannya, kajian ini memperkenalkan cara alternatif untuk menerbitkan dan mengakses data geospatial, terutamanya yang berkaitan dengan oseanografi fizikal yang menggunakan prinsip data terangkai. Pendekatan sedemikian dapat memudahkan pihak yang berkepentingan dan mendedahkan maklumat dalam data besar, sekaligus memperkayakan penemuan maklumat geospatial di laman sesawang.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASCII	-	American Standard Code for Information Interchange
BBC	-	British Broadcasting Corporation
BODC	-	British Oceanographic Data Centre
CSV	-	Comma Separated Value
ETL	-	Extract-Transform-Load
GIS	-	Geographic Information System
GML	-	Geography Markup Language
HDF	-	Hierarchical Data Format
HTML	-	Hypertext Markup Language
HTTP	-	Hypertext Transfer Protocol
INRIA	-	French Institute for Research in Computer Science and Automation
JSON	-	JavaScript Object Notation
KATS	-	Ministry of Water, Land and Natural Resources
KML	-	Keyhole Markup Language
MaCGDI	-	Malaysia Centre for Geospatial Data Infrastructure
MAMPU	-	Malaysian Administrative Modernization and Management Planning Unit
MESTECC	-	Ministry of Energy, Science, Technology, Environment and Climate Change
MyNODC	-	Malaysian National Oceanographic Data Center
MIMA	-	Maritime Institute of Malaysia
MyGDI	-	Malaysia Geospatial Data Infrastructure
NetCDF	-	Network Common Data Form
NOAA	-	National Oceanic and Atmospheric Administration
NOD	-	National Oceanography Directorate
NSDI	-	National Spatial Data Infrastructure
ODI	-	Open Data Initiative
OGD	-	Open Government Data
ORDA	-	Open Data Readiness Assessment

PSO	-	Particle Swarm Optimization
R&D	-	Research and Development
RADS	-	Radar Altimeter Database System
RDF	-	Resource Description Framework
SHP	-	Shapefile
SLA	-	Sea Level Anomaly
SMOS	-	Soil Moisture and Ocean Salinity
SPARQL	-	SPARQL Protocol and RDF Query Language
SQL	-	Structured Query Language
SST	-	Sea Surface Temperature
UNESCO	-	United Nations Educational, Scientific and Cultural Organization
URI	-	Uniform Resource Identifier
USGS	-	United States Geological Survey
UTM	-	Universiti Teknologi Malaysia
W3C	-	World Wide Web Consortium
WH	-	Wave Height
WS	-	Wind Speed
XML	-	Extensible Markup Language

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Since last few years, the ocean resources have been degraded and endangered by human activities such as overfishing, overdeveloped coastal areas, illegal wildlife trade and marine aggregates dredging that destroyed ocean ecosystems, organisms and habitats. In 2015, there were 253,103 metric tons trash fishes landing in Malaysia (Idris, 2017). These trash fishes are mainly caught from trawling and usually have either small or no market values as their quality was too poor to be served as human food but have been used in fish meal production. Thus, the continuous trawling activity has led to overfishing which caused the fish stocks gradually decreasing (Nguyen, 2012; Idris, 2017). Besides that, climate change also has been identified as one of the main threats to the marine environment and Southeast Asia region is facing the highest climate risk (MIMA, 2017). Hence, an effective ocean and coastal management are required to provide more sustainable future for the ocean.

Oceanographic data are important in providing information about the state of the ocean, its phenomenon and processes that occur surround it (Bastos *et al.*, 2016). The observation of these data is important for us to prepare an appropriate action when expecting any possible oceanic related events such as climate change and unseasonal weather to reduce the cost and impact of damages. Nevertheless, the study related to physical oceanographic data collection and management in Malaysia's marine science development was not getting too much exposure compared to the studies associated with the new techniques and technologies within marine pollution or marine biological domains (Ramli, 2008). There are many different methods using different instruments and technologies in collecting oceanographic data either from the seafloor, on land, throughout the water column, at the ocean surface or from the space (Smith and Johns, 2012) as shown in Figure 1.1.

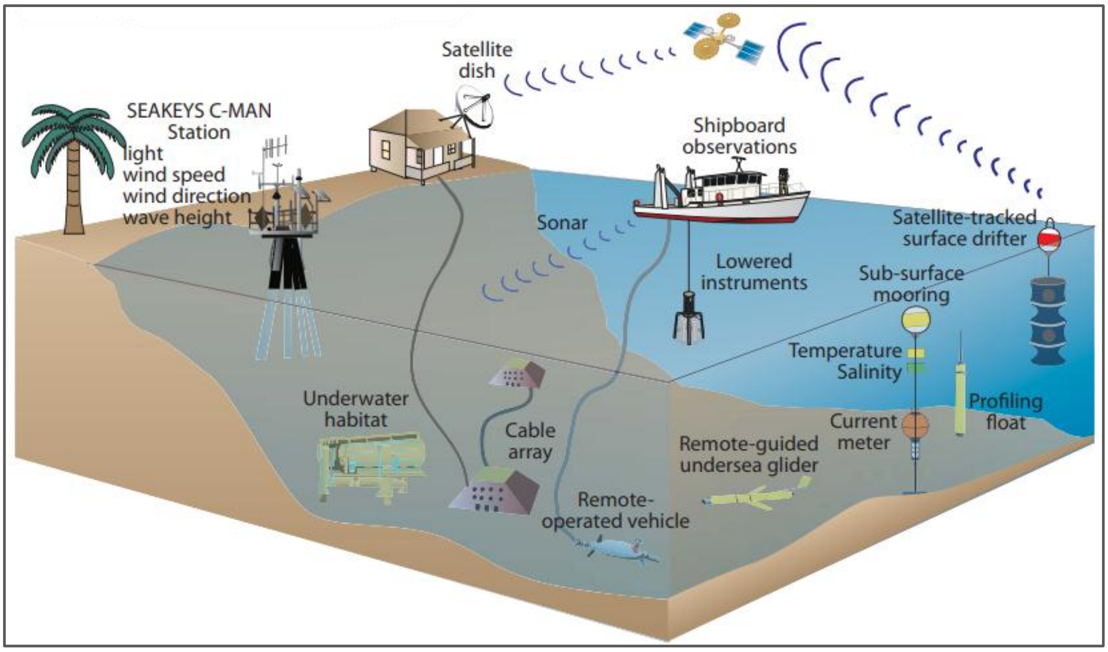

Figure 1.1 Methods collecting oceanographic data (Smith and Johns, 2012)

