10-31-2022

Impact of big data analytics on the privacy rights of seafarers

Adanna Nkiruka Okonkwo

Recommended Citation

Okonkwo, Adanna Nkiruka, "Impact of big data analytics on the privacy rights of seafarers" (2022). *World Maritime University Dissertations*. 2084.
https://commons.wmu.se/all_dissertations/2084

WORLD MARITIME UNIVERSITY
Malmö, Sweden

THE IMPACT OF BIG DATA ANALYTICS ON THE PRIVACY RIGHTS OF SEAFARERS

By

ADANNA NKIRUKA OKONKWO
Nigeria

A dissertation submitted to the World Maritime University in partial fulfilment of the requirements for the award of the degree of

MASTER OF SCIENCE
in
MARITIME AFFAIRS

(MARITIME ENERGY MANAGEMENT)

2022

Declaration

I certify that all the material in this dissertation that is not my own work has been identified, and that no material is included for which a degree has previously been conferred on me.

The contents of this dissertation reflect my own personal views, and are not necessarily endorsed by the University.

(Signature): **...**

(Date): **............**20/9/2022**......................**

Supervised by: **Professor Dr. Aykut I. Ölçer**

Supervisor's affiliation: **Director of Research Nippon Foundation Professorial Chair in Marine Technology and Innovation Head, Maritime Energy Management**

Acknowledgements

First and foremost, I would like to thank the Almighty God for His infinite mercies, unconditional and endless love, guidance, protection and grace throughout the course of this programme.

My sincere appreciation to my sponsors Orient Fond for the opportunity given to me to study at this wonderful institution. I am grateful for your scholarship support.

I am highly indebted to my dissertation supervisor and mentor, Professor Aykut Olcer for his professional perspicacity, his constructive criticism, mentorship, suggestions and guidance towards the accomplishment of this research work. Many thanks to the entire staff of Maritime Energy Management Specialization, Professor Fabio Ballini, Professor Alessandro Schönborn and Professor Momoko kitada for their excellent lectures and incredible support.

I am immensely grateful to my parents Mr. Godwin Okonkwo and Lady. Victoria Okonkwo for their continued prayers, support, love, care and guidance throughout my life and career.

I am very grateful to my darling husband, you are my support system and this wouldn't have been possible without you. To my siblings Uchenna, Edozie, Nnamdi, and Faith, My aunties Ifeoma and Veronica, and my uncle Stanley. I sincerely appreciate you and your efforts.

To my lovely course mates and friends, I appreciate you all for your contributions and support toward the success of this journey. God bless and reward you all.

Abstract

Title of Dissertation: **IMPACT OF BIG DATA ANALYTICS ON THE PRIVACY RIGHTS OF SEAFARERS**

Degree: **Masters of Science**

This research evaluated the impact of big data analytics on the privacy rights of seafarers and how security and privacy infringement issues can be linked to the collection, storing, analysis, processing, destruction, reuse and sharing of data. The research further analysed the relation that exists between big data characteristics and privacy infringement from the standpoints of data collection, storage, processing, sharing and accessibility.

The research employed an exploratory mixed method design. The qualitative exploratory part consisted of semi-structured interviews and the choice of participants for these interviews consisted of a ship-owner, a data controller and two seafarers working on board ships engaged in international and domestic voyages. Responses from the interviews assisted in the design of the questionnaire and the formulation of the research hypothesis. In the quantitative data analysis, questionnaires were distributed which elicited responses from 125 seafarers working on board vessels of shipping companies in 4 different continents (Africa, Europe, Asia, North America, and the United Kingdom) and this questionnaire was reconstructed and its validity and reliability was tested using Exploratory Factor Analysis of the SPSS tool (SPSS 26) including other statistical tests and analyses such as a correlation analysis. This analysis resulted in the extraction of three (3) factors: top management commitment to data safety, consent for the use of personal data and duration of personal data.

The study found that top management commitment is a factor that enables the infringement of seafarer's privacy rights. The study further found that Big data analytics in shipping companies negatively impacts on the privacy rights of seafarers whose consent is not sought and obtained before the use of their personal data. This study recommended that shipping companies adopt a Privacy by Design model as it is crucial for setting boundaries for big data processing and also it incorporates the proper data safety measures at the very centre of the analytics value chain in order to enable all the benefits of analytics without breaching seafarer’s privacy.

KEYWORDS: Data, Big Data, Analytics, Privacy, Infringement, Consent.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABE	=	Attributes-Based Encryption
AIS	=	Automatic Identification System
APEC	=	Asian-Pacific Economic Cooperation
BI	=	Business Intelligence
CUP	=	Consent for the Use of Personal data
CSP	=	Cloud Service Providers
DP	=	Data Privacy
DPC	=	Duration of Personal Data Collection
EFA	=	Exploratory Factor Analysis
EMSA	=	European Maritime Safety Agency
ENISA	=	European Union Agency for Network an Information Security
FTC	=	Federal Trade Commission
GDPR	=	General Data Protection Regulation
GSM	=	Global System for Mobile Communication
GPS	=	Global Position System
IMO	=	International Maritime Organization
IoT	=	Internet of Things
KMO	=	Kaiser- Mayer Olkin
LRIT	=	Long Range Identification and Tracking
MASS	=	Maritime Autonomous Surface Ship
OECD	=	Organization Economic Cooperation and Development
ODP	=	Overall Data Privacy
PETs	=	Privacy Enhancing Technologies
PWC	=	Price Waterhouse Coopers
REC	=	Research Ethic Committee
SOLAS	=	The International Convention for Safety of Life at Sea

TMC	=	Top Management Commitment
VDR	=	Voyage Data Recorder
VHF	=	Very high frequency
VSAT	=	Very small aperture terminal
VTS	=	Vessel Traffic Services

CHAPTER ONE: INRODUCTION

1.1. Background of Study

In a fast-paced world such as ours, the use of data-driven technologies is steadily on the increase, the evolution of data-driven technologies and its impact in modern life is unquantifiable (Cukier and Mayer choenberger, 2013; Neff and Nafus, 2016). Datafication has continually advanced into all spheres of human life, from education to healthcare and more recently, the shipping industry. These advancements in data-driven technologies and the associated reductions in cost have led to the availability of data which can be both in a structured and an unstructured form (Bhatia & Mittal, 2019).

The current trends and technological advancement is shaping a future with ubiquitous connectivity. The penetration of smartphones has been estimated to be approximately 80% in almost all developed countries (Atas and Çelik, 2019), and it is further estimated that approximately 3.2 billion people use mobile internet globally (GSMA, 2017). Finally, it is estimated that the number of Internet of Things (IoT) connections will increase in 3 folds between 2017 and 2025, thereby reaching 25 billion, driven by a proliferation of use cases for smart homes and smart cities (Khan et al, 2020).

Despite the popularity of big data in modern life and its widespread use, the meaning of this concept is still shrouded by conceptual vagueness as there is currently no universally accepted definition of big data (Favaretto et al., 2020). Einav and Levin (2014) defined big data as involving the availability of data in real time, at larger scale, with less structure, and on different types of variables than previously used. Big Data can also be commonly used to describe a plethora of different concepts; from the collection of data to processing of huge amounts of data to other advanced digital techniques which are designed to show patterns related to human behaviour (Favaretto et al., 2020). Big data is also defined as the amount of data which cannot be managed by conventional methods (Jan et al, 2019). According to De Mauro et al (2015), big data involves a range of different concepts; the technological ability to collect, store, aggregate data. The lack of a universally accepted definition has led to existing ambiguity and research into this subject has evolved into multiple, different and inconsistent paths.

The transport industry is one of the industries where quantities of data are generated on a daily basis through transactions, sensors and location services (Zhu et al, 2018). The industry provides an enormous opportunity for the collection of data and the subsequent conversion of these data into valuable insights. The location services used by the transport industry are built on location-sensing data which are extracted from location-relevant signals like Wi-Fi positioning (Shi et al, 2020). This has been adopted by the transport industry for navigational purposes and also the identification of nearby transport options and services. Location services are becoming extremely accurate and this accuracy is supplemented by the ever growing location sensing technologies and the satellite positioning systems (Grejner-Brzezinska et al, 2016). Sensors are more and more embedded in devices and ships, the use of face scanners, radars and cameras with adapted algorithms which are able to detect users and make sense of their environment are increasingly in use (Berman & Stern, 2011).

In the shipping industry, data capturing as well as its analysis are part of the industry (Osekowska et al., 2017). This is evident in the captain's log and engine log book where the captured data in these logs are carefully analysed in the investigation of accidents. More recently, the shipbuilding industry has become more technology-reliant which has led to the adoption of automatic control systems on vessels wherein data is collected, analysed and processed in order to create more efficient ships as well as achieve an optimized workflow and environment-friendly engines (Latifov, 2019).

Similarly, with recent technological advancement, high bunker prices and the need for cost cutting measures to reduce the high operational cost associated with shipping came the new hope for the giant companies – autonomous ships (Burmeister et al., 2014). Remote-controlled and autonomous vessels are currently the biggest development in the shipping industry, this new technology-driven way of operating ships makes use of data analytics. The introduction of MASS to the shipping industry will undoubtedly increase the level of data transmission as the collection, analysis and management of huge volume of unstructured data for example data on voyage performance, machinery, ship structure, cargo, traffic, fuel consumption as well as the weather condition will most certainly be used in providing valuable insights on ship operations, uncover hidden navigation pattern and market trends (Du et al, 2019).

The analysis of this big data will enable ship-owners to predict the likely outcomes of certain voyages. Furthermore, this may result in cost reduction as the shipping industry and ship-owners will most likely be able to identify better and more efficient ways of doing business which will help in making real time decisions on ship safety and efficiency.

1.2. Problem Statement

As the use of data-driven technologies in the shipping industry continues to grow, it is generating a vast amount of data which is also growing at an exponential rate therefore, creating huge opportunities as well as raising new and different challenges. There are considerable uncertainties and lots of unanswered questions, for example, what will be the effect of big data on the privacy rights of seafarers? Will the collection of data infringe on the privacy protection principle? Will the use of big data provoke a backlash from seafarers which will ultimately harm a range of safety use cases? Are there policies in place to protect seafarers from data exploitation? Are the policies adequate to protect seafarer's privacy along the entire data continuum - from data sensing to its extraction, storage, aggregation, transcription, retention, analysis and destruction? Are there insurance policies in place for seafarers to provide a nudge to ship-owners to adopt safer behaviours? It is without doubt that the very nature of data collection is changing, the access to as well as the use of this data can result in dangerous outcomes especially as it relates to the unwanted and unintended erosion of privacy rights.

The shipping industry generates quantities of data on a daily basis through transactions, sensors and location services. The location services are used by the shipping companies for navigational purposes and also the identification of nearby transport options and services. However, location data has the capability of revealing a seafarer’s religion, health status and political affiliations through keeping a log of a seafarer’s repeated visits to a worship place, hospital or political gathering. Therefore, the use of big data in the shipping industry has the capacity of revealing seafarer’s sensitive information, the question to be asked at this juncture is, are there adequate policies in place to protect seafarer’s privacy rights?

Big data has become a key element of a firm's competitiveness as the overall annual economic gains from big data is estimated to be approximately US$610 billion in productivity and cost savings (Manyika et al, 2011)). However, these gains have come at a cost as big data characteristics have been closely linked to certain privacy and security effects on its users as an unethical collection of huge amount of data can lead to privacy violations and security breaches which will likely lead to an amplified technical impact which can be legal liability, reputational damage and other ethical harms (ISACA, 2014).

The privacy and security risks associated with high volume data collected from different sources as well as complex data sharing and accessibility are some of the issues that arise in a big data environment (Kshetri, 2014). Undoubtedly, the existing non-big data solution currently in use in the shipping industry is not designed in such a way that it can handle the scale, variety, speed and complexity of big data in autonomous vessels. Most shipping companies also lack systematic approaches which are important in ensuring appropriate data access mechanisms. Similarly, the time-variant nature of data flow in the shipping industry implies that some of the aforementioned issues will be significant during the peak data traffic, and where the shipping companies lack the capability to securely store and process this collected data during this peak data traffic, the peak data flow will require the shipping company to outsource to cloud service providers (CSPs) thereby compromising the data privacy of the seafarers. According to Trustwave, approximately 64% of data privacy violations and security breaches in 2012 involved the use of outsourcing providers for data collection and processing (Kshetri, 2014). This was because many firms were not capable of building a complete in-house big data environment

(Andrejevic & Gates, 2014), which will be the case with most shipping companies when autonomous vessels will be in full operation thereby placing much reliance on CSPs for storage, processing and analysis needs.

Another issue related to outsourcing is that most CSPs are bigger than the shipping companies and also deal with higher volumes of data which are stored in the cloud, information stored in the cloud are often times potential targets for cybercriminals and secondly, these CSPs might use seafarer's data for their personal benefits thereby compromising seafarer's privacy. For example, circa 2013, the Swedish data protection authority, Datainspektionen directed an organization named Salem Municipality to stop the use of Google Apps, calendar services and e-mail. The reason given for this directive was that Google writes the contract of engagement and also sets the rules for handling information which gives them plenty of room to use data for purposes which are inconsistent with the purpose specified by the municipality (Chander & Le, 2014). Datainspektionen was concerned that the agreement provided Google with too much power to process personal data for its own potential benefit.

Furthermore, the introduction of MASS to the shipping industry will undoubtedly increase the level of data transmission as well as require shipping companies to store all the extracted data in a cloud or place to facilitate analysis, the bigger volume and the higher concentration of this data will most definitely create an opportunity and an appealing target for cybercriminals and hackers. Similarly, the higher volume and concentration of data will likely increase the possibility that data files containing personal, valuable and sensitive information of seafarers will be stored and if used inappropriately will lead to psychological, economic, emotional and social harm to seafarers. For example, big data predictive analysis will most likely result in an accurate prediction of a seafarer's preferences thereby leading to unpleasant, creepy and frightening experiences for seafarers. The above phenomenon can also be known as predictive privacy (Crawford & Schultz, 2013).

1.3. Aims and Objectives

The aim of this research is to critically evaluate the impact of big data analytics on the privacy rights of seafarers as well as the relation that exists between big data characteristics and privacy infringement from the standpoints of data collection, storage, processing, sharing and accessibility.

The above aim will be researched using the case study of ships engaged in both international and domestic voyages. These characteristics will allow the researcher to investigate whether security and privacy infringement issues can be linked to the collection, storing, analysis, processing, destruction, reuse and sharing of data.

In order to achieve the research aim, the study will progress according to the following objectives:

a. Assessing top management commitment to seafarer's data safety as a factor that enable data privacy infringement.

b. Ascertaining whether the consent of seafarers is sought and obtained before their data are used by their companies.

c. Highlighting the costs, benefits as well as the externalities which are associated with the use of big data by shipping companies.

d. Investigating how the inherent characteristics of big data are linked to security and privacy infringements.

e. Recommending the adoption of Privacy by Design Strategy as a solution for Data Privacy issues in shipping companies.

The research objectives will be investigated through the following questions;

a. Does the top management of shipping companies prioritize profit making over data safety?

b. Is the consent of seafarers sought and obtained before their data are used by their companies?

1.4. Research scope

This research evaluated the impact of big data analytics on the privacy rights of seafarers and how security and privacy infringement issues can be linked to the collection, storing, analysis, processing, destruction, reuse and sharing of data. The research further analysed the relation that exists between big data characteristics and privacy infringement from the standpoints of data collection, storage, processing, sharing and accessibility. Although this study evaluated the impact of big data analytics on the privacy rights of seafarers, the study did not discuss the enforcement of these privacy rights as well as the integration of big data analytics in the public policy pedagogy as this falls outside the scope of the research.

The study discussed the failure of shipping companies to create insurance policies to cover for data infringements, however, the research stopped short of discussing the role of big data in maritime insurance as well as the societal and economic advantages of using big data analytics in maritime insurance. The study also did not discuss the regulatory debate on maritime insurance trade-offs and the major issues that have come up in the public and regulatory discourse as all these are beyond the scope of the research.

