Data-Centric Engineering (2023), 4: e8
doi:10.1017/dce.2023.6

CAMBRIDGE UNIVERSITY PRESS

TRANSLATIONAL PAPER

Leveraging Big Data in port state control: An analysis of port state control data and its potential for governance and transparency in the shipping industry

Dimitrios Ampatzidis[1,2]

[1]Faculty of Technology, Linnaeus University, Kalmar, Sweden
[2]Finance and Statistics School, University of Piraeus, Piraeus, Greece
Email: d.ampatzidis@hotmail.gr

Received: 29 August 2022; **Revised:** 01 February 2023; **Accepted:** 24 March 2023

Keywords: Big Data; PSC inspections; Transparency; Governance

Abstract

The International Maritime Organization along with couple European countries (Paris MoU) has introduced in 1982 the port state control (PSC) inspections of vessels in national ports to evaluate their compliance with safety and security regulations. This study discusses how the PSC data share common characteristics with Big Data fundamental theories, and by interpreting them as Big Data, we could enjoy their governance and transparency as a Big Data challenge to gain value from their use. Thus, from the scope of Big Data, PSC should exhibit volume, velocity, variety, value, and complexity to support in the best possible way both officers ashore and on board to maintain the vessel in the best possible conditions for sailing. For the above purpose, this paper employs Big Data theories broadly used within the academic and business environment on datasets characteristics and how to access the value from Big Data and Analytics. The research concludes that PSC data provide valid information to the shipping industry. However, the lack of PSC data ability to present the complete picture of PSC regimes and ports challenges the maritime community's attempts for a safer and more sustainable industry.

Impact Statement

The maritime industry is currently grappling with digitalization, and reliable access to port state control (PSC) inspection data has become vital for the sector. PSC data can offer valuable insights into safety and security factors affecting ships. The present study aims to address the governance and transparency challenges related to PSC data utilization as a Big Data challenge. This study will explore how PSC data can be harnessed for research and business purposes, and it will also serve as a tool for screening datasets for Big Data compliance, not just in maritime but also in other fields. Ultimately, the research intends to investigate how professionals and researchers can use Big Data, the limitations posed by restricted access to databases, and how such restrictions can impact decision-making.

1. Introduction

The port state control (PSC) inspections are crucial for securing that the vessels globally are compliant with the guidelines provided by the International Maritime Organization (IMO) (n.d.); thus, it is one of the main ways of ensuring the safety and security of the vessels' operations. The findings (deficiencies and

detentions) of these inspections are reported by the local PSC Officers (PSCO) to the shipping industry through a structured procedure; the PSCO report to the local port authorities, then to the maritime administration, and finally to at least one of the ten Regimes[1] that each administration belongs to. Then it is the Regime's responsibility to inform the shipping community regarding the PSC outcome for each vessel and make such data publicly available to the industry.

Globally there are more than 110,000 vessels (Equasis, 2020) that can undergo more than one inspection per year, with different results and without a standard timeline. Therefore, we could approach the PSC data described by volume, velocity, and variety following Doug Laney's 3V theory for Big Data principles (Laney, 2001). Viewing the PSC data as Big Data helps us see the broader picture and realize their strategic value, symbolic, or functional (Grover et al., 2018). The only feasible way for Big Data to create any form of value for the users should be based only by actions dependent on Big Data techniques, such as data mining (Manyika et al., 2011).

Data could be seen as symbols of a specific piece of information (Beynon-Davies, 2020, p. 61); it is essential to understand that information and PSC data lack such sharing in the language and representation. For example, the deficiencies codes or wordings are not used universally, leading to wrong inputs and false alarms when using such data. It is observed that certain nations have implemented additional conventions and regulations beyond those widely adopted by other countries. As exemplified by the United States' Jones Act (Merchant Marine Act, 1920), Japan's regulations for the discharge of ballast water (Maritime Safety Agency, Japan, n.d.; Guidelines for the Control and management of Ships' Ballast Water to Minimize the Transfer of Harmful Aquatic Organisms and Pathogens, 2008), the European Union's Ship Recycling Regulation (Regulation [EU] No. 1257/2013 of the European Parliament and Council [2013] and of the Council of 20 November 2013 on ship recycling), and China's stringent sulfur regulations in marine fuels (Ministry of Transport, China, 2019). These variations in regulations may result in different deficiencies identified during PSC inspections. However, not all countries have the same level of transparency regarding PSC inspection reports. For instance, the United States only publicly releases reports of detainable inspections, which constitutes a relatively small percentage of overall cases. Therefore, there are limitations to the current PSC reporting system that affect the Big Data techniques and make it impossible for the shipping community to extract value from them. This paper will explore how the available PSC inspection data does not achieve its potential for enhancing safety and security.