Figure 1.1 shows the oceanographic data such as sea temperature, salinity, wave height, wind speed and direction can be collected and observed in many ways. Besides traditional shipboard surveys, there are several methods that could provide a greater coverage and more economical way. For example, there are moored instruments arrays, surface drifter and remote sensing technology (Smith and Johns, 2012). Malaysia has conducted a number of groundwork observations on the ocean for long term projects that have particular purposes with limited areas covered. However, Ramli (2008) states that these observed data were inaccessible in most occasions as the observations were carried out for military, oil and gas industry or fisheries purposes. These datasets were stored separately according to the institutions that collected the data. In fact, most of the national institutions are not really contented with the data exchange process among them especially for real-time or near-real time data related to sea level information which were important in physical oceanography studies on marine environmental resource monitoring and prediction. Besides that, scientists and researchers also face difficulty to access and obtain the data as the datasets are not maintained and shared in a centralized portal.

In Malaysia, the Ministry of Energy, Science, Technology, Environment and Climate Change (MESTECC), formerly known as Ministry of Science, Technology and Innovation (MOSTI), has established National Oceanography Directorate (NOD) as the national focal point to instigate Research and Development (R&D) policy and agenda for Malaysia's oceanography domain. NOD's mission is to spearhead and exploit the national transcendence through innovations of marine science and oceanography (Mokhtar, 2010). NOD realizes that the government agencies, private sectors and institutions need a mechanism to access and share information among them. Consequently, NOD has initiated Malaysian National Oceanographic Data Centre (MyNODC) in 2010 to provide an open access and supervision for the national resource of oceanographic data (Mokhtar, 2012). MyNODC offers a platform that allows the stakeholders to interact, distribute and share their data and knowledge on oceanography and marine science research (Muslim, 2017). In order to promote the people to actively share their data in making the data to become more reliable and meaningful, the Director of NOD has indicated open data as a critical facilitator towards a sustainable ocean (Singh, 2016).

Besides MyNODC, Malaysia also started implementing Open Government Data (OGD) initiatives organized by the Malaysian Administrative Modernization and Management Planning Unit (MAMPU) which has developed a Malaysian's Open Data Portal in 2014. This portal serves as an official site that provides open data from Malaysian government and enables data sharing among various users that come from both the government and private sectors. Practically, most of the government agencies now have been equipped with the awareness and knowledge of this open data readiness assessment in Malaysia (Nurakmal *et al.*, 2017). The main data providers for the portal are the Department of Statistics Malaysia and the Malaysia Centre for Geospatial Data Infrastructure (MaCGDI). Besides providing open data access through the Malaysian's Open Data Portal, some of the agencies also allow the citizens to access and request the data directly from their websites. On January 2019, there are about 12,800 datasets provided in this portal including the data related to the Malaysian marine.

Apart from the government's open portals and platform, ocean related data are also accessible online. Altaweel (2018) states that most of the available online data are satellite data because of its availability, stored in various formats and coverage. Earth observation satellites are firmly used in observing the ocean to monitor the climate change. The data from satellite could provide essential climate variables affecting the characterization of Earth's climate (OECD, 2016). For example, satellite radar altimetry could measure the distance between the Earth's surface from the spacecraft while able to provide consecutive and precise measures of global sea level. There are several space agencies around the world that provide open access to their data such as National Oceanic and Atmospheric Administration (NOAA), United States Geological Survey (USGS), European Space Agency (ESA), National Aeronautics and Space Administration (NASA) and National Institute for Space Research (INPE). Nevertheless, satellite oceanographic data that are available online mostly provide raw data and can be considered as geospatial big data. Lee and Kang (2015) refer geospatial big data as big as 1 terabyte that exceed the capacity and capability of current computing systems.

Geospatial data have been considered as big data. Data with exceptionally large datasets and computational analytics practically defines as big data (Holt, 2017). Conventionally, geospatial data are categorized within three genres which are vector, raster and graph. Yet geospatial big data are omnipresent in all the genres. Geospatial big data in vector and graph types includes geo-located data of Facebook and Twitter, road or supply chain network data and taxi data. Spatial big data in raster type comprises of the satellite imagery, coordinated and multiple drones imaging system or climate simulation (Yaragal, 2018). The characteristics of the geospatial data fit the big data definition. Geospatial data including vector and raster data have the characteristics of multi-scale, multi-source, dynamic-state and high-dimensional features (Liu *et al*., 2018).

The geospatial big data has made sharing and data analytic become complicated. As the volumes of the geospatial data continually increasing, the current algorithms used in geospatial problem-solving are not competent and need to be optimized where the lacking of techniques, algorithms and specialized systems in supporting the geospatial big data become a challenge in geospatial world. Yaragal (2018) also states that most of the well-developed big data concept and tools such as Hadoop software, HBase, Hive, Spark and Map-Reduce techniques do not directly support big spatial data. The data are typically processed as non-spatial data or by using some wrapper function which might lengthen the processing time of the data. Thus, several products and extensions in Geographic Information System (GIS) have been introduced to handle geospatial big data in efficient manner such ESRI tools for Hadoop, Spatial Hadoop, Hadoop GIS and Parallel DB systems.