1.5. Dissertation Structure

The dissertation is structured into five chapters, as shown in Figure 2:

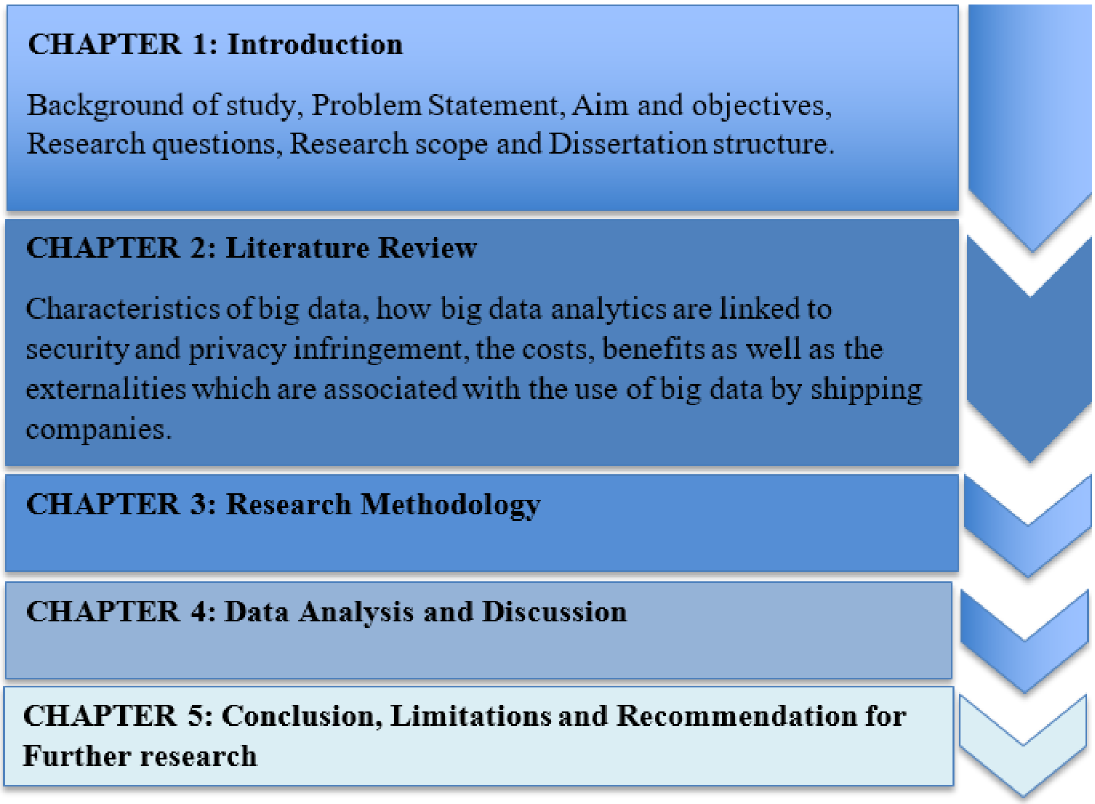
- Chapter one contains the background of study, the problem statement, the aim and objectives, research scope, and the dissertation structure.

- Chapter two focuses on the literature review, mainly on the inherent characteristics of big data, how big data analytics are linked to security and privacy infringements and the costs, benefits as well as the externalities which are associated with the use of big data by shipping companies.

- Chapter three explains the methodology employed in this research and the processes for the personal interviews and survey questionnaire.

- Chapter four contains the analysis of the interviews and questionnaires as well as recommendation of Privacy by design strategy as a solution to privacy rights infringement in shipping companies.

- Chapter five provides the conclusions and limitation of the research including recommendations for further study.



CHAPTER 1: Introduction

Background of study, Problem Statement, Aim and objectives, Research questions, Research scope and Dissertation structure.

CHAPTER 2: Literature Review

Characteristics of big data, how big data analytics are linked to security and privacy infringement, the costs, benefits as well as the externalities which are associated with the use of big data by shipping companies.

CHAPTER 3: Research Methodology

CHAPTER 4: Data Analysis and Discussion

CHAPTER 5: Conclusion, Limitations and Recommendation for Further research

Figure 1: Structure of dissertation (Source: Researcher)

CHAPTER TWO: LITERATURE REVIEW

2.1. General Background of the Concept "Big Data"

The waning level of physical efficiency gains and the quest for a modest economic shipping climate which currently defines the shipping industry explains the current optimization technologies of "soft" assets such as information and data and its increased use in the shipping industry.
The reduced cost of data storage has led to the increased retention of data that would have previously been discarded. This point was further buttressed by science historian George Dyson who asserted that "*Big Data is what happened when the cost of storing information became less than the cost of making the decision to throw it away*." (Dyson, 2013).

The term "Big Data" was first coined by Michael Cox and David Ellsworth (NASA Researchers) in their paper published in 1970 titled "Application-controlled paging for out-of-core visualization" (Sekkesaeter, 2017). The term was used in the paper to describe the problems associated with having too-large datasets over and beyond the storage memory capacity of the main computer.

While there is no broadly agreed definition that exists for big data however, Big Data can be broadly defined as exceptionally large datasets that cannot be acquired, stored and interpreted by traditional data processing software or a typical personal computer or the analytical capacity of commonly used spreadsheet application but through modern technology (Silva et al., 2019).

A well-known definition of Big Data was presented by Doug Laney (2001) who explained the concept of Big Data using the 3Vs model, the 3Vs represents Volume (the amount of data), Velocity (the speed at which data is collected and processed) and Variety (the range of structured and unstructured elements that comprise the data sets) (Laney, 2001). More recently two more Vs have emerged as key characteristics of Big Data and they are Veracity (uncertainty of data) and Value (huge value with a very low density) (Hariri et al., 2019). These characteristics of Big Data are typically used in differentiating Big Data from other data. It is also important to stress that Big Data is not a singular construct; rather, it is a process spanning data acquisition, processing and interpretation (Laney, 2001). Therefore, the lifecycle of Big Data will be described in the following sub-headings. (See Figure 2)

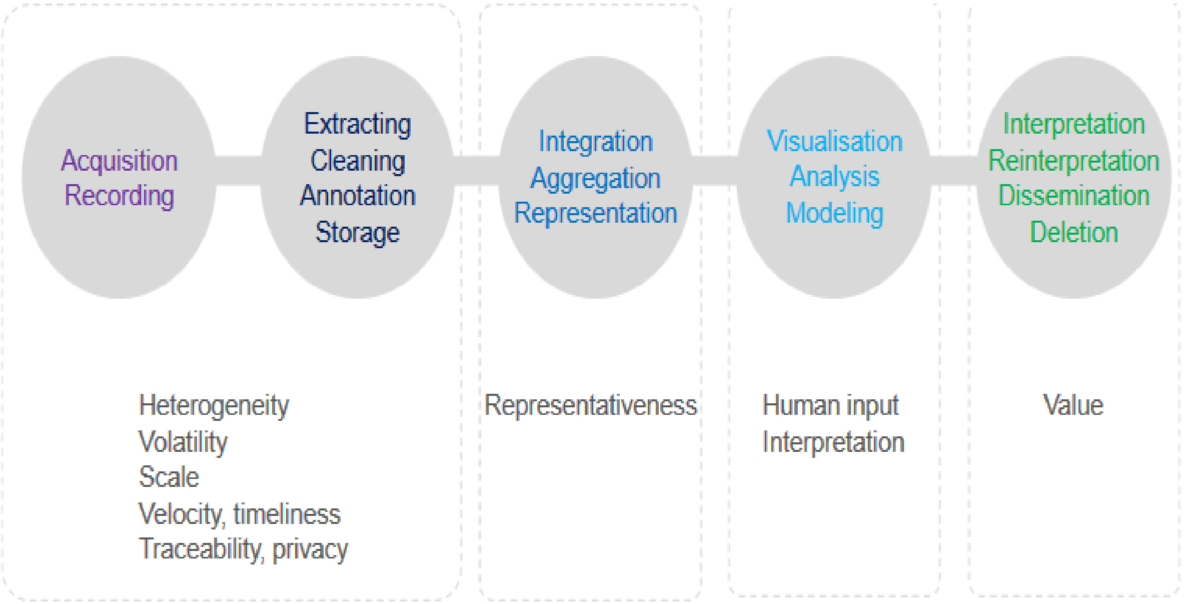


Figure 2: Big Data Collection and Analysis Lifecycle (Crist et al., 2015)
The cycle is iterative to represent the big data lifecycle. The lifecycle represents the step-by-step stages involved in performing analysis on Big Data.

2.2. Big Data Analytics

Big Data Analytics involves the extraction of new and valuable insight from the stored large collection of data (Silva et al., 2019). The ultimate aim of this analytics is to provide data for decision makers using mathematical and statistical methods to enable them understand data, simulate scenarios, validate hypotheses and make predictive forecasts for future incidents (Sarker, 2021).

According to Bendre and Thool (2016) and Burmester et al. (2018), big data analytics circles around five stages namely:

- **Data generation or integration**: This is the first stage where large amounts of data are gathered from several datasets like search engine pages, publishing factual data, world events, social media graphs, analysis of natural language content etc., for future analytics (Sikos, 2015).
- **Data acquisition or management:** This stage of data analytics involves the process of gathering, filtering and cleaning large amounts of data (Lyko et al., 2016).
- **Data storage**: This is the stage where a platform with a clustered network of servers and community hardware are used to store the data (Bendre and Thool, 2016).
- **Data analytics**: this is the stage where useful information are examined from the huge data storage using complicated machine learning and data mining techniques (Chen et al., 2014).
- **Data visualization or presentation**: The graphical format representation of data can easily be understood and represented in a simple way (Bendre and Thool, 2016)

A variety of techniques have been developed over the years to manipulate, aggregate and visualize big data and these techniques for data analytics can be grouped into, although not limited to, the following categories; Data mining, data fusion, and optimization and visualization techniques (McKinsey Global Institute, 2011)

- Data mining is one of the key concept in Big Data Analytics. This technique involves the application of data science techniques in the analysis and exploration of large datasets in order to find useful patterns in those data (Sowmya & Suneetha, 2017), for example extracting data from a large dataset in order to find the relationships between discrete nodes in a transportation network. Data mining involves the use of certain sophisticated algorithms and complex statistical models to perform two types of analytics (the Descriptive and Predictive analytics) (Oatley, 2021). The Descriptive analytics basically means turning collected data into useful insights for monitoring, reporting as well as visualization purposes while the Predictive analytics involves the use of those data to make future predictions and support decisions (Saggi & Jain, 2018).

- As the name implies, Data fusion is an analytic technique used in the consolidation of data produced from numerous sources in order to produce a more consistent, useful and accurate insight more than what will ordinarily be provided by a single data source (Khan et al., 2021). An example is the use of location data produced by mobile phones as well as such GPS-enabled vehicles.
- Optimization is a technique used in the reorganization of complex systems and processes in order to improve performance using one or more parameters (Roy et al., 2018) for example travel time or fuel efficiency. While Visualization is a techniques used for generating images, animations and diagrams in order to better communicate the results of data analysis (Ajibade & Adediran, 2016). Visualization is an important technique for identifying a qualitative understanding which will aid to explore datasets, extract dataset as well as help in the identification of patterns, outliers and corrupt data. This technique can be used during and after data analytics to make sense of the information.

2.3. The Background of Big Data in the Maritime Industry

It is beyond doubt that seaborne trade accounts for an estimated 90 percent of world trade in terms of volume (Matekenya & Ncwadi, 2022). Although, the maritime industry is traditionally not an information intensive industry as a result of the limitations associated with ship-to-shore communication. However, the quest to ensure safety, minimize cost and improve productivity as well as the recent advancement in technology (development of navigation systems, tracking systems and sensors), the maritime industry is beginning to open up to digitalization. The growing amount of available data regarding ship navigation and performance results in a broad spectrum of possibilities, ranging from real-time monitoring of ships to the extraction of knowledge through data analytics.

Few years back, the traditional practice was for merchant ships to totally disconnect from broadband communications at the time of leaving the ports, however, in recent times there is an ever increasing availability of VSAT (Very Small Aperture Terminal) which allows merchant ships to remain connected even after disembarking from the ports. This connectivity helps in providing the ship with high-speed internet and phone access, data on weather, ship reporting, bunker consumption analysis as well as aiding the ship to make order for supplies while at sea with resultant time savings.

A study conducted by COMSYS (2015) revealed that from 2008 to 2014, the number of active maritime VSAT installations quadrupled from about 6000 to approximately 22,000 and the study further predicted that these number of active maritime VSAT installations will most definitely exceed 40,000 by 2018 and expects most classed ships to be broadband capable in 2020 (Sekkesaeter, 2017).

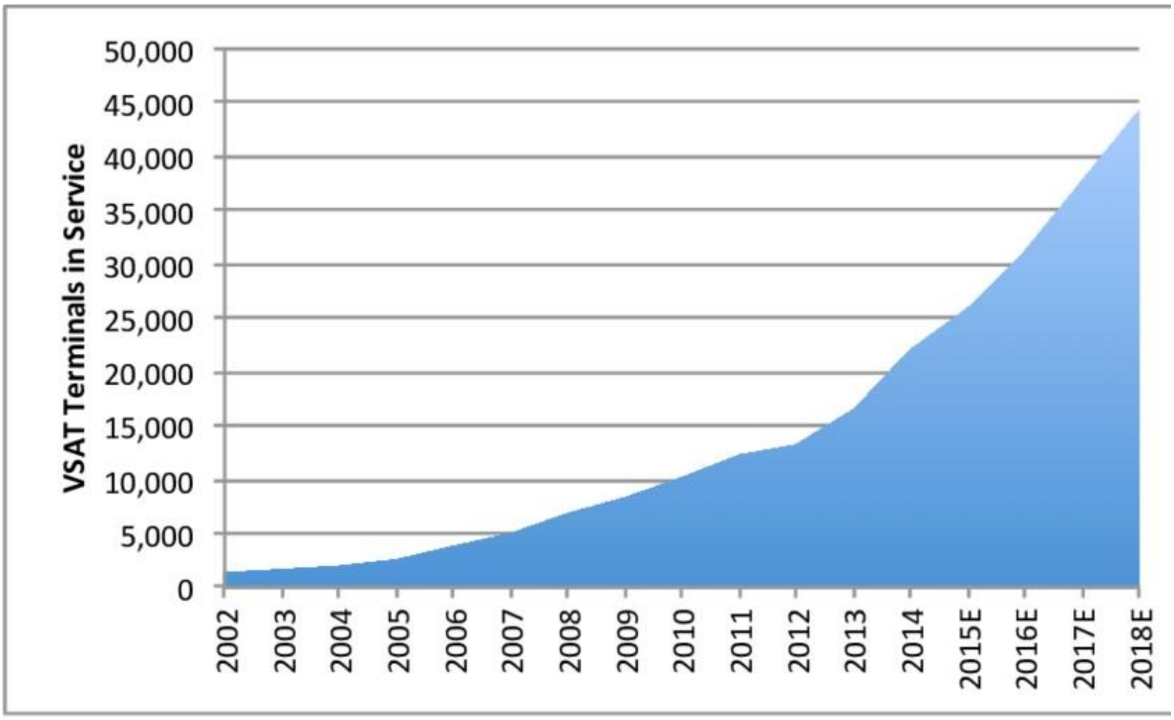


Figure 3: Maritime VSAT Installations on board Vessels (VSAT Terminals) (Sekkesaeter, 2017).

Showing the growth of Maritime VSAT installations on board vessels from the year 2002 to 2018

According to DNV GL (2015), Big Data revolution is beginning to take place in the maritime industry as a broad range of communication technologies is being deployed across the industry. Maritime connectivity is beginning to improve both in coverage and bandwidth, for example VDES (VHF Data Exchange System) and cellular networks are now present in some coastal areas, there are now Wi-Fi networks in ports and satellite communications like VSAT (Very Small Aperture Terminals) equipment are now installed on board ships.to transmit audio, text, and video data using satellite broadband services, thus indicating traffic stream of maritime big data.

According to the European Commission (EU) the existence of big data in the maritime industry is evidenced by the fact that there are approximately 150,000 ships globally that carry transponders and each one of these transponders have the capacity of sending over 10,000 messages on a daily basis (European Union, 2016) about their position to avoid collision. Similarly, the research agency Futurenautics' (2016) survey on ship operators revealed that more than half of these ship operators had at one time or the other undertaken some form of data analytics of data collected either from the on-board sensors or other ship applications. The survey further found that the main areas where big data analytics was undertaken was the operations and IT/Network areas of deployment.

