Big Data Management in the Shipping Industry:
Examining Strengths Vs Weaknesses and Highlighting Relevant Business Opportunities

Dimitrios Dalaklis[1], Georgios Vaitsos[2], Nikitas Nikitakos[3], Dimitrios Papachristos[3], Angelos Dalaklis[4], Esslam Hassan[5]

[1] *World Maritime University, Maritime Safety and Environmental Administration, Malmo-Sweden*
[2] *University of West Attica, Depart. of Industrial Design and Production Engineering, Athens-Greece*
[3] *University of the Aegean, Department of Shipping, Trade and Transport, Chios-Greece*
[4]*University of Reading, Henley Business School, Reading-United Kingdom*
[5]*Egyptian Authority for Maritime Safety, Alexandria-Egypt*
Corresponding author's email: esslamfattah@gmail.com

Abstract

History testifies that there is a dialectic relationship between humans and technology. Especially during the last couple of decades, the shipping industry has benefitted from a very extended number of advanced technology innovations. Today, all systems supporting the conduct of navigation and the various information technology (IT) applications related to ship management activities are heavily reliant upon (almost) real-time information to safely/effectively fulfil their allocated tasks. As a result, truly vast quantities of data -which are often described as "Big Data" in the wider literature- are created and the issue of how to effectively manage all the associated information is clearly standing out. Furthermore, topics such as optimising the conduct of all relevant activities on-board the vessel at sea, identifying the right opportunities in order to further promote business and boost profits, or even contributing to the numerous elements of sustainability by achieving reductions in energy consumption and/or a better environmental footprint for shipping, should all be researched further.
Considering the quite limited capacity of the human brain to process really enormous quantities of data in comparison to modern computers, the trend to use advanced software tools for extracting and processing the "right" information that is often hidden in the vast pool of Big Data, as well as relying on advanced techniques and algorithms to perform the relevant statistical analysis becomes quite obvious. The purpose of this paper, which follows a qualitative approach working in unison with a "Strengths, Weaknesses, Opportunities, and Threats" (SWOT) analysis, is to identify and briefly discuss the most relevant tools and techniques that are associated with Big Data Management. It will also clearly highlight the various benefits that are opening up and will try to explain the notion behind this transition to a new era, characterized by the term "smart shipping". A very important conclusion is that the exploitation of Big Data and the role of certain software applications in accessing and managing this large volume of information are key factors for improving/optimising the conduct of ship operations and management; establishment of a "Data Driven Culture" within a shipping company can clearly improve the current business model and at the same time promote sustainability.

Keywords: *Big Data Analytics, SWOT Analysis, Data Driven Culture, Optimisation, Shipping Industry.*

1. Introduction

The modern business environment is characterized by continuous changes and fierce competition. This situation is creating a pressing need for companies to identify and quickly adopt the right "means" for ensuring their survival, as well as promoting their further development and consolidation within the markets they operate. Informatics, from the very beginning of their inception, have proven to be a powerful tool at the disposal of companies that want to improve their business model, as they can be used as the main strategic enabler for integrating changes in the company's internal structure, functions and processes. Especially in the maritime sector, a large volume of data is produced from a very extended pool of relevant sources (i.e. systems supporting the conduct of navigation and/or ship's machinery, as well as related marine fleet management systems etc.), on a daily basis. The domain of

"Big Data Analytics" examines large amounts of data to uncover hidden patterns, correlations and other insights (i.e. market trends and customer preferences) that can help organizations make informed business decisions; it can be categorized as a special branch of the wider information technology (IT) domain and its main aim is to discover correlations and interactions between different measurable or non-measurable parameters, in order to identify non-clearly defined standards and patterns (Goyal et al., 2020). This research effort aims to clearly highlight that Big Data Analytics have the potential to create a very positive impact upon the shipping industry. To achieve this, a qualitative approach is deployed, working in unison with a "Strengths, Weaknesses, Opportunities, and Threats" (SWOT) analysis. This combinatory methodology will allow to identify and briefly discuss the most relevant tools and techniques that are associated with effective Big Data Management. Furthermore, the various benefits in terms of improving the current prevailing business model will be explained, indicating that the transition to a new era, characterized by the term "smart shipping", has already started.