GIS has been around for fifty years to capture, store, manipulating and display the geospatial data (Chang, 2012). Spatial Data Infrastructure (SDI) is introduced to facilitate the access and availability of spatial data among GIS communities by implementing a framework that consists of technology standards, institutional arrangements and policies (Aalders and Moellering, 2001). However, the increasing amount and multiple formats of collected geospatial big data contribute challenges to the existing GIS tasks of storing, managing, processing, visualizing and verifying the quality of data (Li *et al*., 2016). Shekar *et al*. (2012) states that the volume, variety and update rate of dataset overwhelm the capability of spatial computing technologies. Since then, the advancement of GIS practices and researches have been diversified following the impressive geospatial data's growth and mature evolution of data processing technologies. Hence, the issues associated with geospatial data exchange, publishing and discovery have been one of the major focuses for further research. One of the solutions is on Semantic Web technology which allows anything to be published by anyone at any time, except that the published documents must be in the standard format in order to create typed statements that link unlimited things in the world (Bizer *et al*., 2009).

Semantic Web is not only about putting data on web, but it makes links to allow people or machines to explore the web. Occasionally, they can discover the other related data on web. It introduces new solutions for data sharing and exchange based on the web principles. World Wide Web Consortium (W3C) has proposed standards for data exchange on web which involves Linked Data and Semantic Web technologies (Iwaniak et al., 2016). Linked data and GIS are surely different paradigms in the realm of geospatial data. However, linked data approach is considered as one of the best practices to expose, share, publish and connect data on the web. Linked data has been introduced as an alternative method in geospatial analysis as it is benefited from semantic describing for both qualitative and quantitative characteristics of the spatial entities. Furthermore, most of the data semantics are available on the Web is stored in linked data format. As a part of Semantic Web, linked data is all about creating typed link, a link that computer machine could understand it when it is clicked to connect from one data source to another. Linked data has made a move into geospatial information domain and brings some new innovation in accessing and using geospatial data. It is able to link geographically referenced data to any data sources in enriching data integration across their domains (Kuhn et al., 2014).

1.2 Problem Statement

Ocean offers rich information about the marine life, waves, geology of the sea, ocean currents and many more. Numerous methods have been introduced to monitor and observe the ocean more openly alongside the collection of the oceanography data. Since the ocean is too wide, the measurements from the space have been notably advanced to answer unpredicted scientific question and applications about the ocean (Freeman *et al.*, 2010). Smith and Johns (2012) states that space-based technology is able to cover a large area and provides high resolution with longtime data series as the satellite passes across the ocean regions. In recent years, the number of satellite Earth observation missions that active are increasing as the missions allow the researchers to pay more attention to ocean colour as the crucial indicator to the state of Earth’s environment (Mikelsons and Wang, 2018). Yet, the aspect of Earth observation from space that provides a massive amount of the data that produced daily by the missions and how to provide the easy access to the most appropriate data for the various purposes has been a challenge.

As researchers and scientists continuously monitor the land use and global climate change, there is a huge demand on an open access of using geospatial data, hence several space agencies have allowed their observed satellite data to be accessed on the web (Altaweel, 2018). Usually, the open accesses of geospatial data, particularly satellite data, are shared according to the satellite mission's orbit phasing. While the number of Earth observing missions operated increases, the temporal and optical resolutions also increase. This expansion has caused the increased amount of produced data which caused problems regarding the storage and long-term preservation of the data (OECD, 2016). The gigantic amount of data that are collected over time has now led to the big data which poses challenges to be analysed, stored, processed, displayed and shared as the data are characterized in heterogeneous formats (Abburu *et al.*, 2014) such as in binary format, American Standard Code for Information Interchange (ASCII), Network Common Data Form (NetCDF) libraries or Hierarchical Data Format (HDF) files. These formats bring along the heterogeneous semantics.

The phenomenon of big data has completely drawn major attention these days. Demchenko *et al.*, (2013) have stated that big data are overtaking almost all aspects of human activity either in research, designing, production or digital services. Therefore, many new technologies and architecture have been built to solve the issue of capturing and processing such data. It is because the current techniques of managing the data in existing relational database are no longer effective to extract rich information from large size of data (Katal *et al.*, 2013; Narock and Hitzler, 2013). Relational database has been used widely in most organizations to store data digitally based on the relational model. The data are usually stored in a database table that consists of column and rows which are identified by unique primary key. Although, a relational database can handle the consistency issue within the transaction update functions quite well, it is struggling in managing big data as its scalability is not efficient to handle large volume of data (Reeve, 2012).

Large volume data are usually laborious to be handled and heterogeneous data also complex to be integrated with different types of data. This has contributed a difficulty among various users who want to use, manipulate and share the data for different purposes. Data integration becomes more complicated whereas information sharing and exchange will be limited and difficult to be shared among different communities. The ubiquitous network connectivity and cloud computing have been introduced to automate all the processes in collecting, storing, processing and visualizing the big data in one platform (Demchenko *et al.*, 2013). In Malaysia, most of the government's data are gathered in a centralized database through an open data portal to allow user access and utilize the data freely with applied terms and conditions regarding the data policy. There are about 250 data publishers and 12,827 datasets on this Malaysian Open Data Portal, yet only sea level rise data for the year of 2001-2009 are available on the physical oceanography domain. Other accessible related marine or ocean data are statistical data of fisheries, water quality, cargo ships registration and marine parks. In fact, MyNODC was the only close related platform that is established to allow the open access to oceanographic data (Mokhtar, 2012). Unfortunately, the platform has been temporarily closed due to certain reasons and constraints.