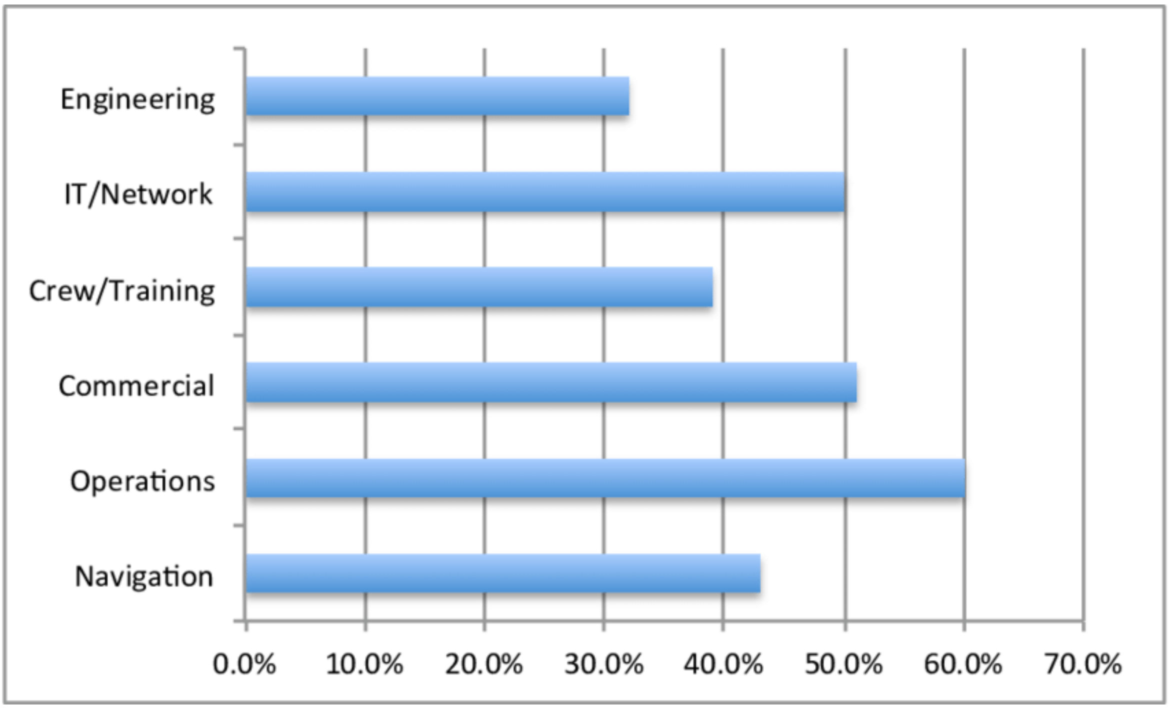
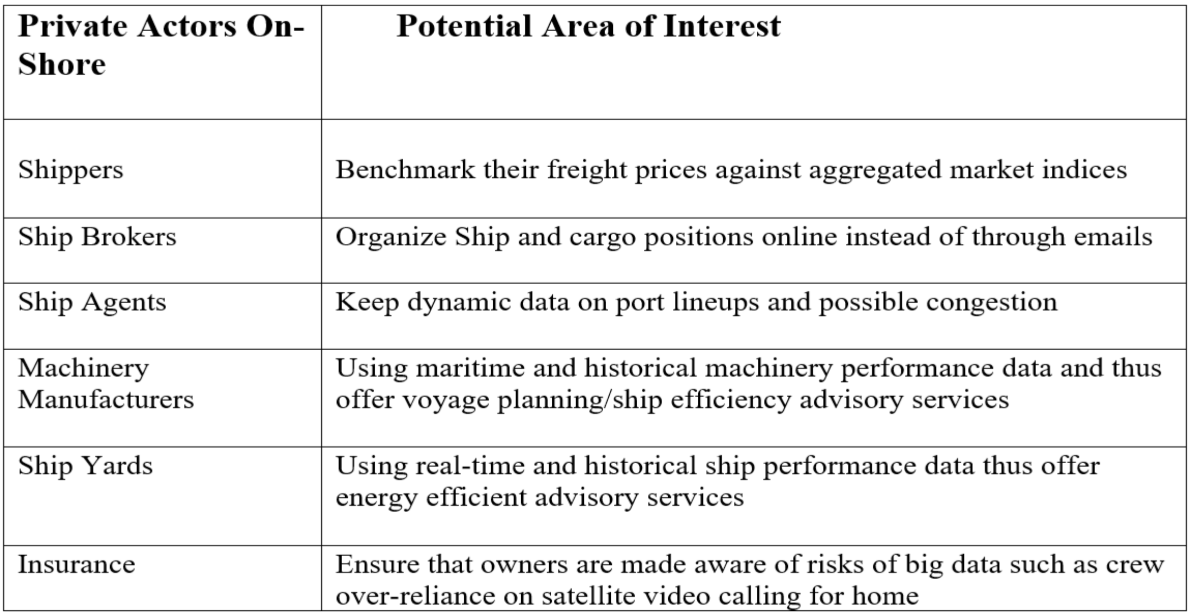


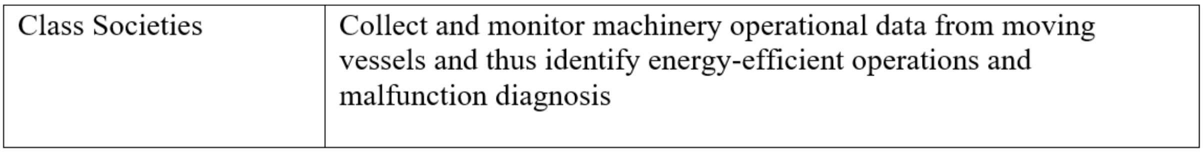
Figure 4: Applications in which Data Analytics are currently performed in the Maritime Industry (Sekkesaeter, 2017)

2.4. Use of Big Data in the Maritime Industry

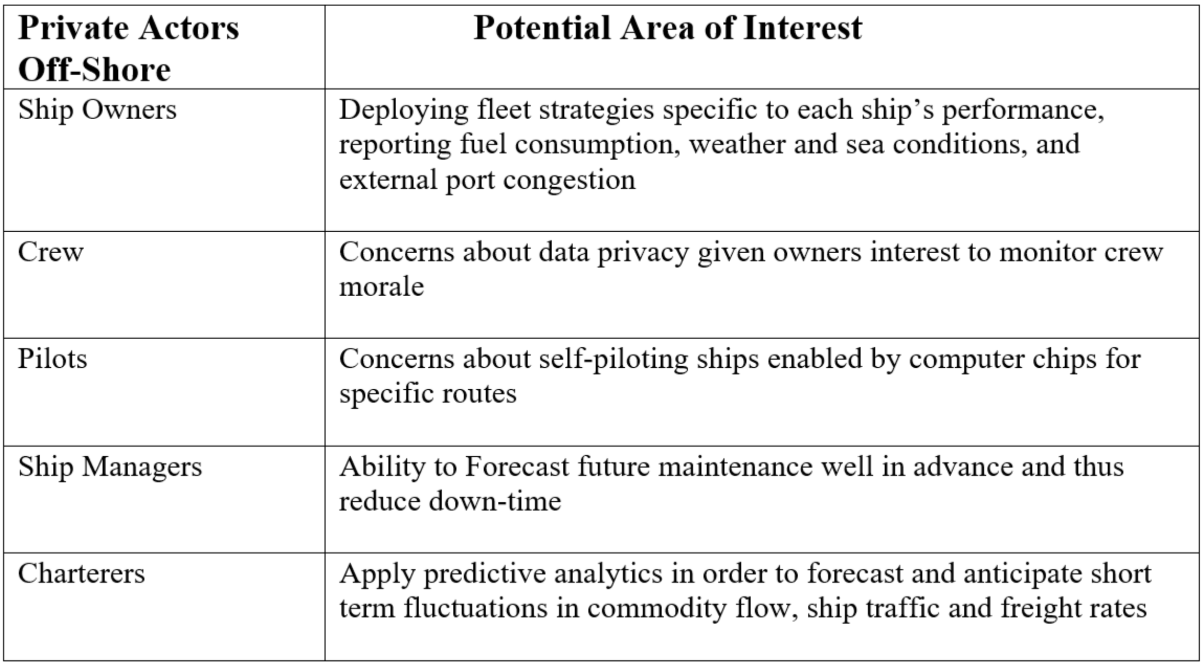
Despite the digital transformation going on in the transport and logistics industry, currently, there is limited research into the concept of big data in the context of maritime shipping, this basically explains the non-existence of a universal framework for maritime big data. According to Koga (2015) there is currently no internationally established convention/rule book on maritime big data, and this can only be explained by the fact that there are a wide range of stakeholders in the maritime industry with different data needs. While the Ship owners will most likely be interested in exploring the extent to which all their ships can provide continual performance reporting the fuel consumption for each ship, sea and weather conditions, and external port congestion, the Charterers will most likely be interested in how Big Data can be utilized for predictive analytics in estimating ship traffic and commodity flows in order to better predict the best time to charter a ship. Below is a non-exhaustive list of private and public stakeholders in the maritime industry and their various Big Data needs.

Table 1: Private Stakeholders in the Maritime Industry and their Big Data Needs (Sekkesaeter, 2017)



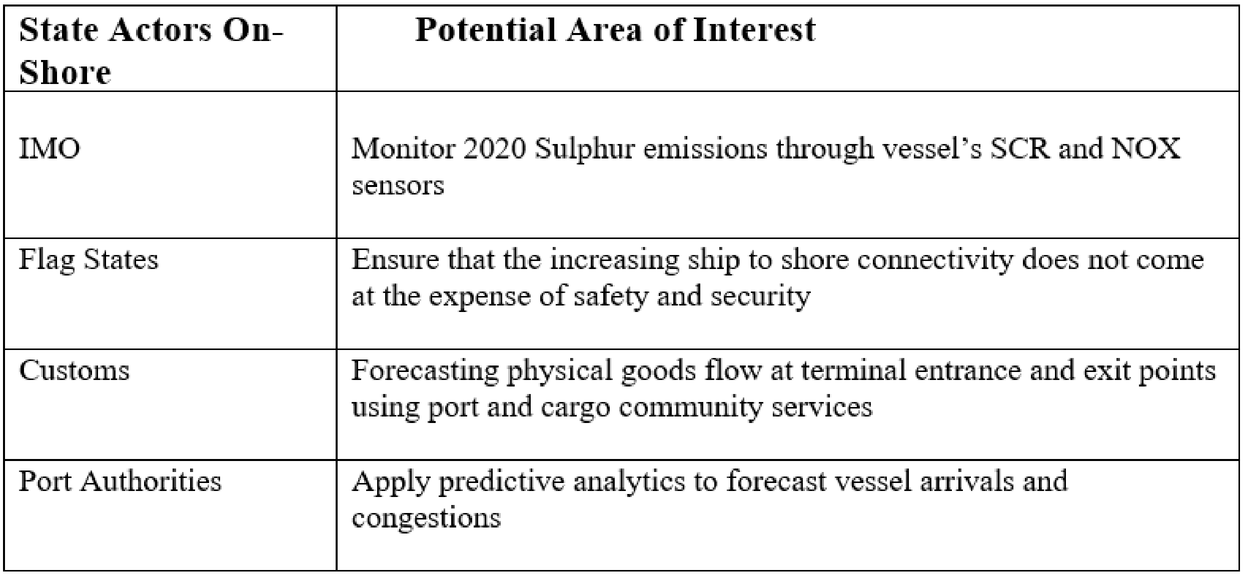
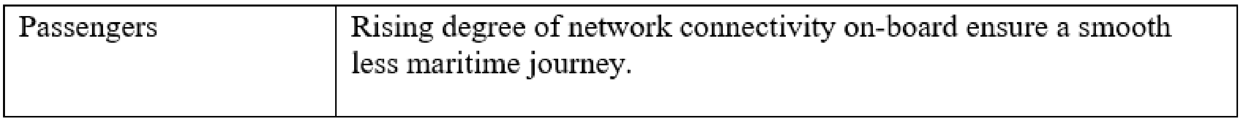


Private Actors On-Shore	Potential Area of Interest
Shippers	Benchmark their freight prices against aggregated market indices
Ship Brokers	Organize Ship and cargo positions online instead of through emails
Ship Agents	Keep dynamic data on port lineups and possible congestion
Machinery Manufacturers	Using maritime and historical machinery performance data and thus offer voyage planning/ship efficiency advisory services
Ship Yards	Using real-time and historical ship performance data thus offer energy efficient advisory services
Insurance	Ensure that owners are made aware of risks of big data such as crew over-reliance on satellite video calling for home
Class Societies	Collect and monitor machinery operational data from moving vessels and thus identify energy-efficient operations and malfunction diagnosis



Private Actors Off-Shore	Potential Area of Interest
Ship Owners	Deploying fleet strategies specific to each ship's performance, reporting fuel consumption, weather and sea conditions, and external port congestion
Crew	Concerns about data privacy given owners interest to monitor crew morale
Pilots	Concerns about self-piloting ships enabled by computer chips for specific routes
Ship Managers	Ability to Forecast future maintenance well in advance and thus reduce down-time
Charterers	Apply predictive analytics in order to forecast and anticipate short term fluctuations in commodity flow, ship traffic and freight rates

Table 2: Public Stakeholders in the Maritime Industry and their Big Data Needs (Sekkesaeter, 2017)

State Actors On-Shore	Potential Area of Interest
IMO	Monitor 2020 Sulphur emissions through vessel's SCR and NOX sensors
Flag States	Ensure that the increasing ship to shore connectivity does not come at the expense of safety and security
Customs	Forecasting physical goods flow at terminal entrance and exit points using port and cargo community services
Port Authorities	Apply predictive analytics to forecast vessel arrivals and congestions
Passengers	Rising degree of network connectivity on-board ensure a smooth less maritime journey.

2.4.1. Voyage Data Recorder (VDR)

The VDR is akin to the 'Black Box' which are installed on airplanes for accident investigation. A VDR is an equipment that is fitted on board a ship which records the various data on a ship which can be used for reconstruction of the ship's voyage details and vital information during an accident investigation.

The Voyage Data Recorder is a complete system which can store information or data in a secure and retrievable form and this data relates to the position, movement, physical status, command and control of a ship over the period and following an incident. This stored and retrievable information is then used for accident or safety investigation in order to identify the cause(s) of the accident. Other than its use in accident investigations, VDRs can also be used for preventive maintenance, heavy weather damage analysis, performance efficiency monitoring, accident avoidance as well as training purposes to improve safety and reduce running costs.

According to the Japanese Big Data Centre, MITI of NYK Group, there are fundamentally 5 examples of maritime data (Monohakobi Technology Insitute, NYK Group, 2016) and they are:

a. AIS Data (Satellite AIS and Shore AIS)
b. Voyage data (automatically collected data; noon report)
c. Weather data (forecast/past data)
d. Machinery data (automatically collected data, manual report data, maintenance data and trouble data
e. Business data (container transport data)

From the above examples of the types of maritime big data, it is clear that apart from voyage data recorders, the other main sources of ship reporting data are in the form of AIS (Automatic Identification System) or LRIT (Long Range Identification and Tracking). These types of ship reporting data are shown in Figure 5 and will be discussed broadly below.

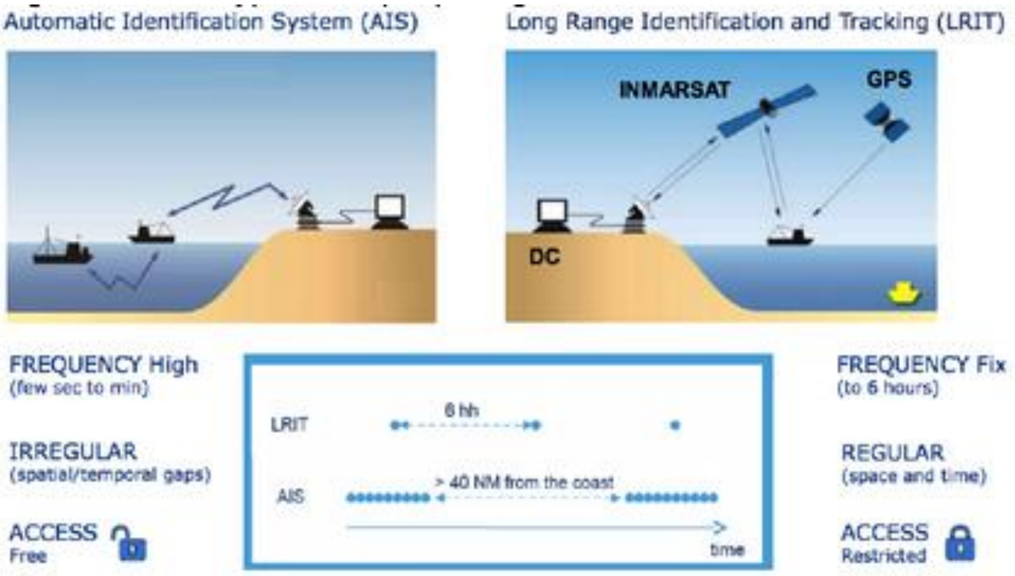


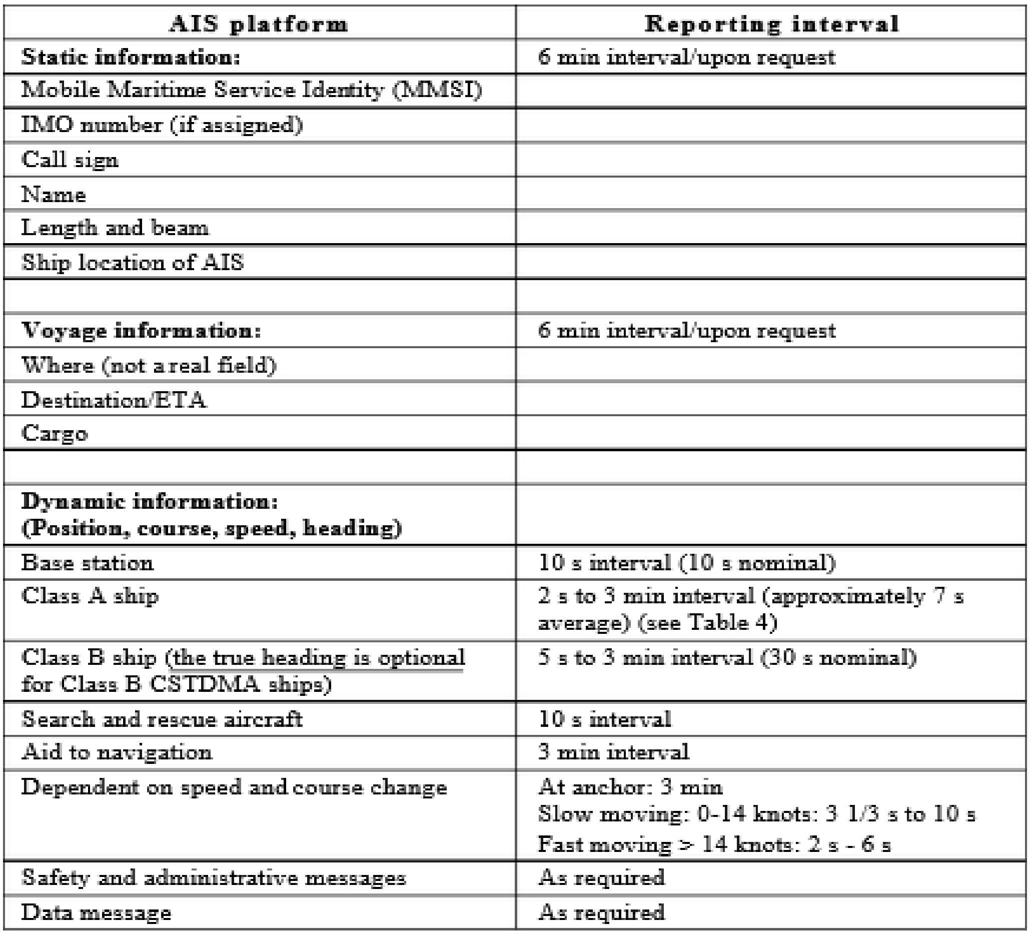
Figure 5: Main types of Ship Reporting Data (European Union, 2016)
Showing the two main types of ship reporting data (Automatic Identification System and the Long Range Identification and Tracking System) and the types of data it reports. These two systems enable the ship to transmit emergency/distress signals as communication to both ships and actors on shore, and has been very instrumental in saving thousands of lives at sea

2.4.2. Automatic Identification Systems (AIS) Data

AIS (Automatic Identification Systems) was introduced in the year 2002 as a transponder system designed in a ship to exchange navigational marine data with other ships as well as control stations and coastal authorities on shore. Traditionally, the fundamental purpose of the AIS was to increase the possibility of detecting other ships even when those ships are out of visual range or have navigational obstruction in their sea path. By the use of GPS technology and VHF (Very High Frequency) radio, AIS displays live ship traffic through RADAR, PC Charting software or ECDIS etc. (Sekkesaeter, 2017).