2. Digitalization on Maritime

In the maritime sector, recently more and more ship-owning or ship-managing companies and organizations take initiatives to be in line with digital solutions for operational and management purposes. This fact can be interpreted as a side effect of the enhanced amount of disk storage; for example, the disk storage's total amount fourfold between 2005 and 2010 (Wilkes et al., 2008) and this is a nonstop enhancement. For the above reasons and because the industry was in a hurry to understand the fundamentals of operational and business matters better, it is possible to find many data-based applications and statistical models within maritime. There are models for detecting the unnormal behavior of vessels (Kowalska and Peel, 2012), piracy (Vanek et al., 2013), accidents (Zhang and Thai, 2016), or fuel consumption (Wang et al., 2018), or at least, applications to enhance the collaboration between people on board and ashore (Varelas and Plitsos, 2020).

Indeed, it is possible to find various uses of Big Data within the shipping industry, whether on an academic or business level. For example, Knapp and Franses (2007) found mathematical proof that the port of inspection affects the total probability of detention, and a few years later, Piniella et al. (2020) verified their results by finding that in Paris MoU, the inspection outcome is affected by the inspectors, the

[1] Paris MoU, Tokyo MoU, Acuerdo de Vina del Mar, Caribbean MoU, Abuja MoU, Black Sea MoU, Mediterranean MoU, Indian Ocean MoU, Riyadh MoU, and United States Coast Guard (IMO, n.d.).

https://doi.org/10.1017/dce.2023.6 Published online by Cambridge University Press

country, and the port. It is understood that it is possible to have Big Data support and effectively assist shipping professionals ashore on their decisions. It is widely accepted that the goal of each action taken in every organization is to minimize the risks for every decision that stakeholders should make; Big Data and Analytics (BDA) could provide essential support to Business professionals on that issue.

One of the key elements to performing statistical analysis or exploring BDA is access to data; this is a significant issue in the maritime industry as companies and organizations store vast amounts of data but do not share it. This issue not only provokes a discussion on transparency and data governance but raises concerns on how the shipping community will be actively involved in the digital world and enhance its safety and security, as well as ensure sustainable operation. It is a fact that having access to data without an ability to gain any insight by it does not present any value to anyone (Sharda et al., 2014), and coming to PSC data, shipping companies should compare their results to others' (Yuan et al., 2020) to understand the real circumstances and improve themselves as IMO and the whole maritime community expects from them (Mitroussi, 2004). As a result, examining the example of PSC data and how companies and organizations have access to them, use them for analytics, could be the necessary boost for a more open data transfer within the shipping community.

2.1. PSC data as Big Data

For the purpose of this discussion, it is more useful to approach the PSC lack of publicity and easy-to-access data as a Big Data issue; by doing so it is possible to address available approaches using the fundamental knowledge of the technology aspect. Thus, based on the above reasoning, it is evident that first we must establish why PSC could be treated as Big Data. There are plenty of theories concerning the characteristics we should expect of information or data to understand them as Big Data; however, one of the first and one of the elements most often found within other approaches, is the 3V Laney's theory (Laney, 2001). Laney insisted that a dataset should have volume, velocity, and veracity to be realized as Big Data set. While Oracle (2019) added the Value that Big Data should bring to the end-user, the shipping professionals in another research found to look for value in the analytics that could be exported from the complexity of the PSC inspections data (Ampatzidis, 2021). In conclusion, the present paper understands PSC data as Big Data under the principles of volume, velocity, variety, value, and complexity (4V&C).

More than 100 vessels are actively sailing across the seas, and daily there are PSC inspections in various ports, with a structured procedure leading to a standardized report. Indeed, if all the information gathered by the inspection is collected—concerning vessel, companies or business entities connected with the ship, inspection authorities, and outcomes—there is a vast amount of information; therefore, PSC data have Volume. Furthermore, as the information mentioned above could be dynamically changed (e.g., the companies connected with the vessel) and new data may occur without a typical pattern, we could assume that PSC data also present Velocity. Within the 3V approach, it could be understood that the PSC data as datasets with structural heterogeneity, it is possible to include information in various formats like text, numbers, and images; therefore, we conclude that PSC data have Variety. According to the criteria of Laney, PSC inspection data could be seen as Big Data as they present characteristics within the 3V's theoretical framework.