Since the availability of Malaysian physical oceanography datasets is insufficient in the government portals, there is a need to access these data openly on a web. These open data especially spatial related ones, are mostly extracted from the satellite observations, hence managing the data in one data centre is quite a challenge particularly on how to structure and share the data with other relevant big data on the Web. The traditional data repository is becoming much smaller and secluded as geospatial big data now become highly accessible on the web. As mentioned previously by Abburu *et al.*, (2014), the data retrieved from the satellites are facing big data issue as they are presented in large scale time series and stored in many types and formats. Apparently, the current GIS computer programs could not scale up with the volumes of data that continually grow. Therefore, the technology of semantic Web is considered as an ideal solution for addressing the big data challenges by enabling semi-automated alignment between data repositories by providing a platform to link data during publication (Narock and Hitzler, 2013). Besides seeing big data as a bad data collection, linked data approach can provide great paradigm to adapt a knowledge sharing environment by enabling information discovery through links and relationships (Holt 2017).

1.3 Research Goal

The purpose of this research is to explore linked data approach in publishing and accessing Malaysian physical oceanographic datasets on the web.

1.3.1 Research Objectives

The objectives of the research are:

(a) To review the existing software tools used in publishing linked data.

(b) To identify an appropriate tool in generating a proper Resource Description Framework (RDF) structure for presenting geospatial data.

(c) To develop physical oceanography online portal based on the linked data principles.

1.4 Scopes and Delimitation of the Study

This section describes the scope of this study that focuses on the method to publish oceanographic data on the web by using different linked data tools.

(a) Study area

Malaysia is surrounded by South China Sea, Straits of Malacca, Sulu Sea and Celebes Sea geographically. All these seas are part of the Pacific Ocean except the Straits of Malacca that is in the Indian Ocean region. This study area is delimited to the Straits of Malacca region as shown in Figure 1.2. This strait is located between the Peninsular Malaysia and The Sumatera (Indonesia) Island. It is considered as one of the busiest straits in the world. By 2020, the number of vessels navigating through this strait is expected to be more than 100,000 as the traffic's density are increasing at the average of more than six thousand per month (Awaludin and Abu, 2017).

Figure 1.2 Straits of Malacca map (Welt-Atlas.de, 2018)

(b) Data

The data used in this study are delimited to those obtained from the soil moisture and ocean salinity (SMOS) satellite mission and multi-mission satellite altimetry. The physical oceanographic data that are successfully derived comprise of daily and monthly time series data of sea level, sea surface current, sea surface temperature, salinity, significant wave height and wind speed over Malaysian seas. Among of these data, only four data are chosen to be used in this study. They are monthly sea level, sea surface temperature, wave height and wind speed. These data are chosen because of their high demand in the sustainable marine environment study. The data consist of 19 years data from 1993 until 2012.

(c) Software/Tools

This study used both commercial and open source software that associated with geospatial data. ArcGIS Desktop is the commercial mapping software that provided by the university and is used in data pre-processing for this study. In generating RDF data process, several open source tools are reviewed, compared and used based on the capabilities and functionalities of the software. The comparison has been made between of these tools:

i. geometry2rdf

ii. TripleGeo

iii. KARMA

iv. Openlink Virtuoso

v. Datalift

1.5 Significance of the Study

As Malaysia is located between the Pacific and Indian oceans, it is considered as a maritime country where its total area of sea coverage is 614, 159 square kms which is almost twice of the country's landmass (MIMA, 2018). Marine resources give great contribution to the Malaysian economic wealth as the petroleum and gas is the largest export commodity of this country (Khalid, 2005). So, it is important for the navy, armed forces, as well as the citizens to realize and start to develop a strategic thinking of the Malaysian future as a maritime nation (Tong, 2018). Hence, it is essential for Malaysia to enhance and sustain the efforts in balancing the maritime sector's growth parallel to the ecologically sustainable development of the marine resources and preservation of the marine environment. Khalid (2005) also states that the coastal area management, pollution prevention, ecosystems and biological conservation should be planned meticulously in protecting the marine resources and environment.

In order to manage the ocean and coastal areas efficiently, oceanographic data are used to provide information about the ocean condition and the events occur surrounds it (Bastos *et al*., 2016). These oceanographic data can be accessed from the government portal or space and marine agency sites which provide their data as an open data. Malaysia is the first country among Association of Southeast Asian Nations (ASEAN) to implement the World Bank's Open Data Readiness Assessment (ORDA), a methodological tool that evaluates the readiness of the private or government agencies and identify the action needed to make progress in the Open Data Initiative (ODI) program (World Bank, 2017). This initiative contributes to the implementation of a greater access to Malaysian official information that could help both the government and private institutions to make more informed decisions on how to invest and utilize resources to the fullest.

Thus, this study contributes to the realization of publishing geospatial big data on the web to improve data integration and enhance the relationship among different data sources through linked data approach. Linked data is considered as a good candidate to manage and formalize the relationship between the features as part of data integration process because of its flexibility and interoperability in spatial data fusion. Furthermore, as part of Semantic Web technologies, linked data approach is able to link SDI services to the established information technology to extend the branch of SDI's application (Schade and Smits, 2012). Therefore, the result of this study allows geospatial big data integration among different users including researchers, oceanographers and citizens who interested in accessing physical oceanographic data on web to utilize and manipulate the data according to their need.