According to IMO, AIS (Automatic Identification Systems) are transponders that are specifically designed with the capability of providing position, identification and other information about a ship automatically to other ships and coastal authorities. AIS is regulated by SOLAS (International Convention for the Safety of Life at Sea) and in 2000, IMO adopted a new requirement as part of a revised new chapter V of SOLAS and this new provision requires all ships of 300 gross tonnage and above engaged in international voyage, all passenger ships irrespective of size be fitted aboard an AIS and this requirement became effective by 31 December, 2004 (IMO, 2016). These ships shall ensure AIS is in operation at all times except where there are international agreements, rules or standard which provides for the protection of navigational information (IMO, 2016).

The AIS can transmit a broad variety of information but generally this information can be categorized into voyage information, static information, and dynamic (kinematic) Information (ITU, 2011). These aspects of information all come with different reporting intervals, and some data is also dependent on the AIS transponder type.



AIS platform	Reporting interval
Static information:	6 min interval/upon request
Mobile Maritime Service Identity (MMSI)	
IMO number (if assigned)	
Call sign	
Name	
Length and beam	
Ship location of AIS	
Voyage information:	6 min interval/upon request
Where (not a real field)	
Destination/ETA	
Cargo	
Dynamic information: **(Position, course, speed, heading)**	
Base station	10 s interval (10 s nominal)
Class A ship	2 s to 3 min interval (approximately 7 s average) (see Table 4)
Class B ship (the true heading is optional for Class B CSTDMA ships)	5 s to 3 min interval (30 s nominal)
Search and rescue aircraft	10 s interval
Aid to navigation	3 min interval
Dependent on speed and course change	At anchor: 3 min Slow moving: 0-14 knots: 3 1/3 s to 10 s Fast moving > 14 knots: 2 s - 6 s
Safety and administrative messages	As required
Data message	As required

Figure 6: AIS Message Data and Reporting Intervals (Sekkesaeter, 2017). ***Showing AIS Message Data and Reporting Intervals, including Rates Vessel's Position and Details that are Transmitted***

2.4.3. Long Range Tracking and Identification System (LRIT) Data

LRIT which means (Long Range Identification and Tracking) is a system that provides for global tracking and identification of ships for the primary purpose of enhancing security of ships as well the safety of ship and protection of the marine environment. This LRIT system makes use of satellite technology as the method of transmission for authorities to track and monitor vessels over a bigger geographical area than what the shore-based AIS can cover. The obligation for ships to transmit LRIT information is embedded in Regulation V/19-1 of the 1974 SOLAS Convention, LRIT requires merchant ships engaged in international voyage of 300GT and above to report their ship identity (IMO number and MMSI), position

(GPS coordinates) and date/time of position to their flag state a minimum of 4 times each day (IMO, 2017), and these data can only be accessible to governmental actors such as the flag state, coastal state and port state.

2.5. Analysis of Maritime Big Data

A well-known definition of Big Data was presented by Doug Laney (2001) who explained the characteristics of Big Data using the 3Vs model. In the analysis of big data to show its application in the maritime industry, the Laney's 3V-model and the subsequent extensions to the model as earlier discussed in this study will be used.

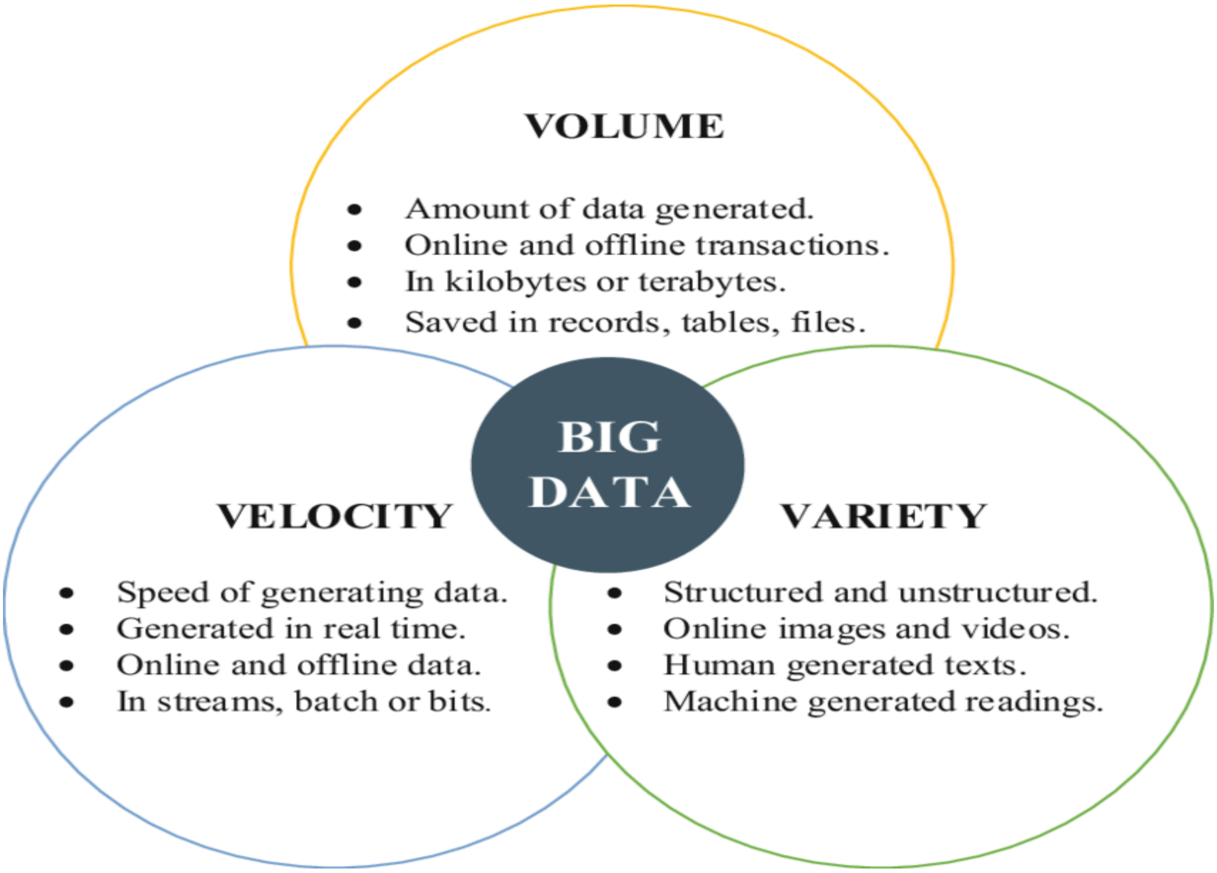

VOLUME
• Amount of data generated.
• Online and offline transactions.
• In kilobytes or terabytes.
• Saved in records, tables, files.
BIG DATA
VELOCITY
• Speed of generating data.
• Generated in real time.
• Online and offline data.
• In streams, batch or bits.
VARIETY
• Structured and unstructured.
• Online images and videos.
• Human generated texts.
• Machine generated readings.

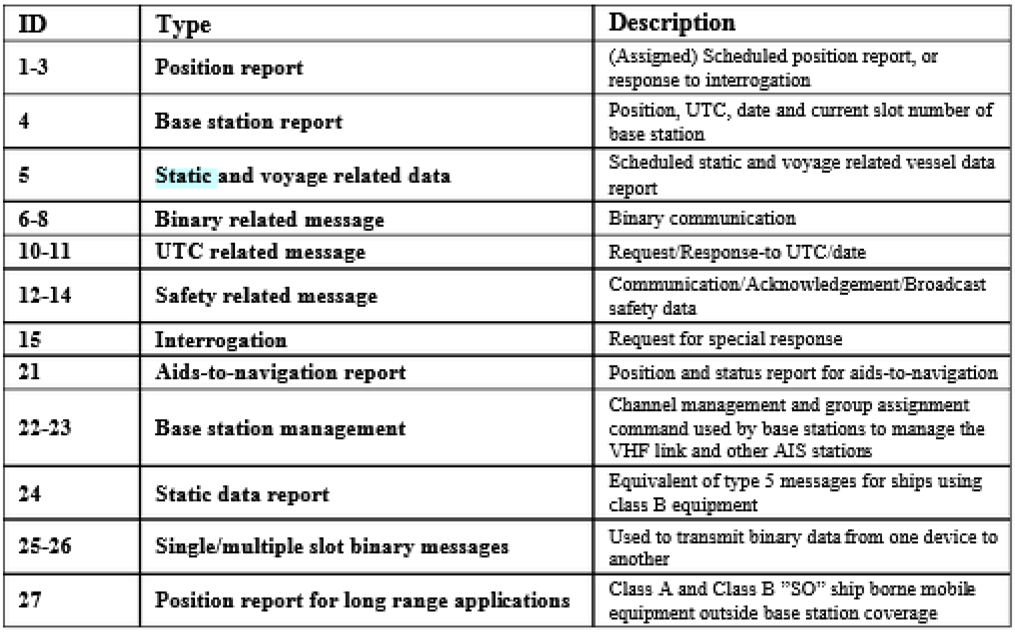
Figure 7: A Representation of the 3Vs Characteristics of Big Data (Yu & Wang, 2020)

2.5.1. Variety

Simply put Variety in big data refers to the diversity of data types. Here an organization may obtain its data from a number of different data sources that might vary in value. Messages (data) from the AIS (Automatic Identification Systems) can be generally divided into 3 categories with regards to the information it contains (Sekkesaeter, 2017):

a. Static (includes vessel's name, IMO number etc.)
b. Dynamic (includes vessel's position, speed over ground, course over ground etc.)
c. Voyage specific (includes destination, ETA, draught etc0)

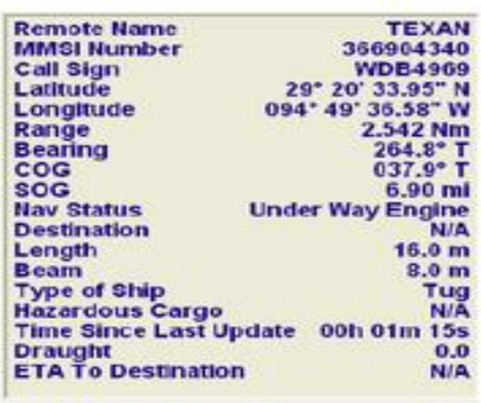
It is crystal clear from the above that the "Variety" of AIS data is significant as this fact is duly established by the wide range of purposes for which ships globally report such data. There are approximately a total number of 27 AIS messages types, out of which "AIS position report message type" is adjudged to be the most frequently used and this is as a result of its relevance for purpose of navigation and data analytics (Tu et al., 2016).



ID	Type	Description
1-3	Position report	(Assigned) Scheduled position report, or response to interrogation
4	Base station report	Position, UTC, date and current slot number of base station
5	Static and voyage related data	Scheduled static and voyage related vessel data report
6-8	Binary related message	Binary communication
10-11	UTC related message	Request/Response-to UTC/date
12-14	Safety related message	Communication/Acknowledgement/Broadcast safety data
15	Interrogation	Request for special response
21	Aids-to-navigation report	Position and status report for aids-to-navigation
22-23	Base station management	Channel management and group assignment command used by base stations to manage the VHF link and other AIS stations
24	Static data report	Equivalent of type 5 messages for ships using class B equipment
25-26	Single/multiple slot binary messages	Used to transmit binary data from one device to another
27	Position report for long range applications	Class A and Class B "SO" ship borne mobile equipment outside base station coverage

Figure 8: Most common AIS message types (PWC, 2013)
Showing the most common AIS message types, the description and its relevance to navigation and data analytics

Notably, in addition to this high variety of AIS data is also the broad range of different types of ships that sail the oceans. This is ranging from minor tugs to larger merchant ships to cruise Ships to oil & LNG tankers to container ships to vehicle carriers to offshore vessels and dry bulk carriers, each transmitting a variety of data.



Remote Name	TEXAN
MMSI Number	366904340
Call Sign	WDB4969
Latitude	29° 20' 33.95" N
Longitude	094° 49' 36.58" W
Range	2.542 Nm
Bearing	264.8° T
COG	037.9° T
SOG	6.90 mi
Nav Status	Under Way Engine
Destination	N/A
Length	16.0 m
Beam	8.0 m
Type of Ship	Tug
Hazardous Cargo	N/A
Time Since Last Update	00h 01m 15s
Draught	0.0
ETA To Destination	N/A

Figure 9: AIS message display of tug boat (Sekkesaeter, 2017)

A cursory look at Figure 10 shows the AIS data from the tug boat “TEXAN”, displayed data that shows that this is the specific data relevant to tug boats. This data might be entirely different from the data on other different types of ships. For example, an oil tanker may display Category X (A) instead of ‘’N/A’’ under ‘’Hazardous Cargo’’ (IMO, 2022). Additionally, most data entries contain text representing a significant portion of AIS data sources therefore corresponding to the widely held notion that unstructured data are abundant in Big Data.

2.5.2. Volume

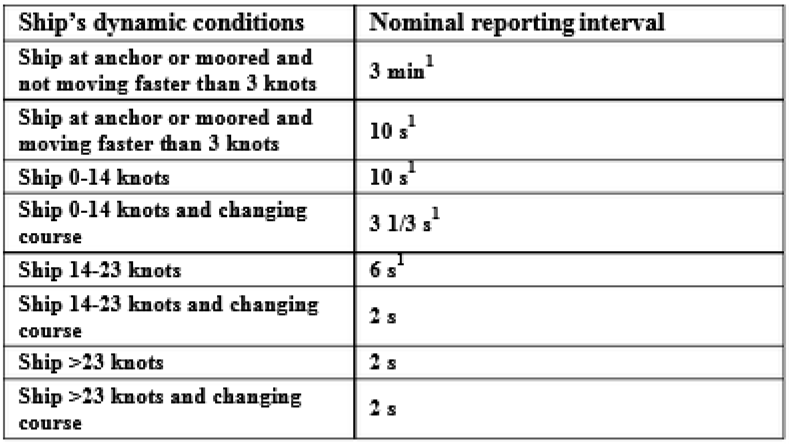
Volume simply put refers to the amount of data that exists. Where the volume of collected data is big enough, it is considered as big data. AIS data have earlier be been defined as Big Data as a result of its volume. In a study conducted in 2015 analysing a 4 months' worth of global S-AIS data revealed that raw AIS data amounted to approximately 35 Gigabytes, which was equated to be Big Data (Smestad, 2015). Similarly, in a study in the United States revealed that real world AIS data covering the vicinity of Los Angeles Harbour and Long Beach between the year 2009-2010 amounted to a data size of 58.8 Gigabytes (Wijaya & Nakamura, 2013). Additionally, it has been reported that the amount of AIS messages significantly increased from approximately 90 million in September 2007 to about 670 million in September 2012 (Cimino et al., 2014). This is as a result of the growing number of global merchant fleet as well as heavy compliance with the installation of AIS transponders by ships. Although, it is a challenging task to quantify the exact amount of AIS data that circulates globally, however, an attempt will be made in this regard.