2. Background

2.1 Big Data in the Maritime Sector

The contemporary era is frequently referred to as "the information age"; it is therefore not a coincidence that modern economic activities are very highly dependent on data. It is quite obvious that the volume of (stored and processed) data has grown exponentially over the course of time and this trend is recorded in literally all economic sectors and related activities. In order to provide a relevant definition, the McKinsey Global Institute, is approaching "Big Data" as those data-sets with a size that exceeds the ability of traditional database software tools to collect, store, manage and analyze these data (Manyika et al., 2011; Saxena, 2016; Al-Sai et al., 2019). This very significant growth in accessibility of data, storage capacity, and computational power has impacted businesses throughout the world. This change involves not only very popular and well known businesses like Yahoo or Facebook that were "*were born and function solely online"*, but also early adopters in more conventional industries such as banking, retail and transport. In the near future, a really impressive data growth is expected because of the related developments in remote sensors, as well as functions like communications, computations and processing activities of an "interconnected world" that will involve very large data collections/handling (Davenport and Harris, 2007; Goyal et al., 2020). It is also true that there are various definitions of Big Data. The prevailing version is usually different from one industry to another, and clearly depending on the type of available software tools and the sizes of datasets that are in common use within that specific industry (Al-Sai et al., 2019).

In simple terms, the notion of Big Data translates into a dataset that continues to grow till it becomes very difficult to effectively manage it by using the prevailing (standard) resources of database management. Data collection, processing, search, sharing, analytics, and visualization can all be viewed as the possible reason of this complexity and ineffective management (Manyika et al., 2011). The META Group analyst Doug Laney (now Gartner) defined data growth challenges and opportunities as being three-dimensional (volume, velocity, and variety) in 2001 (Esteves and Curto, 2013). After just two years, this definition was updated by Gartner: "*high-volume, high velocity and/or high variety information assets that demand cost-effective innovative forms of information processing for enhanced insight, decision making, and process optimization*" (Zainal et al., 2017). SAS also defined Big Data as "*Popular term used to describe the exponential growth, availability, and use of information, both structured and unstructured*"; IBM added more input towards the Big Data concept: "*Data, coming from everywhere; sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction record, and cell phone GPS signal to name a few*", "*Big Data is defined as large set of data that is very unstructured and disorganized*", "*Big Data is a form of data that exceeds the processing capabilities of traditional database infrastructure or engines*" (Al Nuaimi et al., 2015; Quintero, 2015; Al-Sai et al., 2019).

A very large number of researchers and associated practitioners approach Big Data as a term that describes large volumes of high velocity, complex/variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information. This domain is characterized as a new generation of technologies and architectures, designed to economically extract value from massive volumes of a wide variety of data, by enabling high-velocity capturing, discovery, and analysis (Esteves and Curto, 2013). Several studies, such as those by Manyika et al., (2011), Widyaningrum (2016), Bronk and Khan (2017), as well as Drus and Hassan (2017), have characterized Big Data by a multi-V (4Vs) model (volume, variety, velocity and value). At the same time, other approaches emphasize the so-called 5Vs model (Fig. 1). Specifically, the Five Vs (volume, velocity, variety, veracity, and value) exhibit certain differentiating characteristics that have been formulated by the difference between using traditional data and effectively exploiting Big Data (Saxena, 2016).

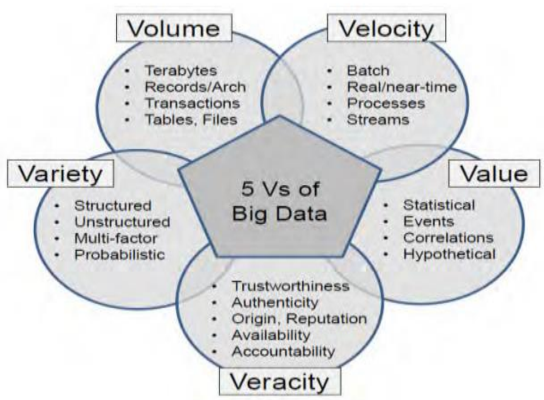

Fig. 1. Big Data – 5Vs definition model (Ishwarappa, & Anuradha, 2015)