1.6 Thesis Organization

This thesis outlines the study conducted for exploring linked data approach to access and publish physical oceanographic data on the web. There are five chapters divided in this thesis and each chapter describes the introduction, review of related literature, methods, findings, and conclusion of this study. Chapter 1 is an introductory chapter highlighting the research gaps in accessing physical oceanographic data in Malaysia. It includes the background of the study, problem statement, aims and objectives, scope and significance of this research. Chapter 2 presenting the literature review of the previous works related to this study in order to have a clear understanding in research background, methods used and any other theoretical findings. It is also to find unanswered questions from previous studies that need more findings. Then, the methodology to accomplish the aim and objectives of this study is carried out and described in Chapter 3. Chapter 4 displays and discusses the results from this study along with the procedure of the analysis used. Lastly, chapter 5 concludes this study with some inference and recommendations given for the further study on the research topic.

The next section provides a literature review for oceanography, geospatial big data, linked data and the tools for publishing physical oceanographic data on web as RDF.

REFERENCES

Aalders, H. G. L. and Moellering, H. (2001), 'Spatial Data Infrastructure', *Research and Development Center for Geo-Information Technology of Delft University of Technology*. Available at: http://www.gdmc.nl/publications/2001/Spatial_data_infrasructure.pdf (Accessed: 10 December 2018)

Abbott, M. (2013) 'From the President: The Era of Big Data Comes to Oceanography', *Oceanography*, 26(3), pp. 7–8. doi: 10.5670/oceanog.2011.65.

Abburu, S., Dube, N. and Nayak, M. R. (2014) 'Ocean Data Techniques and Standards Tools and Analysis Using Semantic Sensor Ontology An Emerging Area – Indian Scenario', *ISPRS International Symposium Commission TC VIII*, Hyderabad, India.

Alexakis, M. (2015) 'TripleGeo : An open-source tool for extracting geospatial features into RDF triples', *Notes from Institute for the Management of Information Systems*, Athena Research Center, Greece.

Altaweel, M (2018) 'Accessing Real-time Satellite Imagery and Data', *GIS Lounge*, 1st August. Available at https://www.gislounge.com/accessing-real-time-satellite-imagery/ (Accessed: 20 November 2018).

Andersen, O. B. and Scharroo, R. (2011) 'Range and Geophysical Corrections in Coastal Regions: And Implications for Mean Sea Surface Determination', Coastal Altimetry, pp. 103-145.

Awaludin, N. A. and Abu, M. I. H. (2017) 'Malaysia's Emergency Response In The Straits Of Malacca', *Maritime Institute of Malaysia*.

Bastos, L., Bio, A. and Iglesias, I. (2016) 'The Importance of Marine Observatories and of RAIA in Particular', *Frontiers in Marine Science*, 3, pp. 1–15. doi: 10.3389/fmars.2016.00140.

Berners-Lee, T. (2009) 'Linked Data - Design Issues', *World Wide Web Consortium (W3C)*, doi: 10.1201/b13877.

Berners-Lee, T., Fielding, R. and Masinter, L. (2005) 'Uniform Resource Identifier (URI): Generic Syntax Status'. Available at: https://tools.ietf.org/pdf/rfc3986.pdf (Accessed: 10 January 2016).

Bizer, C., Heath, T. and Berners-Lee, T. (2009) 'Linked Data - The Story So Far', *International Journal on Semantic Web and Information Systems*, 5(3), pp. 1–22. doi: 10.4018/jswis.2009081901.

Boneva, I., Labra Gayo, J. E. and Prud'hommeaux, E. (2017) 'Semantics and Validation of Shapes Schemas for RDF', *16th International Semantic Web Conference*, Vienna, Austria, pp. 104–120. doi: 10.1007/978-3-319-68288-4_7.

Chandler, C.L., Allison, M.D., Groman, R.C., West, P., Zednik, S. and Maffei, A.R. (2010) 'Developing an Ontology for Ocean Biogeochemistry Data', *AGU Fall Meeting*. San Francisco.

Chang, K.-T. (2012) 'Introduction to Geographic Information System', Sixth Edition, Singapore: McGraw Hill Higher Education.

Cheney, R. E. (2001) 'Satellite Altimetry', *Encyclopedia of Ocean Sciences*, 2nd Edition, pp. 58-64.

CIESIN, FAO, UNEP, WHO, and CGIAR (2004) 'Global Spatial Data and Information: Development, Dissemination and Use', *Global Spatial Data and Information User Workshop*, 21-23 September 2004, Lamont-Doherty Earth Observatory, Columbia University Palisades, New York, USA.

Clarke, R. A. (2006) 'Physical Oceanography in Atlantic Canada', *Proceedings of the Nova Scotian Institute of Science*, 43 (2), pp. 111-127.

De León, A., Saquicela, V., Vilches, L.M., Villazón-Terrazas, B., Priyatna, F. and Corcho, O. (2010) 'Geographical Linked Data: a Spanish Use Case', *Proceedings of the 6th International Conference on Semantic Systems*, p.36. doi: 10.1145/1839707.1839753.

Demchenko, Y., Grosso, P., De Laat, C. and Membrey, P. (2013) 'Addressing Big Data Issues in Scientific Data Infrastructure', *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems*, CTS 2013. IEEE, pp. 48–55. doi: 10.1109/CTS.2013.6567203.

Dimitrov, M. (2011) 'Semantic Technologies and Triplestores for Business Intelligence', *1st European Business Intelligence Summer School (eBISS 2011)*, pp. 139-155.

Din, A. H. M., Ses, S., Omar, K. M., Naeije, M., Yaakob, O. and Pa'suya, M. F. (2014) 'Derivation of Sea Level Anomaly Based on The Best Range and Geophysical Corrections For Malaysian Seas Using Radar Altimeter Database System (RADS)', *Jurnal Teknologi*, 71(4), pp. 83–91. doi: 10.11113/jt.v71.3830.

Falkowski, P. (2012) 'Ocean science: The power of plankton', *Nature*, 483(7387), pp. S17–S20. doi: 10.1038/483S17a.