Notably, an AIS message type 5 approximately amounts to about 424 bits (Fluit, 2011), and this encompasses both static and voyage related data which relays information on the ship itself including its current voyage. In a similar vein, the maritime ship database, Equasis (2015), estimates that the global merchant fleet of ships over 500GT stood at 87,233 in 2015 (it is important to also bear in mind that not all these ships may be installed with an AIS transponder), similarly, another study conducted in 2015 stated that approximately 85,108 of the world merchant fleet is installed with an AIS transponder and this number was arrived at by deciphering unique MMSI numbers across all AIS message types received by satellite AIS stations(Smestad, 2015).

The United States Coast Guard has stated that AIS message type 5 is transmitted every 6 minutes from ships (Sekkesaeter, 2017). Therefore, it can be safely argued that a merchant ship with a class A AIS transponder will most likely transmit approximately 87,600 AIS type 5 messages annually and this equates to about 4,642,800 bytes, or 4.64 Megabytes of data annually. Consequently, if the world fleet of merchant ships is estimated at 85,000 ships, it can be safely argued that AIS message type 5 alone reaches an estimated 394,638 Megabytes, or 394.6 Gigabytes of data on annual basis. The above analysis did not include data connected with the transmission of other common AIS message types such as type 1, type 3 and type 4, which is obviously known to be the most widely used AIS message type. This goes to show the volume of data that exist in the maritime industry.

2.5.3. Velocity

Velocity simply put refers to how quickly the data is generated as well as how quickly the data moves. Generally, the speed at which AIS data is transmitted is dependent on the type of AIS message that is been transmitted because they have differing reporting intervals. It is common knowledge that both voyage and static AIS data has a reporting interval of 6 minutes while dynamic AIS data has a shorter reporting interval which obviously depends on the speed and course alteration of the vessel (ITU, 2014).



Ship's dynamic conditions	Nominal reporting interval
Ship at anchor or moored and not moving faster than 3 knots	3 min[1]
Ship at anchor or moored and moving faster than 3 knots	10 s[1]
Ship 0-14 knots	10 s[1]
Ship 0-14 knots and changing course	3 1/3 s[1]
Ship 14-23 knots	6 s[1]
Ship 14-23 knots and changing course	2 s
Ship >23 knots	2 s
Ship >23 knots and changing course	2 s

Figure 10: Class A shipborne mobile equipment reporting intervals (ITU, 2014)

Showing a Class A shipborne mobile equipment reporting interval. The following can be noted:

(1) When a mobile station determines that it is the semaphore, the reporting interval should decrease to 2 s
(2) 1 Nautical mile= 1,852 metres. 1 knot= 1,852 m/h. 3 knots= 5,556 m/h. 14 knots= 25,928 m/h. 23 knots= 42,596 m/h

With regards to band frequency, most AIS stations operate on the VHF maritime mobile band, and this generally includes a bandwidth of 25 kHz. Further, there have been 4 international channels that have been designed for the use of AIS namely;

AIS 1, AIS 2 as well as 2 other channels namely; channel 75 and 76 for long range AIS (see Annex 4). The AIS channel 1 and 2 have a default setting of 161.975 MHz and 162.025 MHz respectively, this implies that all AIS transponders and receivers are to be set to this default setting when transmitting data except otherwise specified to be a channel management command (87B or 88B). In a study conducted by EMSA (European Maritime Safety Agency) a body that oversees the European Union *SafeSeaNet* VTS system, revealed that Europe's network of over 700 AIS data stations process 100,000,000 AIS positions on a monthly basis (EMSA, 2017). Similarly, the Norwegian Coastal Administration reported in the year 2016, the country's network of about 58 data across their coastline received about 6.7 billion AIS messages and notably this figure excluded automatic duplicates, the above figure would have been significantly higher (about 40% higher) ships (Sekkesaeter, 2017).

2.5.4. Other V's of Relevance

More recently two more Vs have emerged as key characteristics of Big Data and they are Variability and Value which are typically used in differentiating Big Data from other data and are relevant to AIS data.

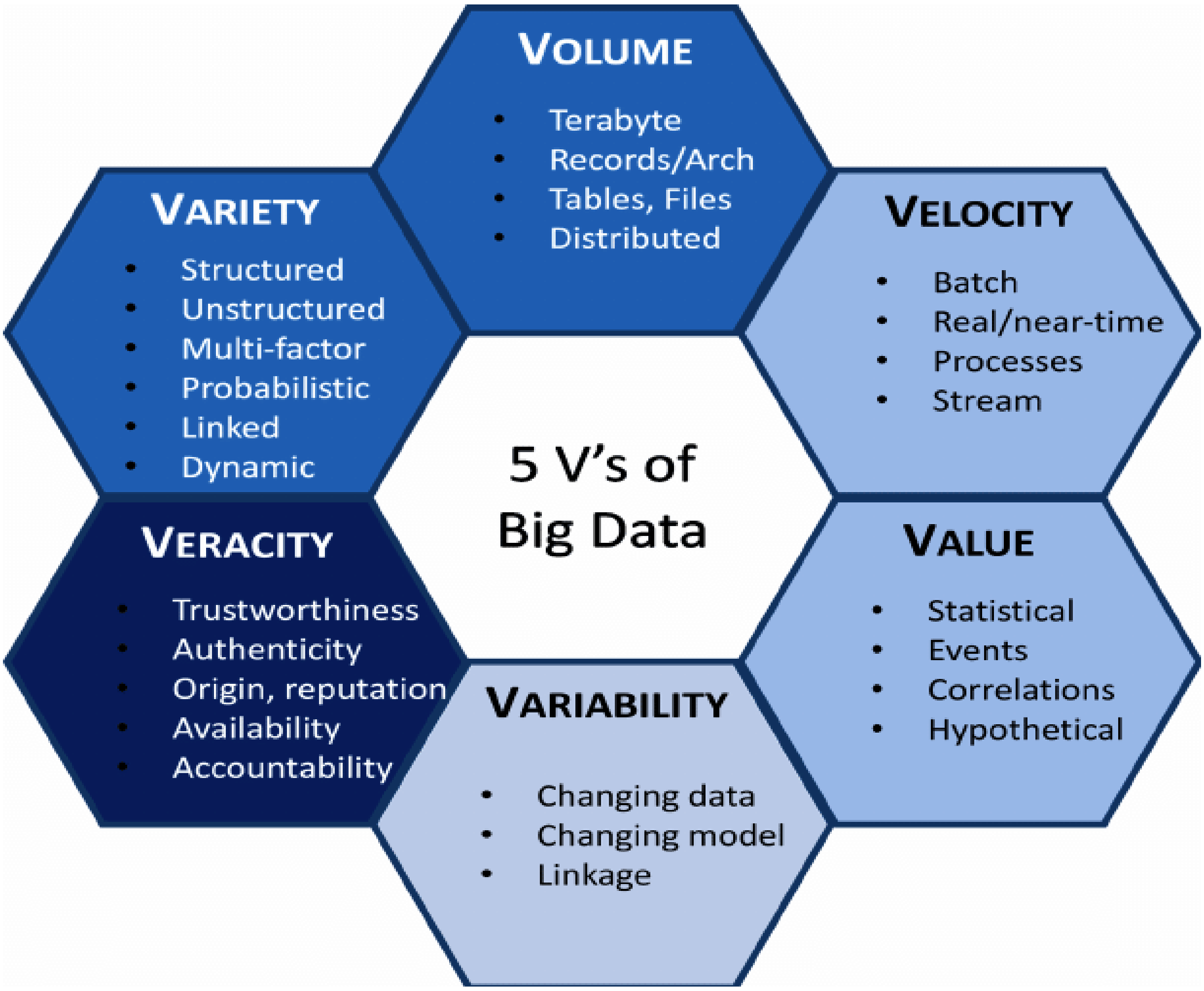


Figure 11: showing the 5Vs Characteristics of Big Data (Moura & Serrao, 2019)

Variability basically refers to the data that keeps changing constantly and Variability as it relates to AIS data is an important topic in view of the changing ship traffic (seen through the rate of AIS message transmissions by vessels trafficking through a given area). In a study conducted in the year 2011 by the Norwegian Costal Authorities in order to assess the viability of a proposed sea tunnel through the peninsula of Stad (Lampe, 2011) revealed that in Big Data visualization of how the ship traffic changed according to varying wind speeds proficiently proved that ships are inclined to sail nearer towards the coast when wind speeds increased. As the name implies, Value refers to the usefulness of the collected data and the Value dimension of AIS data is also a model example that is in agreement with the theory of Big Data. Typically, the Value of AIS data rises logarithmically in accordance with the volume of AIS data as it increases the scope for ‘’data capture, analytics and conversion into actionable insights’’ (Gudivada et al., 2017).

2.6. Big Data, Personal Data and Privacy

The continued and ever increasing growth of data production and storage has undoubtedly been accompanied by growing concerns about the adequacy of the regulatory framework ensuring privacy. These ever growing concerns is triggered by the personally identifiable nature of most of these collected data bearing in mind that oftentimes this data are collected without the express consent and full knowledge of the data object (Hajli et al., 2021). It can be argued that even “anonymous” data” whose intent is privacy protection – for example unencrypted transmissions from tire pressure monitoring systems is now easily cross-referenced with various other sources of contextual data in a bid to link individuals to locations and trajectories (Schneier, 2015). This situation without doubt can compromise reasonable expectations of personal privacy most especially location-based data which is particularly vulnerable to privacy breaches.

In as much as big data analytics raises generic privacy threats, however, it is pertinent to note that this threat is quite different from threats which pertains to cyber-security breaches. Privacy and cyber-security are two distinct but related concepts as both concepts tend to define and enforce policies that relates to the use of computer and electronic communications, however, privacy basically refers to “the claim by individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others, and the right to control information about oneself even after divulgating it” (Steenbruggen et al., 2013), while cybersecurity intends to assess the following (PCAST, 2014):

- Identity and authentication: Are you who you say you are?
- Authorization: What are you allowed to do to which part of the system?

- Availability: Can attackers interfere with authorized functions?
- Confidentiality: Can data communications be passively copied by someone not authorized to do so?
- Integrity: Can data or communications be actively manipulated by someone not authorized to do so?
- Non-repudiation, auditability: Can actions later be shown to have occurred?

The increased use of connected services as well as network-based information in the shipping industry undoubtedly poses significant privacy and cybersecurity risks (Ervural & Ervural, 2018). Although, poor cyber-security practices will ultimately result to exposure, gathering and malicious use of personal data, however, privacy risks occurs even in fully secured systems as the misuse or inappropriate use of personal data in a fully secured system by an authorized operator can lead to a violation of privacy policy other than a security policy (Churi et al., 2021). Also, data fusion may lead to privacy violation across multiple and secured systems (Tawalbeh et al., 2020). It is important to make this distinction in order to demonstrate that focusing solely on cybersecurity as a means to ensure personal data protection is not enough as the misuse of personal data by authorized data operators can also lead to equal privacy risks (PCAST, 2014)

2.7. Privacy and Location-based Data

As earlier discussed, it is without doubt that personal data includes information like name, address, Gender, marital status, religion, employment history, finances, and unique identifiers such as passport or identity card numbers revealing information about an individual that such individual may ordinarily not want to share. It is evident that all privacy frameworks which will be discussed subsequently are particularly concerned about this type of data, it is however important to point out that location data can also be seen as very personal.

The application of sensor technology is rapidly growing and advances in sensor platform architectures in the shipping industry are likely to increase the amount of location-tagged data produced (Dabiri & Heaslip, 2018). In the shipping industry, it is no longer necessary to visit certain remote locations in a bid to upload data or collect data samples for analysis because currently remote sensing technologies collects this data autonomously by deploying a network of remote sensors capable of communicating as well as transmission of data in real time. Although, what is tracked is not data directly linked to a specific individual but sensor based platform, however, the geo-spatial data collected by these devices has the capability of revealing a lot about individuals (Bradlow et al., 2017). These platforms such as the ship-based navigation system are linked to an individual's pattern of activities. These pattern of daily activity can be extremely repetitive and predictable thereby capable of revealing an individual's religion (repeated visits to a place of worship), their political affiliation (visits to political gatherings) as well as other information that can be inferred from where they go and where they spend time. This location data can, in conjunction with similar data on other individual reveal an individual's network of friends, colleagues and acquaintances particularly when cross-referenced with volunteered data on social media and networking sites thereby compromising an individual's privacy (Graupner et al., 2012).

Currently, there exist a dilemma between the value and contribution of large-scale flows of non-anonymized data to individuals and society and the privacy risk associated with it, this is coupled with the fear that regulatory backlash that may occur against the collection and use of big data may hinder the tremendous value in it, this may as well curb innovation (Crist et al., 2015). This is also partly as a result of the fact that uses for this data may emerge only after the data has been collected or combined with other data, rendering notification of intent silent (Data, 2015). Similarly, there is also a realization that the existing regulatory framework governing the collection and use of data (especially mobility-related data) is poorly adapted to current changes in the volume and velocity of data collection.

2.8. Personal Data Protection Frameworks

2.8.1. Organization Economic Cooperation and Development (OECD) Personal Data Protection Framework and Guidelines

The OECD adopted the first internationally- agreed “non-binding” privacy principle in 1980 (updated in 2013) issued with the intent to strike a balance between the benefit of international flow of information and associated privacy concerns (Mosquera, 2017). The guideline provided rules that “*personal data is processed in a transparent manner for legitimate purposes to deliver the relevant mission and work programme. Personal data are to be adequate, relevant, kept up to date, limited to what is needed and retained for no longer than necessary*” (OECD, 2013)

Although, this guideline is not law, it has over the years served as the underlying framework for data protection and privacy laws in many jurisdictions albeit without uniformity in their adoption into national law and implementation. The OECD guidelines provides for 8 principles addressing the collection and use of data. These principles are (OECD, 2013):

- Limitations to data collection.
- Data accuracy and relevance to stated use.
- Communication of the purpose for data collection and limitations of data use to that purpose.
- Restrictions on data disclosure.
- Data safeguard and security measures.
- Transparency regarding the use, and changes in use, of personal data.
- The right of individuals to have access to or control the use of their data.
- Accountability of data controllers regarding the above principles.

Although, the OECD guiding principles has relatively been useful over the last decades, however, its continued implementation has become problematic as a result of the following factors in (OECD, 2013):

- The growing ubiquity of data collection across multiple platforms.

- The volume and velocity of data produced and collected.
- Data fusion and aggregation efforts that potentially de-anonymize data.
- The range of analytical methods and techniques that reveal information regarding individuals, their behaviour and associations and their interests.
- Ex-post data mining and re-use of data in ways that were not originally intended.
- The growing range of threats to the privacy of personal data.
- The number of actors that can either compromise personal data privacy or act to protect it.
- Citizens' insufficient or ill-informed knowledge regarding the complexity of interactions relating to the collection and use of personal data.
- The ease of access to and global availability of personal data.

Other challenges of the OECD guidelines include the fact that personal data that has been de-identified, removed or modified fall outside the scope of the guidelines, the guideline only provided for personal data that identifies an individual's name, registration number and address. This challenges were re-assessed and a revised set of principles were adopted in 2013 and this revised guideline notably maintained the original 8 principles of data privacy. There have also been a number of subsequent national and international data privacy protection guidelines that modelled the OECD principles while emphasizing or de-emphasizing certain aspects. (Cate, 2006).

2.8.2. European Union Personal Data Protection Frameworks

The European Union has over the years adopted a set of data privacy principles which basically incorporated the principles in the OECD guidelines. The 1995 EU Data Protection Directive (Directive 95/46/EC) further incorporated two additional principles namely: independent oversight of data controllers and processors; and an outline of the legally enforceable rights of individuals against data collectors and processors (Bygrave, 2017).

There is also The ePrivacy Directive of 2002 (and revised in 2009), this is an EU directive on data protection and privacy and inculcated more principles on location data. The ePrivacy Directive of 2002 defined location data as *"The latitude, longitude and altitude of the user's terminal equipment, to the direction of travel, to the level of accuracy of the location information, to the identification of the network cell in which the terminal equipment is located at a certain point in time and to the time the location information was recorded."* (Directive 2009/136/EC). One significant shortcoming of the ePrivacy Directive is that it only considers location data in the context of binding rules for telecoms operators (Cheung, 2014) thereby failing to consider other collectors, aggregators and users of location data for example the shipping industry.