The last step is to establish their value as it is a crucial element for Big Data utility to clarify that PSC data possess characteristics that they belong to the Big Data family. McKinsey Global Institute (Manyika et al., 2011) proposed a way to examine the value of Big Data to organizations and institutions based on five elements: transparency, experimentation, customizing actions, decision-making, and exploring new business models and rationally coming by having all the information for the PSC outcome for every vessel and company worldwide provides transparency to the industry and among the shipping professionals on many levels. It is essential to find what is challenging for vessels and where the maritime industry should place its focus for the future. So far, management companies, or even broader organizations such as classification societies, P&I clubs, and other business entities, have faced various issues in collecting such data even for the fleet that they relate to; this problem seriously affects their research on how vessels could be more efficient on resolving any issues that arise regarding safety and security following the applied

https://doi.org/10.1017/dce.2023.6 Published online by Cambridge University Press

regulations. Thus, access to PSC data would make it possible for any organization or shipping professional to deeply investigate issues connected with the inspections of vessels, to have a more data-driven approach and finally, to customize their actions accordingly. Exploring new business models is essential to fully understand the current trends within business operations. Shipping professionals cannot perform accordingly without having a holistic view of the safety and security of the existing fleet. The last element is Complexity, as the shipping professionals on shore stated they require the complete picture of each PSC inspection outcome; with such information, it will be possible to apply mathematical algorithms and analytical methods to determine the vessels' characteristics related to the inspections' findings.

In conclusion, we could state that indeed the lack of PSC inspections data can be treated as a Big Data challenge to handle, as PSC exhibit the fundamental characteristics of Big Data, and their value could be an outcome of their analysis, followed by Big Data techniques (like data mining).

2.2. Challenges to address

Defining the Big Data nature of PSC inspection information makes it easier to address and handle the issues arising from the existing published data as Big Data challenges. For those involved in PSC inspections, it is known that the MoUs use an extensive checklist of deficiency items divided into broader categories and groups. After each inspection, the shipping company is informed about the outcome from the PSCO with a typical procedure of a PSC report. In this report, the shipping professionals can access the findings from PSCO and a brief description of what was wrong, and with the help of this report, the DPOs (designated person ashore) and HSQE (health, safety, quality, and environment) Officers try to resolve any issues to prepare the vessel for the continuation of its voyage. In the meantime, the officers ashore try to make sure themselves that the vessel is efficient and operatable based on the regulations; for that reason, they use any possible source of information, as the PSC reports. However, since the MoUs and the PSCO worldwide do not follow the same coding of the deficiencies, or they do not focus on all the aspects similarly, it is common for the shipping professionals to face completely different issues between ports and regimes; an issue thought to be resolved in one port, may be reported as a detainable item in the next one. This results in problematic preparation of PSC inspections from the viewpoint of ship managers as they do not have the complete picture from all the ports and regimes for every vessel, as such data are not widely published.

The ten PSC Regimes operate under the IMO's regulations and are responsible for addressing any issue concerning the safety, security, health, and quality of the vessels and crew on board. The IMO does impose specific regulations for the vessels and how they sail concerning nature and people. However, the PSC Regimes are responsible for screening the vessels for compliance with the legislation, without however following the same procedures or protocol. This happens mainly because the authority for the inspections lies on national bodies (PSCO), and each country has a different approach to the same topic; thus, the IMO has shown no responsibility or intention so far to standardize the procedure for all countries to have the same reports published. For the above reasons, it is easy to understand that the ten PSC Regimes present considerable differences in their published statistics (when they publish such reports), or even that the same vessels inspected within a few weeks in ports between MoUs, present entirely different results, as if they were different ships.