Friedman, U. (2012) 'Big Data: A Short History', *Foreign Policy*, 8th October 2012. Available at: https://foreignpolicy.com/2012/10/08/big-data-a-short-history/ (Accessed: 13th January 2016).

Freeman, A., Zlotnicki, V., Liu, T., Holt, B., Kwok, R., Yueh, S., Vazquez, J., Siegel, D. and Lagerloef, G. (2010) 'Ocean Measurements from Space in 2025', *Oceanography*, 23(4), pp. 144–161. doi: 10.5670/oceanog.2011.65.

Goodwin, J., Dolbear, C. and Hart, G. (2008) 'Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web', *Transactions in GIS*, 12(1), pp. 19–30. doi: 10.1111/j.1467-9671.2008.01133.x.

Haag, J. (2015) 'Semantic Web Technology, Linked Data and RDF'. Available at: https://www.w3.org/community/xapivocabulary (Accessed: 22 July 2016).

Halevi, G. (2012) 'Special Issue on Big Data', *Research Trends*, 1(30), pp. 1–40. doi: 10.1080/07350015.2016.1197681.

Hallo, M., Luján-Mora, S. and Chávez, C. (2014) 'An Approach To Publish Scientific Data Of Open-Access Journals Using Linked Data', E*DULEARN14 Proceedings*, pp. 1145–1153.

Harris, R. (2010) 'HTTP Protocol'. Available at: https://tools.ietf.org/html/rfc2616 (Accessed: 22 July 2016).

Hart, G. and Dolbear, C. (2016) 'Linked data: a geographic perspective', *CRC Press*.

Heath, T. and Bizer, C. (2011) 'Linked Data: Evolving The Web Into A Global Data Space', *Synthesis Lectures On The Semantic Web: Theory And Technology*, 1(1), pp.1-136.

Hietanen, E., Lehto, L. and Latvala, P. (2016) 'Providing geographic datasets as linked data in SDI', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, Prague, Czech Republic, pp. 583–586. doi: 10.5194/isprsarchives-XLI-B2-583-2016.

Holt, I. (2017) ‘Geospatial Big Data’ *GIM International*, 26 June 2017. Available at: https://www.gim-international.com/content/article/geospatial-big-data-2 (Accessed: 08 April 2018)

Hyland, B. and Villazón-Terrazas, B. (2011) ‘Linked Data Cookbook’, *Government Linked Data (GLD) Working Group Wiki*. Available at: http://www.w3.org/2011/gld/wiki/ Linked_Data_Cookbook (Accessed: 15 July 2015).

Hyland, B., Atemezing, G. and Villazon-Terrazas, B. (2014) ‘Best Practices for Publishing Linked Data’, *W3C Working Group Note*. Available at: http://www.w3.org/TR/ld-bp/ (Accessed: 02 September 2016).

Hu, Y. and Li, W. (2017) ‘Spatial Data Infrastructures’, *The Geographic Information Science & Technology Body of Knowledge*, John, P. W. (ed.), doi: http://dx.doi.org/10.22224/gistbok/2017.2.1.

Idrees, M. O. (2015) ‘Challenges in Coastal Spatial Data Infrastructure implementation: A review’, *South African Journal of Geomatics*, 4(4), pp. 369-383.

Idris, M. SM. (2017) ‘Seeking A More Sustainable Future For Our Oceans’, *Free Malaysia Today*, Available at: http://www.freemalaysiatoday.com/category/ opinion/2017/06/07/seeking-a-more-sustainable-future-for-our-oceans/ (Accessed: 10 January 2018).

INSPIRE (2012) ‘Lift your data – hands on session’, *Datalift Workshop*, Instanbul.

Iwaniak, A., Kubik, T., Paluszyński, W., Tykierko, M. and Kaczmarek, I., (2012) ‘Semantic based extension of search capability of SDI’, *Global Geospatial Conference*, pp. 14-17.

Iwaniak, A., Kaczmarek, I., Strzelecki, M., Lukowicz, J. and Jankowski, P. (2016) ‘Enriching and improving the quality of linked data with GIS’, *Open Geosciences*, 8(1), pp. 323–336. doi: 10.1515/geo-2016-0020.

Kaisler, S., Armour, F., Espinosa, J.A. and Money, W. (2013) ‘Big Data: Issues and Challenges Moving Forward’, *2013 46th Hawaii International Conference on System Sciences*, pp. 995–1004. doi: 10.1109/HICSS.2013.645.

Katal, A., Wazid, M. and Goudar, R. H. (2013) ‘Big Data: Issues, Challenges, Tools and Good Practices’, *2013 Sixth International Conference on Contemporary Computing (IC3)*, Noida, India: IEEE, pp. 404–409. doi: 10.1109/IC3.2013.6612229.

Kepeklian, G., Bihanic, L. and Troncy, R. (2014) ‘Datalift: A Platform for Integrating Big and Linked Data’, *Big Data from Space (BiDS’14)*, Frascati, Italy, pp. 267–270. doi: 10.2788/1823.

Khalid, N. (2005) ‘Maritime economic activities in Malaysia: Planning towards sustainable development’, *Research Workshop on Environment and Rural Poverty: Linking Resources Use, Governance and Sustainable Policies*, 22 February 2005, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia,

Knoblock, C.A., Szekely, P., Ambite, J.L., Goel, A., Gupta, S., Lerman, K., Muslea, M., Taheriyan, M. and Mallick, P. (2012) ‘Semi-Automatically Mapping Structured Sources into the Semantic Web’, *Proceedings of the ESWC 2012 Workshop on Linked APIs,* pp. 375-390.