Circa 2012, the EU adopted a new directive to replace the ePrivacy directive of 1995. The new directive named – the General Data Protection Directive – incorporated most of the text in the 1995 directive but however, added several enhanced privacy protections which include (Hoofnagle et al., 2019):

- The "right to be forgotten" to help manage online data privacy risks. Individuals can request that data pertaining to them be deleted if there are no legitimate grounds for keeping it.
- Improved visibility and access to one's personal data and the right to transfer one's personal data from one service provider to another.
- Requirements for clear and explicit consent to collect and use personal data.
- Improved administrative and judicial remedies in cases of violation of data protection rights.
- More robust responsibility and accountability for those collecting and processing personal data – e.g. through data protection risk assessments, data protection officers, and the principles of "Privacy by Design" and "Privacy by Default". (European Commission, 2012)

Notably, the General Data Protection Directive took a step further in simplifying the definition of personal data to mean "any information related to a data subject", this in itself clearly includes location data. Similarly, this directive upheld the need for a robust notice and an express consent before the collection and processing of personal data from data sets while also providing that the collection of data must be within its stated purpose as well as robust enforcement mechanisms for transgressions regarding personal data rules (as inscribed in national laws).

2.8.3. Asia-Pacific Economic Cooperation and Other Asian Personal Data Protection Frameworks

The APEC forum in 2005 adopted the APEC Privacy Framework which is a framework that promotes a very flexible approach to data privacy protection from all APEC member states while ensuring there is a seamless flow of information devoid of unnecessary barriers. The APEC framework builds on the OECD Guidelines however, the framework extends the principle of robust notice and express consent to include a call for "clear and easily accessible" statements, accompanied with a call for practical steps be taken to provide consent either before or at the time of collection – or soon thereafter (Barth et al., 2022). It is important to note that this concession was made because of the difficulty in putting out notice and obtaining consent in certain situations particularly situations which involves real-time or machine-to-machine digital data collection. The APEC Privacy Framework further added another principle dealing with the prevention of harm – this principle stresses the need for efforts in data protection to be commensurate with the potential for harm from personal data release or discovery (Kessler et al., 2014). This is basically because all personal data does not pose the same risk for creating negative outcomes for data subjects therefore, efforts in data protection put this into consideration (APEC, 2004)

2.8.4. United States' Personal Data Protection Frameworks

There is no single principal data protection legislation in the US, rather, there are over a hundred federal and state legislations enacted with the aim of protecting the personal data privacy of the citizens.

In response to the need for data privacy protection in the United States, the US Federal Trade Commission (FTC) in 1998 and 2000 formulated and transmitted to the US Congress the principles the Commission believes should the US data protection policy. These principles relate to (Cate, 2006):

- Notice given to citizens regarding data collection and intended uses
- The choice framework offered to citizens relating to personal data collection and use and the need to obtain consent
- The possibility for citizens to access data about themselves and to contest this data on grounds of inaccuracy or incompleteness
- The responsibility for data controllers to keep personal data secure and accurate, and
- The need for enforcement and redress mechanisms to secure the above.7

Generally, it is clear that there is a difference in approach employed by the EU and the United States in the protection of personal data. While the European Union employs a single framework for addressing personal data protection, the United States prefers the sectoral approach thereby distinguishing between the private and public sector when making rules governing the collection, processing and use of data.

2.8.5. The General Data Protection Regulation (GDPR)

The GDPR is arguably the toughest security and privacy law globally. The law was drafted and passed by the EU in 2018 to impose strict obligations on organizations globally with respect to the collection of data of persons in the EU. The GDPR includes harsh fines for persons and organizations who violates the privacy and security standards (Gruschka, 2018).

Article 5.1-2 of the GDPR outlines 7 protection and accountability principles and they are (Sharma, 2019):

- **Lawfulness, Fairness and Transparency:** This principle provides that, processing of data must be fair, lawful and transparent to the data subjects.
- **Limitation of Purpose**: this principle provides that data must be processed for the legitimate and specific purpose which was expressly specified to the data subject when it was collected
- **Data Minimization**: This principle provides that data collected and processed must be the amount absolutely necessary for the specified purpose
- **Accuracy**: This principle provides that personal data must be accurately kept and up to data
- **Storage Limitation**: This principle provides that personal data can only be stored for as long as necessary and for the purpose specified.
- **Integrity and Confidentiality**: This principle provides that data processing must be done in a way as to ensure the security, confidentiality and integrity (this can be done by the use of encryption)
- **Accountability**: This principle provides that the data controller is responsible for ensuring that there is GDPR compliance in the collection and processing of data.

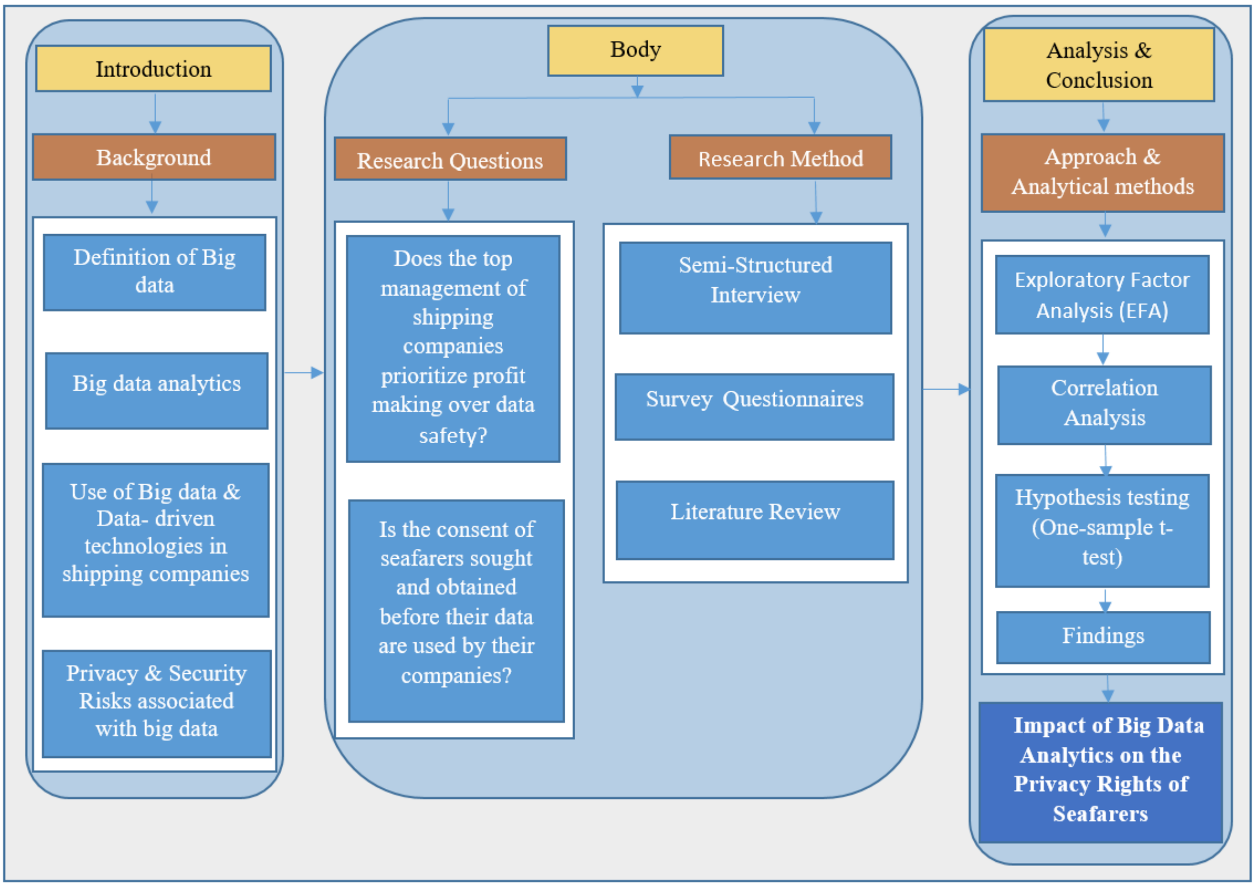
CHAPTER THREE: RESEARCH METHODOLOGY

3.1. Introduction

A Research Methodology is a strategy to systematically tackle the research problem (Kothari, 2004) and different research methodologies are offered by the associated science. However, the ability of a method to fulfil the goals and objectives of the research plays a major role in the selection of a suitable approach (Gray, 2013). This chapter discusses the research's recommended methodology, including the research plan, ethical concerns, data collecting, data analysis, validity and reliability, and the research's limitations.

3.2. Research Outline

The Illustration below in Figure 12 shows the research outline



Introduction
Background
Definition of Big data
Big data analytics
Use of Big data & Data- driven technologies in shipping companies
Privacy & Security Risks associated with big data
Body
Research Questions
Does the top management of shipping companies prioritize profit making over data safety?
Is the consent of seafarers sought and obtained before their data are used by their companies?
Research Method
Semi-Structured Interview
Survey Questionnaires
Literature Review
Analysis & Conclusion
Approach & Analytical methods
Exploratory Factor Analysis (EFA)
Correlation Analysis
Hypothesis testing (One-sample t-test)
Findings
Impact of Big Data Analytics on the Privacy Rights of Seafarers

Figure 12: Research Outline (Source: Researcher)

3.3. Research Strategy

This study investigated the impact of big data analytics in the shipping industry and its role in the infringement of seafarer's privacy rights in shipping companies. In order to achieve this, a case study of maritime companies engaging in both domestic and international voyages is used and a qualitative method based on the use of interviews and survey questionnaires has been chosen because it provides a thorough understanding of the participant's personal experience, perceptions, and knowledge on the relation between big data characteristics and privacy infringement in maritime sector.

This study was carried out using an "exploratory mixed technique." First, a qualitative exploratory study using semi-structured individual interviews was carried out to generally examine the variations between the organizational data safety and privacy practices. The exploratory part attempted to pinpoint how security and privacy infringement issues can be linked to data collection, storage, analysis, processing, sharing and accessibility. Also, the relation that exist between big data characteristics and privacy infringement that needed to be investigated further in the study. In fact, in addition to the related literatures, the study has explored the findings of the exploratory survey to assist in the creation of a quantitative data privacy questionnaire.

Following the research design approach shown in Figure 14, the data collected—both qualitatively and quantitatively—served to address the study's aim and objectives.

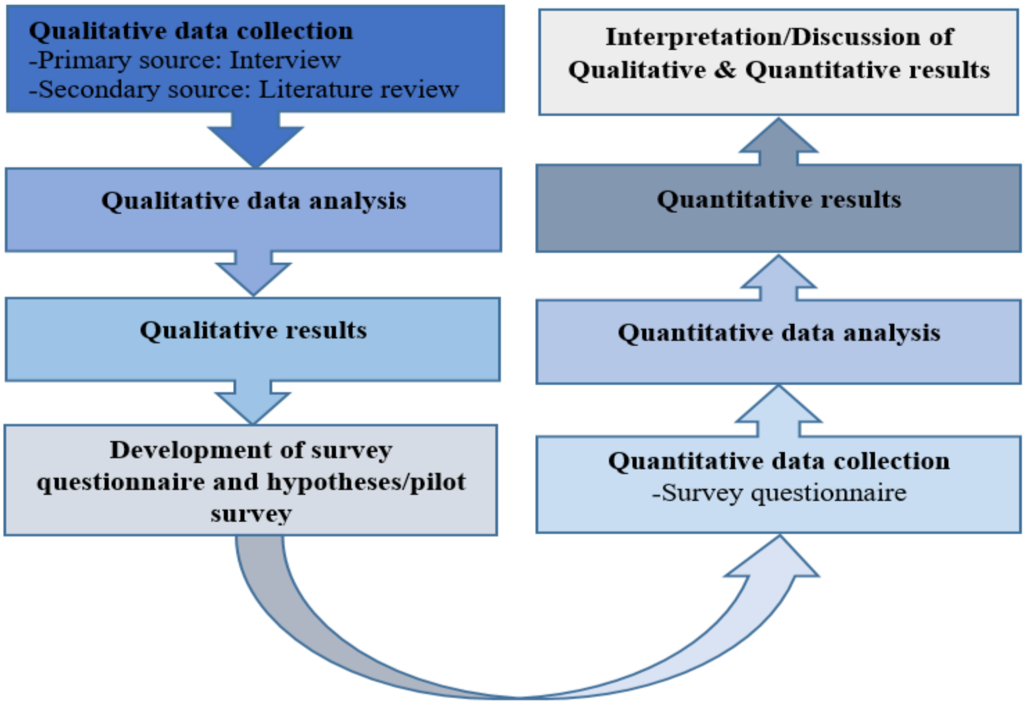

Qualitative data collection
-Primary source: Interview
-Secondary source: Literature review
Qualitative data analysis
Qualitative results
Development of survey questionnaire and hypotheses/pilot survey
Quantitative data collection
-Survey questionnaire
Quantitative data analysis
Quantitative results
Interpretation/Discussion of Qualitative & Quantitative results

Figure 13: Research Design – Exploratory mixed method (Researcher)

3.4. Ethical consideration

Following the involvement of persons in interviews and questionnaires in this study, Ethical considerations were adopted. Research Ethics Committee had to thoroughly review and approve the survey's questionnaire and interview questions. Additionally, the research took into account concerns like confidentiality, anonymity, data protection, and the option to withdraw from the study in addition to safeguarding the participants' rights and privacy. There were no payments involved with the participants' participation in the study, and their contributions were purely voluntary. Last but not least, no alterations or additions to the data obtained were made, and all data will be erased after the dissertation's submission.

3.5. Data Collection

The data collection process began by interviews and survey questionnaires on 9 July 2022 and 15 July 2022 respectively, the process was finalized on the 15 August

2022. The data collection process using questionnaires and interviews is then thoroughly discussed.

3.5.1. Qualitative Method: Personal Interview

Both primary and secondary qualitative sources of data were used in this study to gather information. Semi-structured interviews were the primary sources while previous research from the other sector such as aviation as well as international publications and journals were the secondary sources.

Semi-structured interviews were conducted to gather information about various aspects of big data privacy infringement issues in shipping companies involved in both domestic and international voyages. Given the chance to speak freely, the participants provided relevant information for the study because of their experience and expertise in the field. Both direct and open-ended questions were used in the interviews (shown in Appendix A: Personal Interviews and Appendix B: Big data Survey Questionnaire). The precise and consistent phrasing of the interview questions, the careful selection of knowledgeable participants, and the ethical consideration all served to maximize the benefits and the gathering of data that could best answer the research questions.

Four relevant participants were involved in the research question- Two seafarers, one ship owner, one data controller. Amongst the seafarers is a chief engineer with considerable experience of more than 15 years serving in both international and domestic shipping companies. Having selected the relevant participants, they were contacted through email to describe the research's aim and objective and to seek for their participation in the research. Further information about the ethical concerns was provided to those who consented to participate, and a date for the interview was then set. The interviews were done in English language.

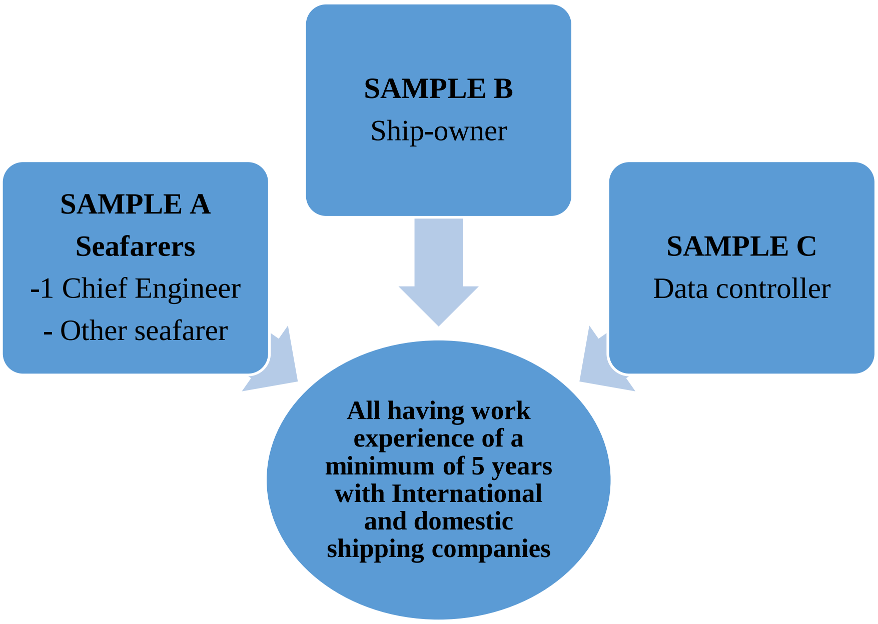

SAMPLE B
Ship-owner
SAMPLE A
Seafarers
-1 Chief Engineer
- Other seafarer
SAMPLE C
Data controller
All having work experience of a minimum of 5 years with International and domestic shipping companies

Figure 14: Composition of the interview participants (Researcher)

3.5.2. Quantitative Method: Survey Questionnaire

The seafarers employed by shipping firms that operate both domestic and oceangoing ships were the focus of the survey. However, due to the challenge of directly targeting the sample of interest using an online survey, the questionnaire was first made available to any seafarers regardless of operation of their shipping companies and their years of experience.

In the data analysis process, the targeted sample were then filtered using their background information provided. There were four samples that made up the survey population as shown in the figure 15.