As a result, the industry lacks transparency of the PSC inspections data. It may be possible to find basic information for vessels after specific research but extracting information regarding Ports or Countries is not feasible. If the IMO published a standardize procedure on the PSC inspections and on how the outcomes should be reported to the PSC Regimes and the IMO, it would be possible to combine information on how the PSC inspections assist shipping worldwide in having better-maintained vessels in the seas. If such a practice prevails, it is rational to expect that even the IMO will have the authority and the ability to access what is happening across different ports, find correlations between vessels' characteristics and deficiencies reported or correlations between PSCO's background or skills and deficiencies and define better guidance to the operators and more data-driven training to the PSCO's. Complete transparency on PSC inspections could lead to a better understanding of the fundamental

https://doi.org/10.1017/dce.2023.6 Published online by Cambridge University Press

reasons that lead to a poorly maintained ship, a detailed focused inspection could enlighten the whole maritime community and upscale the operation following that knowledge. The limited availability of information regarding the quality of vessels approaching the United States is demonstrated by the United States Coast Guard's publication of only detainable PSC inspections. This lack of information deprives the maritime industry's interested parties and stakeholders of valuable data to assess the potential maintenance challenges posed by these vessels. On the other hand, vessels that belong to either the Paris memorandum of understanding (MoU) or the Tokyo MoU have access to more comprehensive information and are thus better equipped to understand the risks associated with their fleets. The publication of all PSC inspection results as soon as they are conducted would provide the maritime industry with a clearer understanding of the condition of vessels sailing in the vicinity.

3. Conclusion

In recent years, the vast technological expansion has made it possible to set new standards on how we communicate information and explore new opportunities on how we work and operate daily tasks. The scope of this paper was to theoretically approach in this light the needs of the shipping professionals and, in general, the maritime community and indicate how more transparency on PSC inspections data and a more standardized procedure among PSC regimes could lead to a better outcome for more safe, secure, and sustainable sailing across seas. In the last few years, it has become feasible to find many Big Data solutions within the shipping industry providing exemplary services in various fields like fleet operation, oil handling, emissions estimations, and analytics. Researchers and companies have used the available data and information to make it possible to have better vessels sailing, profoundly understanding the concerns of shipping professionals or even seafarers in the previous century. The way these services affect the shipping industry and the marine research, highlights the fact that it is essential to gain more details from the PSC inspections to be better informed on how the authorized PSCO judge the vessels regarding safety, security, health, and environment. To better approach this challenge, it is useful to understand the fundamental elements of PSC data and treat them as Big Data. Applying the knowledge that stems from Laney's 3V theory (Laney, 2001), the McKinsey Global Institute (Manyika et al., 2011), and qualitative research on PSC data characteristics (Ampatzidis, 2021), this approach paves the way to tackle issues concerning the transparency of PSC in a more efficient way. In conclusion, it is obvious that a more standardized procedure and report on PSC can lead to a better understanding of how we can make better vessels, maintain them more successfully, and finally, have a more sustainable industry serving humanity.

The IMO should take proactive measures to establish a standardized procedure for the publication of all available information pertaining to PSC inspections by all MoUs. Such a measure would ensure that the maritime industry has access to comprehensive and up-to-date information on the condition of vessels and the results of PSC inspections on a daily basis.

Abbreviations

3V	volume, velocity, and variety
4V&C	volume, velocity, variety, value, and complexity
BDA	Big Data and Analytics
DPA	designated person ashore
HSQE	health, safety, quality, and environment
IMO	International Maritime Organization
MoU	memorandum of understanding
PSC	port state control
PSCO	port state control officer

https://doi.org/10.1017/dce.2023.6 Published online by Cambridge University Press

Author contribution. Writing—original draft: D.A.; Writing—review and editing: D.A.

Competing interest. The author declares no competing interests exist.

Data availability statement. Data availability does not apply to this article as no new data were created or analyzed in this study.

Funding statement. This research was supported by a grant from the Lloyd's Register Foundation.

References

Ampatzidis D (2021) Big Data Usage in the Maritime Industry: A Qualitative Study for the Use of Port State Control (PSC) Inspection Data by Shipping Professionals (Dissertation). Available at http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-107630 (accessed 18 June 2022).

Beynon-Davies P (2020) *Business Information Systems*, 3rd Edn. London: Red Globe Press.

Equasis (2020) The 2020 World Merchant Fleet Statistics from Equasis. Available at https://www.equasis.org/Fichiers/Statistique/MOA/Documents%20availables%20on%20statistics%20of%20Equasis/Equasis%20Statistics%20-%20The%20world%20fleet%202020.pdf (accessed 8 June 2022).

European Parliament and Council (2013) Regulation (EU) No. 1257/2013 of the European Parliament and of the Council of 20 November 2013 on ship recycling. Official Journal of the European Union, L 334, pp. 74–89.

Grover V, **Chiang RHL**, **Liang T-P and Zhang D** (2018) Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems 35*(2), 388–423.