The 5Vs model is summarizing the basics of all those characteristics that clearly define Big Data. Volume-Velocity-Variety are standing out first, with Veracity-Value coming next (Chen and Zhang, 2014; Ishwarappa and Anuradha, 2015). A short description of the above mentioned term is described with the help of Fig.1, along with some additional clarifications that follow: a) *Volume:* Big Data volume refers to the data storage space. Storage space for such data imposes Terabyte or Petabyte requirements; b) *Velocity:* It refers to the rate at which data is generated or processed and stored and further analyzed in real time; c) *Variety:* Data can be stored in various formats such as databases, excel (or csv) files, or even as simple text files. Sometimes the data under discussion does not follow a traditionally tabular format; basic types of this data are divided into structured, semi-structured and unstructured. The greatest challenge for shipping companies is to handle properly these large volumes of unstructured or semi-structured data and manage to convert them into a structure format; d) *Veracity:* It makes sense that among all this data there is "*dirty data"*, which should either be corrected or excluded from the analysis process. The quality of the data stored is important, as the analysis of erroneous data can lead to invalid or unreliable results. The main factor that affects the veracity of the data is their source; e) *Value:* this characteristic was added later on the features of Big Data definition, but it is still very important. The potential value of Big Data is truly huge and it is associated with the fact that it can help companies to identify new opportunities.

Big Data analysis is performed where advanced data techniques and technologies work in large sets of data, as defined above, in order to deconstruct complex information and semi-structured or non-structured data into structured information (Goyal et al., 2020). Specifically, Big Data Analytics (BDA) is a process for examining the variety of data with the aim to improve (or even optimize) relevant decision-making. Indicative benefits of BDA are: a) Reduction of costs. Big data technologies such as Hadoop and cloud-based analytics offer substantial cost savings when it comes to processing large amounts of data; b) Really effective, better decision-making. By combining Hadoop's speed and in-memory analysis the ability to evaluate new data sources is provided; companies can then quickly analyze knowledge and make "better" decisions based on what they have found; c) Improving provision of goods and services. Through the use of analytics, the designated analysts can measure and better define the needs of the customers; improved, or completely new products can be created to effectively meet the needs of consumers.

2.2 Big Data Applications

The volume of data coming from vessels is truly huge. It is indicative for someone to just think the large amounts of data from the Marine Traffic portal that indicate the exact locations of all vessels, at the worldwide level. Also, effective implementation of existing maritime regulations requires continuous data from vessels, which must be collected, stored in a format that allows monitoring of the vessel, such as position/location, speed, course, the weather at the respective location or even data generated each time in real time and collected by sensors such as main engine telemetry data and many more. Data recording must be continuous and maintained on vessel's applications, offering detailed information that will contribute into "easier" decision making. An additional feature of Big Data is its wide variety, as it is stored in many different formats. Data coming from vessels can be collected from dozens of different devices and with different formats. Accuracy and validity in shipping systems may have evolved considerably, but the risk of some data errors remains. An incorrect measurement, or an incorrect entry, can lead to erroneous results and consequently to the wrong decision (Rodseth et al., 2016; Zaman et al. 2017). Sources of Big Data on board a vessel are (Rodseth et al., 2016): Bridge data network; Conventional ship's automation; New cyber-physical systems; Ship performance monitoring systems; Ship voyage monitoring and reporting systems; External ship monitoring (AIS and VTS);

Weather data, among others. Data associated with ship operations can be collected from various sources. However, in order to extract the "right" information and be able to make meaningful decisions, the stage of analysis is essential; a rigorous (Big Data) analysis is crucial. The Big Data value chain also consists, in the case of shipping data, of collecting, processing and storing data, analyzing it with innovative and cost effective technologies and techniques, and formatting it in such a way as to either create a better understanding, or improve insight and decision making (Rodseth et al., 2016).