- Sample A: Seafarers from shipping companies within the African continent.

- Sample B: Seafarers from shipping companies within the European continent.

- Sample C: Seafarers from shipping companies within the Asian continent.

- Sample D: Seafarers from shipping companies within the North American continent.

- Sample D: Seafarers from shipping companies within the United Kingdom

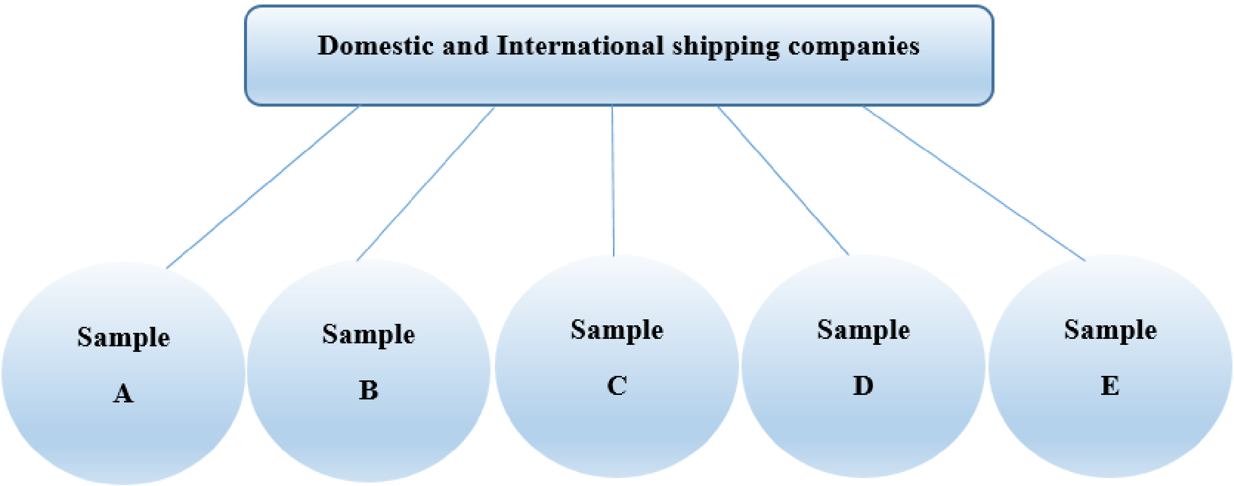


Figure 15: Survey Questionnaire's Targeted Samples (Researcher)

The survey questionnaire was developed on the following basis: the initial exploratory study and literature review from previous studies of shipping and other sectors. In actuality, both new and old survey items were employed in the development of this questionnaire. and the questions were well phrased so that seafarers of all ranks could comprehend the meaning. The initial draft of the survey questionnaire had 41 questions that covered a variety of topics related to the data privacy infringement issues. The number of questions were reduced with the aid of the exploratory study analysis leaving the final draft with 32 questions composed in 4 sections including the background information section as shown in Appendix B: Big Data Survey Questionnaires.

The distribution of the questionnaire was to quantitatively gather information from seafarers, its purpose was to get opinions on several data privacy infringement issues related to seafarers. Thus, a series of statements, such as "Consents of seafarers are not sought for or obtained before the use of their data by my company," were included in the questionnaire. Participants were asked to indicate their agreement with each statement using a Likert scale that ranged from "strongly disagree" to "strongly agree." A seven-point Likert scale (Strongly disagree, Disagree, slightly disagree, Neutral, Agree, slightly-agree, Strongly Agree) was employed in the pilot survey. The length of the question phrases was reduced by the researcher in light of comments from the pilot study and the fact that the questionnaire was administered online. Additionally, the length of the Likert scale was shortened to 5 points (Strongly disagree, Disagree, Neutral, Agree, strongly agree).

3.6. Data Analysis

The phase of personal interviews was primarily designed to act as an exploratory survey. It was meant to aid in the development of hypotheses, the construction of questionnaire parts, and the discussion section. Two research questions aided the direction of the research interviews:

RQ1: Is Top management commitment to data safety a factor that contributes to data privacy rights infringement in the context of big data analytics (the entire data continuum from collection of their data to storage, aggregation, transcription, retention, analysis and destruction).

RQ2: Are the consent of seafarers sought for and obtained before their data are used by their companies (Both domestic and international shipping companies).

Both content analysis and manual coding were used to analyse the interview data. SPSS was used to analyse the quantitative questionnaire's data.

3.7. Validity and Reliability

The mixed-method approach adopted in this research provided substantial insights and information about the research's purpose and goals. The semi-structured interview questions were thoughtfully crafted with regard to both structure and content and they were approved by the supervising professor with in-depth expertise in the marine industry.

Due to the questionnaire's mix of old and new items, which no prior study has examined for validity or reliability, the validity and reliability assessment of the questionnaire was carried out in the data analysis for the quantitative phase. Additionally, by including key stakeholder like the Ship owner, Chief Engineers as well as a sizable number of seafarers (163 in total) in the study has helped to provide a reliable and valid research outcome.

CHAPTER FOUR: DATA ANALYSIS AND DISCUSSION

This study in chapter 3 above has succeeded in providing a theoretical framework as well as the methodology which will be used in the conduct of this research. Therefore, this study will take a step further to discuss and analyse the result from the data analysis. This analysis will be divided into two parts; the first part will consist of qualitative data analysis while the second part of this chapter will analyse the quantitative data.

4.1. Qualitative Data Analysis

Generally, a qualitative data analysis involves the gathering, structuring and interpretation of qualitative (non-numerical and unstructured data) to understand what it represents. These data include text, interviews or open ended responses to survey questions. The survey conducted for this study was primarily to explore the topic of this research as well as help in the formulation of possible hypotheses. Four interviews were conducted for this study and the four respondents are all stakeholders in the maritime industry. The interview was conducted in order to garner in-depth information on the impact of big data analytics on the privacy rights of seafarers and whether the consent of seafarers is sought and obtained along the entire data continuum from data extraction, storage, aggregation, transcription, retention and analysis. The duration of each recorded interview took between 20 to 30 minutes, after which the interview findings were categorized with identifiable themes which helped the researcher in developing the measurement tool and hypothesis.

4.1.1. Analysis of the Interviews

There exist different steps in qualitative data analysis and these steps were thoroughly followed in the analysis of the interviews conducted for this study. The first step employed by the researcher was to prepare and organize the data, this was done by transcribing the interviews, printing out transcripts. This was followed by reading, listening to the recordings, reviewing and exploring the data to ensure accuracy between information provided by the respondents and the transcript. The next step was to create initial codes and labels to identify themes. These codes were then reviewed and combined into themes and the last step was to present these identified themes in a cohesive manner.

4.1.2. Qualitative Results

The researcher carefully selected the interviewee to provide as much in-depth information on the impact of big data analytics on the privacy rights of seafarers and whether the consent of seafarers is sought and obtained along the entire data continuum from data extraction, storage, aggregation, transcription, retention and analysis. In total, 4 interviews were conducted and analysed, the interviewee consisted of two seafarers with over 10 years working on board vessels sailing on international waters. Furthermore, a data controller was interviewed based on the fact the data controller is the person who oversees how data is used, controls and supervises the duties of the data processor, and ensures that data is used, stored, and processed by the guidelines of the GDPR (Rustad & Koenig, 2019) and finally, a ship-owner was interviewed to garner information on whether privacy rights of seafarers is a priority for shipping companies. In the course of these interviews, 3 categories were identified with 10 related themes to the topic of this research. These categories include: organizational data management, impact of big data analytics on the privacy rights and consent for the use of data along data continuum. These 3 categorization, their description and associated themes are discussed below:

Category 1: Organizational data Management

This category was developed from the response of the interviewees discussing the data policy of the company and the manner in which their organization manages these data to ensure the safeguard of seafarers rights. From the discussion in this category, 3 themes were extracted: (1) Non-effectiveness of data policy (2) prioritization of profit making over data safety (3) mechanism in place to guard against data theft and cyber-attacks.

Category 2: Impact of big data analytics on the privacy rights

This category arose from the interviewee discussing how big data analytics by the organization impacts on the privacy rights of seafarers and how personal information is used for online profiling, tracking and for ads. From the discussion in this category, 8 themes were extracted: (1) unlimited data hunger (2) data exploitation (3) extensive use beyond legitimate purpose (4) abuse of personal information (5) personal profiling (6) personalized information for online tracking (7) inadequate data safety training (8) No use of data encryption

Category 3: consent for the use of data along data continuum

This category arose from the interviewee discussions on the failure of the organization to seek and obtain the consent of seafarers before their personal data is used and how there are no insurance policies to guard against such breach. From the discussion in this category, 5 themes were extracted: (1) right to privacy (2) lack of consent to use personal data (3) lack of insurance policies (4) no knowledge of the data controller (5) extensive use of data beyond the realms of consent.

4.1.3. Development of Main Hypothesis

Flowing from the interviews conducted, it was crystal clear that the 4 interviewees were of the unanimous view that big data analytics in their organizations have an impact on the privacy rights of seafarers, it was further unanimously agreed that the requisite consent is not usually sought and obtained before seafarer's personal data are used by the organizations. Therefore, to better address the research aim, the researcher has formulated a hypothesis and this hypothesis is tagged H1 and formulated as follows:

Hypothesis (H1): *"Big data analytics in the shipping companies negatively impacts on the privacy rights of seafarers."*

4.2. Quantitative Data Analysis

As previously mentioned, a survey questionnaire was developed by the researcher in order to assess top management commitment to seafarer's data safety as a factor that enable data privacy infringement and also to determine whether the consent of these seafarers are sought and obtained before their personal data are used. In the quantitative data analysis, the validity and reliability of these survey questionnaires were tested. The quantitative data analysis is divided into 2 sections, the first section was the preliminary analysis, including data collection and preparation for analysis as well as demographic characteristics, reliability and validity while the second section include the results of analysis and other statistical tests. This section specifically follows the sequence of the research objective and the supporting analytical methods, as shown in Figure 16.

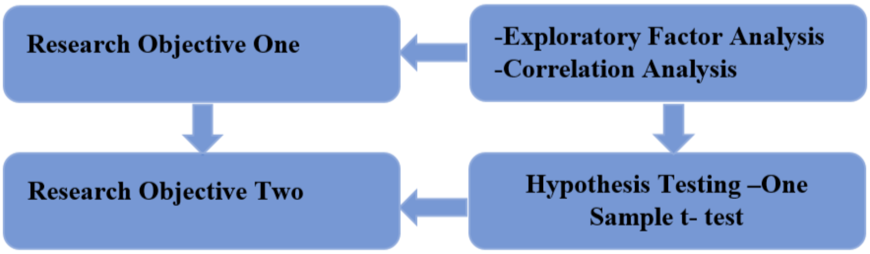

Research Objective One
-Exploratory Factor Analysis
-Correlation Analysis
Research Objective Two
Hypothesis Testing –One Sample t- test

Figure 16: The Analytical Methods employed in the Quantitative Data Analysis (Source: Researcher)

4.2.1. Data Collection and Preparation for Analysis

The survey questionnaire was distributed to respondents who consist of seafarers working on board vessels sailing on international waters and the survey questionnaire was launched online using QuestionPro Survey software between the 15th of July to the 15th of August, 2022. The survey questionnaire was viewed by 163 respondents, for which 38 respondents partially completed the survey and a total number of 125 respondents fully completed the survey. For accuracy of the analysis, the researcher filtered the result using time spent in answering the survey as a yardstick to determine accuracy of the response, hence, surveys completed under 3 minutes were filtered out. After these invalid questionnaires were filtered out, a total number of 125 responses was assumed as valid and therefore analyzed. The valid data was then organized and prepared for analysis and coded into SPSS 26.0 file. The researcher then performed different descriptive analyses using the SPSS software to ensure that data was entered correctly and respondents' demographic characteristics were extracted.

4.2.2. Respondents Demographic Data

A total of 125 valid responses from 125 respondents were analysed and the demography of the respondents were grouped according to the continent for which their companies belonged. The result showed that 53 respondents representing 42.4% are from shipping companies within the African continent, 29 respondents representing 23.2% are from shipping companies within the European continent, 20 respondents representing 16% are from shipping companies within the Asian continent, 15 respondents representing 12% are from shipping companies within North America and 8 respondents representing 6.4% are from shipping companies in the United Kingdom (see figure. 17)

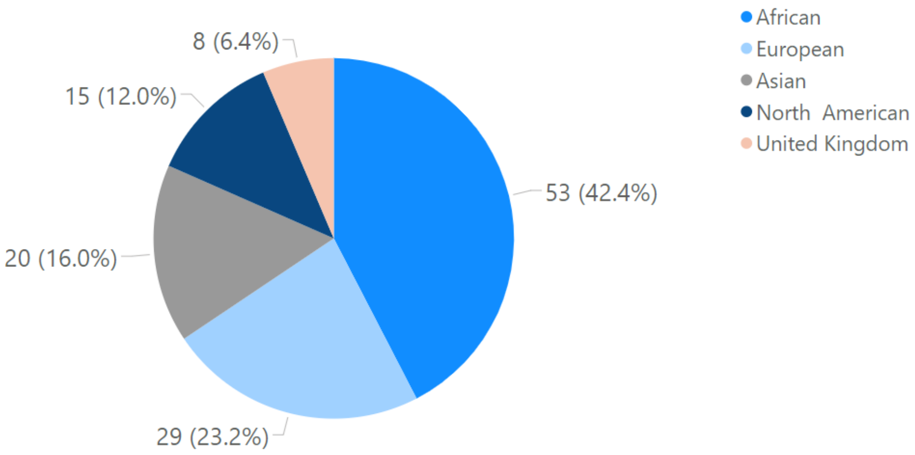


Figure 17: Graphical representation of Respondent's Continents for which the shipping companies belong to. (Source: Researcher)

The respondents were further sampled according to their personal profiles which showed essential and relevant characteristics. The personal profile showing ranks of the respondents showed that 9 respondents representing 7.2% were captains of their vessels, 15 respondents representing 12% were Chief engineers, 10 respondents representing 8% were ship engineers, 12 respondents representing 9.6% were second engineers, 10 respondents Representing 8% were third engineers, 26 respondents representing 20.8% were deck officers, 10 respondents representing 3.2% were Cadets, 12 respondents representing 9.6% were officers, 1 respondent representing 0.8% was Nautical surveyor, 26 respondents representing 20.8% were Seamen. (see figure. 18).

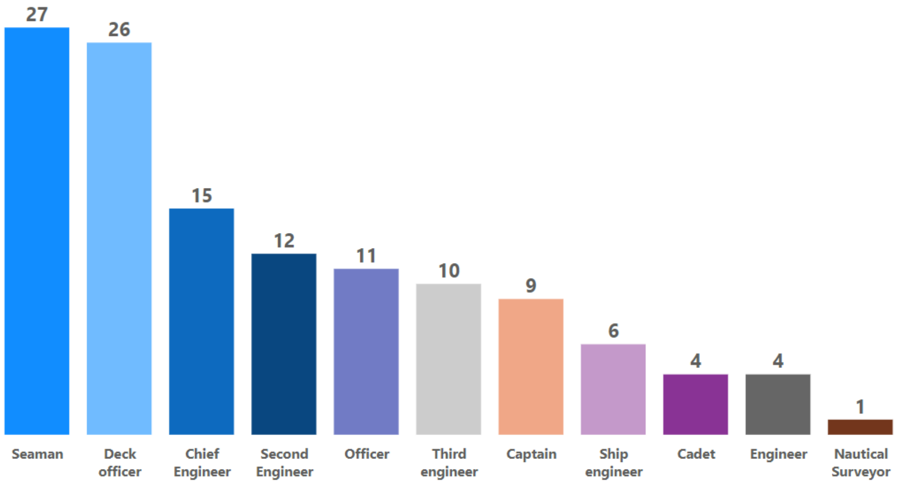


Figure 18: Graphical representation of respondent's position/rank (Source: Researcher)

Furthermore, the respondents' profile showed meaningful characteristics such as years of experience as 10 respondents representing 8% had seafaring experience below 5 years, 90 respondents representing 72% have 5-10 years' experience while 25 respondents representing 20% have above 15 years' experience (see figure 19). Similarly, 24 respondents representing 19.2% have worked in their current company for less than 5 years, 93 respondents representing 74.4% have worked in their current company between 5-10 years while 8 respondents representing 6.4% have worked in their current company for more than 15 years. (see figure 20). The demography of the respondents shows that there is a representation of both current and the future leaders of the maritime industry thus giving more credence to this study.