Konstantinou, N. and Spanos, D. E. (2015) ‘Materializing the web of linked data’, *Materializing the Web of Linked Data*, pp. 1–133. doi: 10.1007/978-3-319-16074-0.

Kuhn, W., Kauppinen, T. and Janowicz, K. (2014) ‘Linked Data - A paradigm shift for Geographic Information Science’, *Proceedings of The Eighth International Conference on Geographic Information Science (GIScience2014)*, pp. 173–186.

Kumar, D. and Shekhar, S. (2014) ‘Research Prospects in Spatial Data Infrastructure for Handling Energy Linked Disaster’s’, *IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSRJESTFT)*, 8(5), pp.53-67.

Kyzirakos, K., Vlachopoulos, I., Savva, D., Manegold, S. and Koubarakis, M. (2014) ‘GeoTriples: A tool for publishing geospatial data as RDF graphs using R2RML mappings’, *CEUR Workshop Proceedings*, 1401(611141), pp. 33–44.

Leadbetter, A., Arko, R., Chandler, C., Shepherd, A. and Lowry, R. (2013) ‘Linked Data: An Oceanographic Perspective’, *Journal of Ocean Technology*, 8(3), pp.7-12.

Lee, J. G. and Kang, M. (2015) ‘Geospatial Big Data: Challenges and Opportunities’, *Big Data Research*, 2(2), pp. 74-81.

Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A. and Cheng, T. (2016) ‘Geospatial big data handling theory and methods: A review and research challenges', I*SPRS Journal of Photogrammetry and Remote Sensing*, 115, pp. 119-133.

Liu, P., Di, L., Du, Q. and Wang, L., (2018) 'Remote Sensing Big Data: Theory, Methods and Applications', *MDPI Remote Sensing – Open Access Journal*, 10 (5), 711.

Masser, I. (2005) 'GIS Worlds: Creating Spatial Data Infrastructures', Vol. 338, Redlands, CA: ESRI Press.

Muslim, A. M. M. (2017) 'Ocean Data Management Update in Malaysia', *The 10th GEOSS Asia-Pacific Symposium*, Hanoi, Vietnam, 18-20 September 2017.

Mikelsons, K. and Wang, M. (2018) 'Interactive Online Maps Make Satellite Ocean Data Accessible', *Earth & Space Science News*, Available at: https://eos.org/project-updates/interactive-online-maps-make-satellite-ocean-data-accessible (Accessed: 30 December 2018).

MIMA (2017) 'Experts Examine Marine Environment Protection at the National, Regional, and Global Levels', *MIMA News*, 01 February. Available at http://www.mima.gov.my/news-list/174-experts-examine-marine-environment-protection-at-the-national,-regional,-and-global-levels.html (Accessed: 30 November 2018).

MIMA (2018), 'Malaysia's Maritime Resources', *Introduction to MIMA*. Available at: http://www.mima.gov.my/about-us/about-mima/mima-overview.html (Accessed: 13 November 2018)

Mokhtar, N.A. (2010) 'Oceanography Research Needs and Priorities', *Workshop on Climate Change and Biodiversity: Mobilizing The Research Agenda*, 13-14 December 2010, UKM Bangi.

Mokhtar, N.A. (2012) 'MyNODC: Spatial Data Info-Structure for Marine Spatial Data Planning', *Simposium Maklumat Geospatial Kebangsaan Ke-5*, 21 Mei 2012, Pusat Konvensyen Antrabangsa Putrajaya (PICC).

Moore, G. (2014) 'On the use of RDF in records management and archiving', *Research Report for Riksarkivet*.

Najar, C. and Giger, C. (2006) 'Spatial data and metadata integration for SDI interoperability', *International Journal of Spatial Data Infrastructures Research*.

Narock, T. and Hitzler, P. (2013) 'Crowdsourcing semantics for big data in geoscience applications', *AAAI Fall Symposium - Technical Report*. Arlington, Virginia.

Nguyen, K.T. (2012) 'What Are The Consequences of Overfishing in WestAfrica, And How Can Sustainable And Flourishing Fisheries Be Promoted?'. *Master Thesis*. University of Agder; 2012.

NOAA (2009) 'How many oceans are there ?', *National Oceanic and Atmospheric Administration*. Available at: https://oceanservice.noaa.gov/facts/howmanyoceans.html (Accessed: 03 January 2016).

Nurakmal, M., Hanum, F., and Hamid, S. (2017). 'Post-adoption of Open Government Data Initiatives in Public Sectors', *Twenty First Pacific Asia Conference on Information Systems*, Langkawi, 2017.

OECD (2016) 'Could big data from satellites play a major role in climate change management?', *Chapter 4 of Space and Innovation*, OECD Publishing. Available at: http://dx.doi.org/10.1787/9789264264014-en (Accessed: 13 May 2017)

Open Geospatial Consortium (OGC) (2015) 'Geometry2RDF'. Available at: http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/technologies/151-geometry2rdf/index.html (Accessed: 02 May 2016).

Patroumpas, K., Alexakis, M., Giannopoulos, G. and Athanasiou, S. (2014) 'TripleGeo : an ETL Tool for Transforming Geospatial Data into RDF Triples', *Workshops on EDBT/ICDT 2014 Joint Conference*, Athens, Greece, pp. 275–278.

PCAST (2014) 'Big Data and Privacy: A Technological Perspective', *The President's Council of Advisors on Science and Technology (PCAST)*.

PO.DAAC (2018) 'Ocean Surface Topography', *Physical Oceanography Distributed Active Archive Center*. Available at: https://podaac.jpl.nasa.gov/OceanSurfaceTopography (Accessed: 11th November 2018)

Ramli, M. R. (2008) 'Physical oceanography activities in Malaysian waters', *Hydro International*, 6(8), pp. 30–33.