The adoption of Big Data Analytics is facilitated by the willingness of the shipping industry to move from a traditional culture to its operational management into a new Data Driven culture, as it is moving forward towards digitalization. According to various sources (IMO 2014; ISO 2015a, 2015b; IEC, 2015; Rodseth et al., 2016), indicative areas of interest include (but, are not limited to): a) *Vessel Performance Monitoring*. The vessel's energy efficiency management plan requires vessels to collect information on the vessel's performance and conduct of navigation from various on-board systems and associated Data Acquisition Systems. These systems are designed to collect, store and communicate large amounts of data relating to vessel performance and navigation data through complex processes. Data obtained from analytics allows a better understanding of the vessel's performance. Data on, for example, fuel consumption may show that in extreme weather conditions a significant reduction in speed can lead to disproportionate fuel reduction, leading the vessel to consume more fuel to move and indicate a change in speed at the point of optimal consumption. Data analytics in relation to a specific vessel might also "predict" a repair or a possible malfunction; b) *Navigation Data*. In 2000, the International Maritime Organization (IMO) defined the need for an Automatic Identification System (AIS), capable of automatically providing information about the vessel to other (nearby) vessels as well as to relevant authorities ashore. Since then, that data have been created and turned into a common object of research in the field of marine studies. This data mainly consists of the ship's identity, location, speed and direction. Big data analytics could provide a better understanding of the vessels' movements, indicating, for example, the most popular points of trade or helping them to avoid collisions and ensure a better level of safety at sea; c) *Green Shipping*. IMO has introduced several regulations regarding the reduction of emissions and improvement of energy efficiency. Big Data analytics can provide important information and lead to optimizations in relation to vessels' emissions, allowing a deeper understanding of the problem; d) *Safety*. Another area of interest for the world of Big Data analytics is that of safety in the operational management of the vessel. The information generated clearly creates new perspectives on maritime risk management and accident prediction; e) *Autonomous or Remote Controlled Ships*. BDA could provide a better decision-making for people performing remote control and allow for a smoother integration of these types of vessels.

3. Methodology and Findings

A SWOT analysis was used as the research tool. SWOT Analysis can be used both for strategic planning and strategic management; it can be used to build organizational strategy and competitive strategy (Gurel and Tat, 2017; Sammut-Bonnici and Galea, 2014). Thompson et al. (2007) point out that "*SWOT Analysis is a simple but powerful tool for sizing up an organization's resource capabilities* *and deficiencies, its market opportunities, and the external threats to its future*”. The acronym SWOT stands for “strengths, weakness, opportunities and threats”. The SWOT Analysis, also referred to as “SWOT Matrix”, can also be formulated as “TOWS Analysis” or “TOWS Matrix” (Gurel and Tat, 2017). Specifically, this internal analysis can be used to identify resources, capabilities, core competencies, and competitive advantages inherent to an organization. Similarly, the external analysis identifies market opportunities and threats by looking at competitors’ resources, the industry environment, and the general environment. The objective of a SWOT analysis is to use the knowledge an organization or area or field has about its internal and external environments and to formulate its strategy accordingly (Sammut-Bonnici and Galea, 2014). As internal and external sources of the SWOT analysis, input of relevant papers from google scholar and science direct databases was used. Maritime environment related keywords were used to search articles for a period of the last ten (10) years, to ensure the relevance of the identified academic papers. The term “maritime” was always included in the search, along with the key terms “Big Data” and “Big Data Analytics”. Certain abbreviations and/or combinations were also used, such as BDA in shipping, AIS and big data etc. Fig. 2 demonstrates the methodology approach; a summary of the findings is provided via Tab. 1:

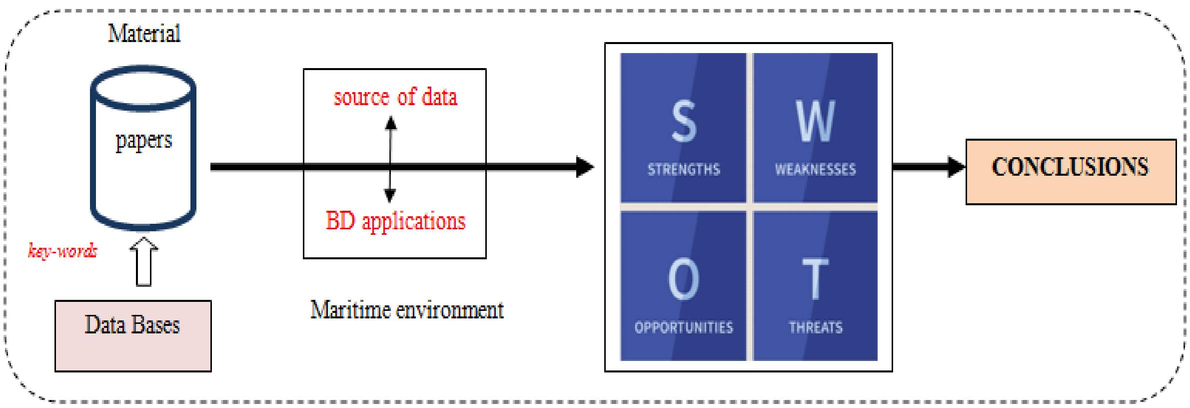

Fig. 2. Methodology Framework

Tab. 1. Final Results

INTERNAL ENVIRONMENT	EXTERNAL ENVIRONMENT
Strengths Data Quality Consistency Data Reliability Data Availability Data Confidentially Data Set Scalability	***Opportunities*** Data protection Business Model Human Factors and Practice AI using (machine & deep learning) IoT application Energy Management Environmental legislation monitoring Performance management Autonomous ship

Weakness	***Threats***
Data management Data transfer Accidents (from IT errors)	Hackers / cybersecurity Data ownership Ethics issues (decision making from autonomous shipping or smart shipping)

4. Conclusion

It is a self-explanatory fact that the volume of data is increasing over the course of time and at the same time there is also an increase in the related speed of transfer (upload and download) and its diversity. The value that is “hidden” in these huge volumes of data, when properly discovered and verified, can provide access to improved knowledge in relation to the business environment and even lead to the notorious “competitive advantage” over other competitors. Shipping companies can no longer rely their business decision-making processes in anachronistic ways (old mode of “paper and pen”)", but on the contrary, it seems quite important to invest in new technologies and techniques that will ensure the optimization of their processes and associated results. Big Data is impacting our daily lives for good with applications like Facebook and various Google services; huge volumes of data (from various sources) are being processed and analyzed in relation to topics like operations of vessels at sea, loading/unloading their cargo, as well as to serve the needs of fleet management. Shipping, despite its peculiarities in relation to companies in other sectors, should view as a priority the best possible utilization of this type of data. It is more than clear that truly vast volumes of data are obtained from both in-vessel systems such as bridge data or on-board automated systems. Additional examples include vessel performance monitoring data from sensors on the vessel or relevant daily reports, as well as external sources like AIS and/or weather data. That data can be collected, processed, analyzed and stored in such a way as to provide the “right” information.

All the above can be facilitates by big data analytics. Following this approach, shipping companies can extract, with a relatively low cost of implementation, meaningful information and improve their decision-making in areas like reduction in fuel consumption and improving the vessels’ environmental footprint. In addition, the use of spatial-temporal data (i.e. vessel identity, location, speed, direction, etc.) can provide opportunities for a better risk management and even contribute into accident prediction. The exploitation of Big Data and the role of certain software applications in accessing and managing this large volume of information are key factors for improving/optimizing the conduct of ship operations and management; establishment of a “Data Driven Culture” within a shipping company can clearly improve the current business model and at the same time promote sustainability. Big Data analytics can be deployed as a powerful tool and facilitate the transition towards “smart shipping” and at the same time help the shipping companies that perform the investment and effort to enjoy the benefits of a more “safe” and more “green” operating environment. In any case, topics such as optimizing the conduct of all relevant activities on-board the vessel at sea, or contributing to the numerous elements of sustainability by achieving reductions in energy consumption and/or a better environmental footprint for shipping, should all be researched further.

References

Al Nuaimi, E. Al Neyadi, H. Mohamed, N. and Al-Jaroodi J., "Applications of big data to smart cities," *J. Internet Serv. Appl.*, vol. 6, no. 1, p. 25, 2015.

Al-Sai, Z. A. and L. M. Abualigah, "Big data and E-government: A review," *2017 8th Int. Conf. Inf. Technol.*, pp. 580–587, 2017.

Al-Sai, Z.A. Abdullah, R. Husin, M.H. (2019). Big Data Impacts and Challenges: A Review. 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), pp. 150-5.

Braun, H. "Evaluation of Big Data Maturity Models: A Bench-Marking Study to Support Big Data Maturity Assessment in Organizations," 2015.

Brock, V. and Khan H.U., "Big data analytics: does organizational factor matters impact technology acceptance," *J. Big Data*, vol. 4, no. 1, p. 21, 2017.