Guidelines for the Control and Management of Ships' Ballast Water to Minimize the Transfer of Harmful Aquatic Organisms and Pathogens (2008) Tokyo, Japan: Maritime Safety Agency, Japan.

International Maritime Organization (IMO) (n.d.) Port State Control. Available at https://www.imo.org/en/OurWork/IIIS/Pages/Port%20State%20Control.aspx/ (accessed 6 June 2022).

Knapp S and Franses PH (2007) A global view on port state control: Econometric analysis of the differences across port state control regimes. *Maritime Policy and Management 34*(5), 453–482.

Kowalska K and Peel L (2012) Maritime anomaly detection using Gaussian process active learning. In *ISIF (International Society on Information Fusion), 15th International Conference on Information Fusion, Singapore, 3–12 July 2012*. Singapore: IEEE.

Laney D (2001) 3D Data Management: Controlling Data Volume, Velocity, and Variety. Technical Report, META Group.

Manyika J, **Chui M**, **Brown B**, **Bughin J**, **Dobbs R**, **Roxburgh C and Byers AH** (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* Washington, DC: McKinsey Global Institute.

Maritime Safety Agency, Japan (n.d.) Act on the Prevention of Marine Pollution by Ships [Japanese Legislation].

Merchant Marine Act (1920) 46 U.S.C. §§ 55101–55110. 66th United States Congress, United States of America.

Ministry of Transport, China (2019) *Notice on the Implementation of the Control on the Sulfur Content of Ship Fuel Oil in Main Sea Areas of the People's Republic of China*. Beijing, China: Ministry of Transport.

Mitroussi K (2004) Quality in shipping: IMO's role and problems of implementation. *Disaster Prevention and Management 13*(1), 50–58.

Oracle Corporation, 2019. What is Big Data? [pdf] Oracle Corporation. Available at: <https://www.oracle.com/a/ocom/docs/what-is-bigdata-ebook-4421383.pdf> [Accessed: 07 March 2021].

Piniella F, **Alcaide JI and Rodriguez-Diaz E** (2020) Identifying stakeholder perceptions and realities of Paris MoU inspections. *WMU Journal of Maritime Affairs 19*(1), 27–49.

Sharda R, **Delen D and Turban E** (2014) *Business Intelligence and Analytics: Systems for Decision Support*, 10th Edn. London: Pearson Education Limited.

Vanek O, **Jacob M**, **Hrstka O and Pechoucek M** (2013) Agent-based model of maritime traffic in piracy-affected waters. *Transportation Research Part C: EmergingTechnologies 36*, 157–176.

Varelas T and Plitsos S (2020) Real-time ship management through the lend of big data. In *Big Data Service, Sixth International Conference on Big Data Computing Service and Applications. Oxford, United Kingdom, 3–6 August 2020*. Oxford, United Kingdom: IEEE.

Wang S, **Ji B**, **Zhao J**, **Liu W and Xu T** (2018) Predicting ship fuel consumption based on LASSO regression. *Transportation Research Part D: Transport and Environment 65*, 817–824.

Wilkes J, **Hoover C**, **Keer B**, **Mehra P and Veitch A** (2008) *Storage, Data, and Information Systems*, 5th Edn. Palo Alto, CA: Hewlett-Packard Laboratories.

Yuan CC, **Chung WH**, **Cai C and Sung ST** (2020) Application of statistical process control on port state control. *Journal of Marine Science and Engineering 8*(10), 746.

Zhang G and Thai VV (2016) Expert elicitation and Bayesian network modeling for shipping accidents: A literature review. *Safety Science 87*, 53–62.

Dimitrios Ampatzidis, a native of Athens, Greece, has a lifelong passion for statistics and information systems. He earned a B.Sc. degree in Statistics and Insurance Science from the University of Piraeus, Greece in 2017, followed by a M.Sc. degree in Information Systems from Linnaeus University, Sweden in 2021. In August 2022, he began a M.Sc. degree in IT Project Management at Stockholm University, Sweden. With a strong understanding of the maritime industry, acquired through academic and professional experience since 2019, Dimitrios currently works as a Compliance and Data Analyst at Lloyd's Register, a position he has held since 2021. In this role, he applies his expertise in statistics and information systems to ensure compliance with safety and security regulations, making valuable contributions to the industry.

Cite this article: Ampatzidis D (2023). Leveraging Big Data in port state control: An analysis of port state control data and its potential for governance and transparency in the shipping industry. *Data-Centric Engineering*, 4, e8. doi:10.1017/dce.2023.6