Reeve, A. (2012) 'Big Data and NoSQL: The Problem with Relational Databases', *Big Data*, 7th September. Available at:

https://infocus.dellemc.com/april_reeve/big-data-and-nosql-the-problem-with-relational-databases/ (Accessed: 20 December 2015).

Révillion, C., Bouche, D., Catry, T., Padeau, J., Guyard, S., Kazmierski, M., Faure, J.F., Desconnets, J.C., Brou, Y.T. and Sand, A. (2015) 'SEAS-OI Spatial Data Infrastucture: A tool for the diffusion of satellite imagery in the South Western Indian Ocean', *ISPRS Geospatial Week 2015*.

Schade, S. and Smits, P. (2012)'Why Linked Data Should Not Lead to Next Generation SDI', *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pp. 2894–2897.

Scharffe, F., Bihanic, L., Képéklian, G., Atemezing, G., Troncy, R., Cotton, F., Gandon, F., Villata, S., Euzenat, J., Fan, Z. and Bucher, B. (2012) 'Enabling Linked Data Publication with The Datalift Platform', *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Sengupta, I. (2014) 'Spatial Data Infrastructure-An Evolutionary Approach to GIS Technologies', *3rd International Conference on System Modelling & Advancement in Research Trends (SMART)*, College of Computing Sciences and Information Technology (CCSIT), Teerthanker Mahaveer University, Moradabad. pp. 345-349.

Shekhar, S., Gunturi, V., Evans, M. R., and Yang, K. (2012) 'Spatial Big-Data Challenges Intersecting Mobility and Cloud Computing', *Proceedings of the 11th ACM International Workshop on Data Engineering for Wireless and Mobile Access - MobiDE '12*, pp. 1-6.

Singh, I. (2016) 'Open ocean, open data for Malaysia', *Malaysia Oceanography: Open data leads to sustainable development*, 27th November. Available at: https://www.geospatialworld.net/blogs/malaysia-oceanography-open-data-leads-to-sustainable-development/ (Accessed: 07 August 2017).

Smith, R. H. and Johns, E. M. (2012) 'Oceanographic data are collected in many different ways', in Kruczynski W.L. and Fletcher P.J., (eds.) *Tropical Connections: South Florida's marine environment*. Chapter 2, Oceanographic Connectivity, pp. 66.

Souza, A., Corcho, O., Salles, P. and Vilches-Blázquez, L. (2014) 'An application of geographical and statistical linked data to ecology: The Brazilian cerrado ontology network and qualitative reasoning models', *Proceedings - 7th*

International Congress on Environmental Modelling and Software: Bold Visions for Environmental Modeling, iEMSs 2014, San Diego, California.

Stadler, C., Lehmann, J., Höffner, K. and Auer, S. (2012) 'LinkedGeoData: A core for a web of spatial open data', *Semantic Web*, 3(4), pp.333-354.

Stewart, R. H. (2008) 'Introduction To Physical Oceanography', *American Journal of Physics*. Texas A & M University.

Sullivan, D. (2017) 'The Limits of Relational Database', Advanced NoSQL for Data Science, 14th February. Available at: https://www.linkedin.com/learning/advanced-nosql-for-data-science/the-limits-of-relational-databases (Accessed: 30 September 2017).

Tole, A. A. (2013) 'Big Data Challenges', *Database Systems Journal*, IV(3), pp. 31–40.

Tong, L. C. (2018) 'Human Capital Transformation', *Royal Malaysian Navy's Seminar*, 27th November. Available at: https://www.liewchintong.com/2018/11/27/malaysia-as-a-maritime-nation/ (Accessed: 11 January 2019)

USGS (2014) 'How Much Water is There On, In, and Above the Earth? ', *The USGS Water Science School*. Available at: http://water.usgs.gov/edu/earthhowmuch.html (Accessed: 11 November 2016).

Vilches-Blázquez, L.M., Villazón-Terrazas, B., Saquicela, V., de León, A., Corcho, O. and Gómez-Pérez, A. (2010) 'GeoLinked data and INSPIRE through an application case', *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 446-449. ACM.

Welt-Atlas.de (2018) 'Map of Strait of Malacca', Available at: https://www.welt-atlas.de/map_of_strait_of_malacca_6-847 (Accessed: 28 April 2018).

Wiemann, S. and Bernard, L., (2016) 'Spatial Data Fusion in Spatial Data Infrastructures Using Linked Data', *International Journal of Geographical Information Science*, 30(4), pp. 613-636.

World Bank (2017) 'Malaysia to Use Open Data to Improve Public Service Delivery and Promote Innovation', *Press Release*, 25th May. Available at: http://www.worldbank.org/en/news/press-release/2017/05/25/malaysia-to-use-open-data-to-improve-public-service-delivery-and-promote-innovation (Accessed: 08 March 2018).

Yahaya, N. A. Z., Musa, T. A., Omar, K. M., Din, A. H. M., Omar, A. H., Tugi, A., Yazid, N. M., Abdullah, N. M. and Wahab, M .I. A. (2016) 'Mean Sea Surface (MSS) Model Determination for Malaysian Seas Using Multi-Mission Satellite Altimeter', *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, pp. 247-252.

Yaragal, S. (2018) 'Big Data in GIS Environment' *Geospatial World*, 10th February. Available at: https://www.geospatialworld.net/blogs/big-data-in-gis-environment/ (Accessed: 30th November 2018).

Zhang, Y., Chiang, Y. Y., Szekely, P. and Knoblock , C. A. (2013) 'A Semantic Approach to Retrieving, Linking, and Integrating Heterogeneous Geospatial Data', *Joint Proceedings of the Workshop on AI Problems and Approaches for Intelligent Environments and Workshop on Semantic Cities - AIIP '13*, pp. 31–37.