Davenport, T.H. and Harris J.G., Competing on Analytics: The New Science of Winning. Harvard Business School Press, 2007.

Chen, P.C. and Zhang, C.Y. (2014). "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences,* 275, 314–347

Drus M. and Hassan N.H., "Big Data Maturity Model – a Preliminary Evaluation," *ICOCI Kuala Lumpur. Univ. Utara Malaysia*, no. 117, pp. 613–620, 2017.

Esteves, J. and Curto J., "A risk and benefits behavioral model to assess intentions to adopt big data," *J. Intell. Stud. Bus.*, vol. 3, no. 3, pp. 37–46, 2013.

Goyal, D. Goyal, R. Rekka, G. Malik, S. and Tyagi, A.K. (2020). Emerging Trends and Challenges in Data Science and Big Data Analytics. 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), pp.1-8.

Gurel, T. and Tat M. (2017). SWOT Analysis: A Theoretical Review. *The Journal of International Social Research*, Vol.10, Issue 51,

IEC (2015), *Maritime navigation and radio communication equipment and systems - Digital interfaces - Part 460: Multiple talkers and multiple listeners - Ethernet interconnection - Safety and security,* IEC 61162-460.

IMO (2014), Sub-committee on Navigation, Communications and Search and Rescue, Report to the Maritime Safety Committee. NCSR 1/28, Annex 7: *Draft e-Navigation Strategy Implementation Plan*

ISO (2015a), ISO/NP 19847 *Shipboard data servers to share field data on the sea*, ISO/TC 8 / SC 6N 359.

ISO (2015b), ISO/NP 19848 *Standard data for shipboard machinery and equipment of ship*, ISO/TC 8 / SC 6N 360.

Ishwarappa, K. and Anuradha, J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. *Procedia Computer Science, 48, 319–324.*

Lloyds Register, QinetiQ, University of Southampton, Global shipping technology 2030, UK, 2015.

Malik, P. "Governing Big Data: Principles and practices," *IBM J. Res. Dev.*, vol. 57, no. 3/4, p. 1:1-1:13, 2013.

Manyika J., Chui M, Brown B., Bughin J., Dobbs R., Roxburgh C., and Hung Byers A, "Big data: The next frontier for innovation, competition, and productivity," *McKinsey Glob. Inst.*, no. May, 2011.

Press, G. "12 Big Data Definitions: What's Yours?," *Forbes*. 2014.

Priestley, T. 2015. The 3 Elements The Internet of Things Needs to Fulfil Real Value. Available at:http://www.forbes.com/sites/theopriestley/2015/07/16/the-3-elementsthe- internet-of-things-needs-to-fulfil-real-value/.

Rodseth, O., Perrera, L., Mo, B. (2016). "Big Data in Shipping – Challenges and Opportunities" Proceedings of the 15th International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT 2016), Italy.

Quintero, D. "IBM Software Defined Infrastructure for Big Data Analytics Workloads." 2015.

Sammut-Bonnici, T. and D. Galea (2014). SWOT analysis. In: *Wiley Encyclopedia of Management*, Sir C. L Cooper (Eds). John Wiley & Sons, Ltd.

Saxena, S. "Integrating Open and Big Data via e-Oman: prospects and issues," *Contemp. Arab Aff.*, vol. 9, no. 4, pp. 607–621, 2016.

Thompson, A. A., Strickland, A. J. & Gamble, J. E. (2007). Crafting and Executing Strategy-Concepts and Cases, (15th Edition), USA: McGraw-Hill/Irwin.

Widyaningrum, D.T. "Using Big Data in Learning Organizations," *Proc. 3Rd Int. Semin. Conf. Learn. Organ.*, vol. 45, no. Isclo, pp. 287–291, 2016.

Zainal, N.Z.B. Hussin, H. and Nazri M.N.M., "Big data initiatives by governments - Issues and challenges: A review," *Proc. - 6th Int. Conf. Inf. Commun. Technol. Muslim World, ICT4M 2016*, pp. 304– 309, 2017.

Zaman, I., Pazouki, K., Norman, R., Younessi, S., & Coleman, S. (2017). Challenges and Opportunities of Big Data Analytics for Upcoming Regulations and Future Transformation of the Shipping Industry. *Procedia Engineering, 194, 537–544.*