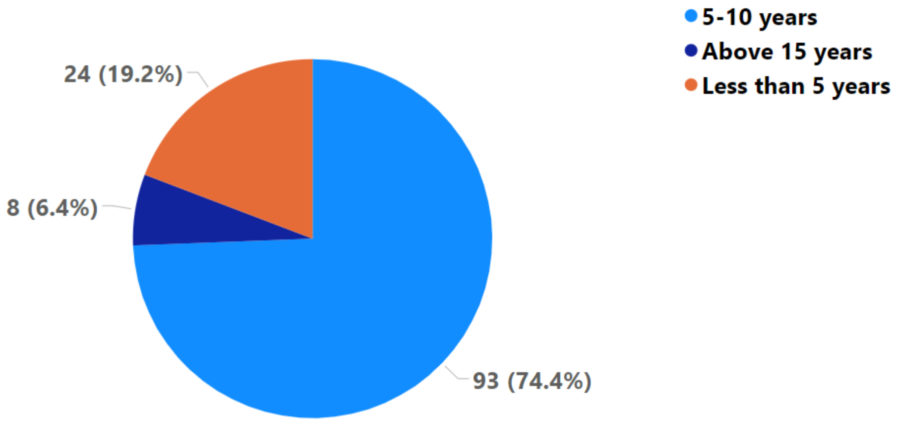


Figure 19: Graphical representation of the respondent's years of service with current company (Source: Researcher)

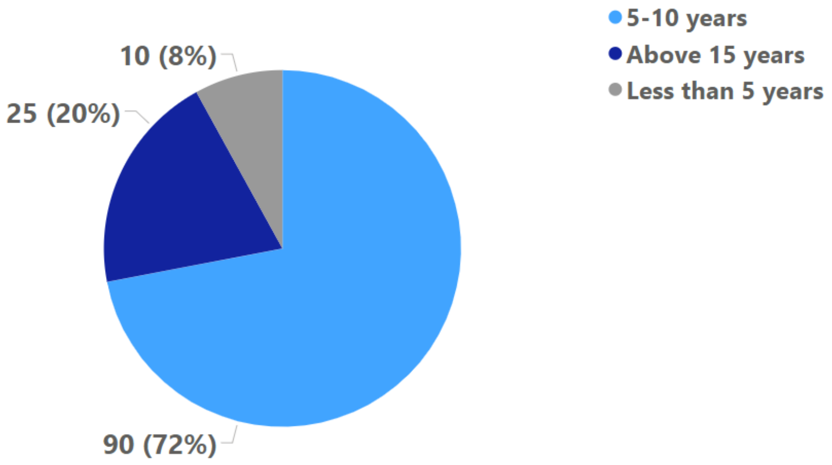
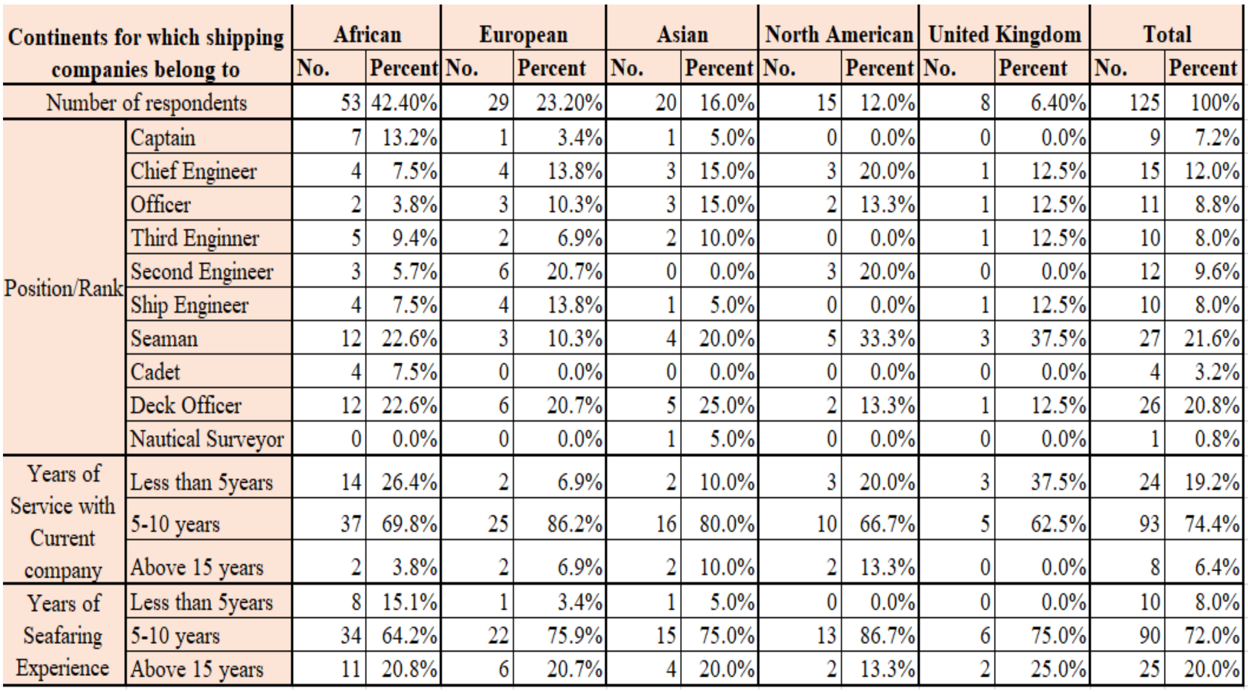


Figure 20: Graphical representation of respondent's seafaring experience (Source: Researcher)

Table 3: General characteristics of Respondents (Source: Researcher)



Continents for which shipping companies belong to		African		European		Asian		North American		United Kingdom		Total	
		No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent
Number of respondents		53	42.40%	29	23.20%	20	16.0%	15	12.0%	8	6.40%	125	100%
Position/Rank	Captain	7	13.2%	1	3.4%	1	5.0%	0	0.0%	0	0.0%	9	7.2%
	Chief Engineer	4	7.5%	4	13.8%	3	15.0%	3	20.0%	1	12.5%	15	12.0%
	Officer	2	3.8%	3	10.3%	3	15.0%	2	13.3%	1	12.5%	11	8.8%
	Third Enginner	5	9.4%	2	6.9%	2	10.0%	0	0.0%	1	12.5%	10	8.0%
	Second Engineer	3	5.7%	6	20.7%	0	0.0%	3	20.0%	0	0.0%	12	9.6%
	Ship Engineer	4	7.5%	4	13.8%	1	5.0%	0	0.0%	1	12.5%	10	8.0%
	Seaman	12	22.6%	3	10.3%	4	20.0%	5	33.3%	3	37.5%	27	21.6%
	Cadet	4	7.5%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	4	3.2%
	Deck Officer	12	22.6%	6	20.7%	5	25.0%	2	13.3%	1	12.5%	26	20.8%
	Nautical Surveyor	0	0.0%	0	0.0%	1	5.0%	0	0.0%	0	0.0%	1	0.8%
Years of Service with Current company	Less than 5years	14	26.4%	2	6.9%	2	10.0%	3	20.0%	3	37.5%	24	19.2%
	5-10 years	37	69.8%	25	86.2%	16	80.0%	10	66.7%	5	62.5%	93	74.4%
	Above 15 years	2	3.8%	2	6.9%	2	10.0%	2	13.3%	0	0.0%	8	6.4%
Years of Seafaring Experience	Less than 5years	8	15.1%	1	3.4%	1	5.0%	0	0.0%	0	0.0%	10	8.0%
	5-10 years	34	64.2%	22	75.9%	15	75.0%	13	86.7%	6	75.0%	90	72.0%
	Above 15 years	11	20.8%	6	20.7%	4	20.0%	2	13.3%	2	25.0%	25	20.0%

4.3. Reliability and Validity

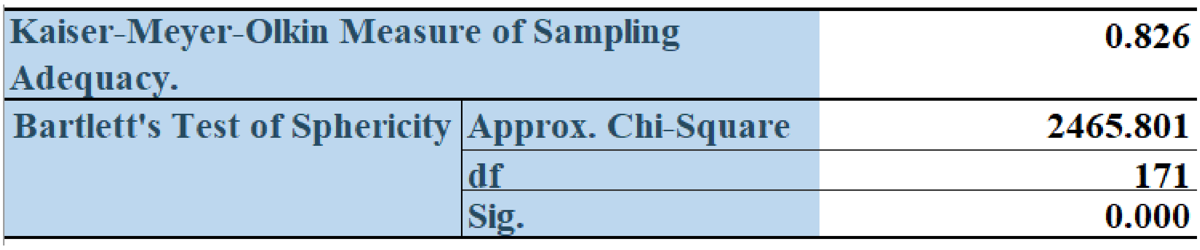
4.3.1. Exploratory Factor Analysis (EFA)

The whole essence of a survey questionnaire is to gather all relevant information from the respondents, however, the validity and reliability of these measurement tools is of paramount importance when designing questionnaires and performing statistical analysis. While validity is important to confirm the accuracy of the survey questionnaire, reliability is essential to show the consistency of the survey tool. Hence, where a measurement tool is shown not to be reliable, it cannot be said to be valid.

The researcher performed an exploratory factor analysis (EFA) using SPSS 26. In order to explore underlying factors as well as structure the dataset. Firstly, the Bartlett test of sphericity and Kaiser-Mayer Olkin (KMO) test were performed by the researcher in order to determine the data suitability to conduct an EFA.

The obtained values for both tests were satisfactory (i.e., more than 0.8 for KMO's test and less than 0.01 for the Bartlett test of sphericity), as shown in Table 4.

Table 4: KMO and Bartlett Tests Result (Source: Researcher)



Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.826
Bartlett's Test of Sphericity	Approx. Chi-Square	2465.801
	df	171
	Sig.	0.000

This study used SPSS's component extraction and Varimax rotation in the analysis of the data and also considered the below criteria: (1) Pairwise deletion (2) a minimum of 3 items in the factor with minimum loading of 0.5 and (3) Eigenvalue more than 1.0 proposed by Awang (2014) who stated that for the established items, the factor loading for every item must exceed 0.5.

The result of the exploratory analysis showed the extraction of 3 factors which cumulated in 72.939 of the entire variances with factor 1 ,2, and 3 having %variance of 36.773, 29.074 and 7.092 respectively. Also, the items were further categorized based on their inter-correlation and relationships and the result is shown in table 6 below. Therefore, the 3 extracted factors which represents all aspects relating to big data analytics and its impact on seafarer's privacy rights are as follows:

Factor 1: *Top management commitment to data safety* – this reflects the perception of seafarers on the commitment of the top management towards privacy rights protection and this reflects the perception of seafarers on the existing data management policies in place aimed at protecting seafarer's privacy rights along the entire data continuum – from data sensing to its extraction, storage, aggregation, transcription, retention, analysis and retention.

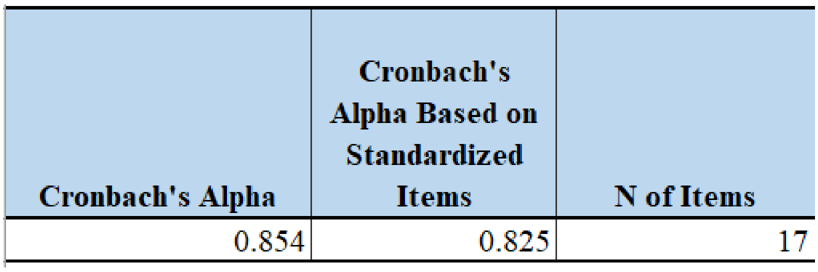
Factor 2: *Consent for the use of personal data* - this reflects the perception of seafarers on whether their consent is sought and obtained by the organization before their personal data is used and if their personal data can be used without consent.

Factor 3: *Duration of personal data collection* –this reflects the perception of seafarers on the number of years their personal data has been collected, analyze and processed by the shipping companies. This period takes into account the number of years their personal data has been kept and retained by the company without their consent.

4.3.2. Reliability and Validity Test

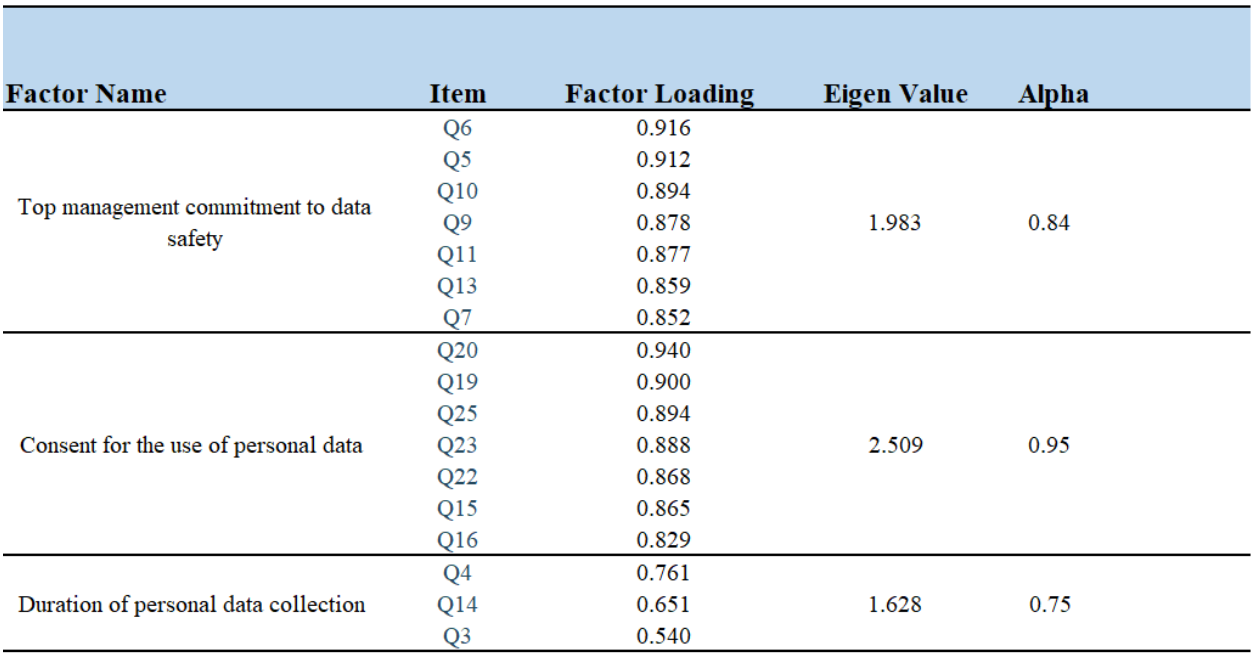
Reliability test estimates the internal consistency of items combined in the same factor using the Cronbach Alpha value, a scale is reliable if the Cronbach alpha value is greater than 0.70 (Hair et al., 2013). The result of reliability test as shown in table 6 revealed that the *Top management commitment to data safety* scale with six items (Alpha = 0.84), *Consent for the use of personal data* scale with seven items (Alpha = 0.95) and *Duration of personal data collection* scale with three items (Alpha = 0.75) were found reliable. In carrying out the Exploratory factor analysis six items were removed due to failure to produce result for the KMO and Bartlett test of sphericity. Similarly, two items were removed in the reliability test step due to failure to produce an alpha value of a minimum of 0.7 as required for reliability. In total, eight items were removed and a re-run of the test confirmed that all other 17 items satisfied the reliability requirement with a Cronbach alpha value of 0.854. Also, the rule for minimum loading of items in each factor that says the factor loading for every item must exceed 0.5 (Awang ,2014) was followed, as shown in Table 6.

Table 5: Reliability Statistics (Source: Researcher)



Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
0.854	0.825	17

Table 6: Exploratory Factor Analysis and Reliability Test Result (Source: Researcher)



Factor Name	Item	Factor Loading	Eigen Value	Alpha
Top management commitment to data safety	Q6	0.916	1.983	0.84
	Q5	0.912		
	Q10	0.894		
	Q9	0.878		
	Q11	0.877		
	Q13	0.859		
	Q7	0.852		
Consent for the use of personal data	Q20	0.940	2.509	0.95
	Q19	0.900		
	Q25	0.894		
	Q23	0.888		
	Q22	0.868		
	Q15	0.865		
	Q16	0.829		
Duration of personal data collection	Q4	0.761	1.628	0.75
	Q14	0.651		
	Q3	0.540		

4.3.3. Normality Test

Statistical tests are usually used when testing hypothesis. These tests are used in determining if a predictor variable maintains a statistically significant relationship with an outcome variable and also to determine the difference between two or more groups.

In conducting this statistical analyses, there may be a need to use different statistical test in order to achieve a correct result and the nature of the distribution of data determines the choice of test to be utilized.

Clearly, these various tests utilized to check the normal distribution of data are provided for in various literatures and they include methods like graphical assessment of normality, Kolmogorov-Smirnov test, Shapiro-Wilk test, Lilliefors corrected K-S test, D’Agostino skewness test, and Jarque-Bera test (Kahlouche, 2021). However, the Kolmogorov Smirnov test seems to be the most popular among the aforementioned tests and the method of doing this is that a p-value is greater than 0.05 (Kahlouche, 2021). So, in order to determine if the sample comes from a population with a specific distribution the Kolmogorov Smirnov test for independent samples was used in addition to visual observation.

The Kolmogorov Smirnov test for independent samples that was performed using SPSS to test that the null hypothesis formulated from the data flows from a normal distribution. The result showed that there was no notable violation of normality. Thus, a parametric test was used in the following data analysis rather than a non-parametric tests. The figure 21 below shows sample of the normally distributed data.

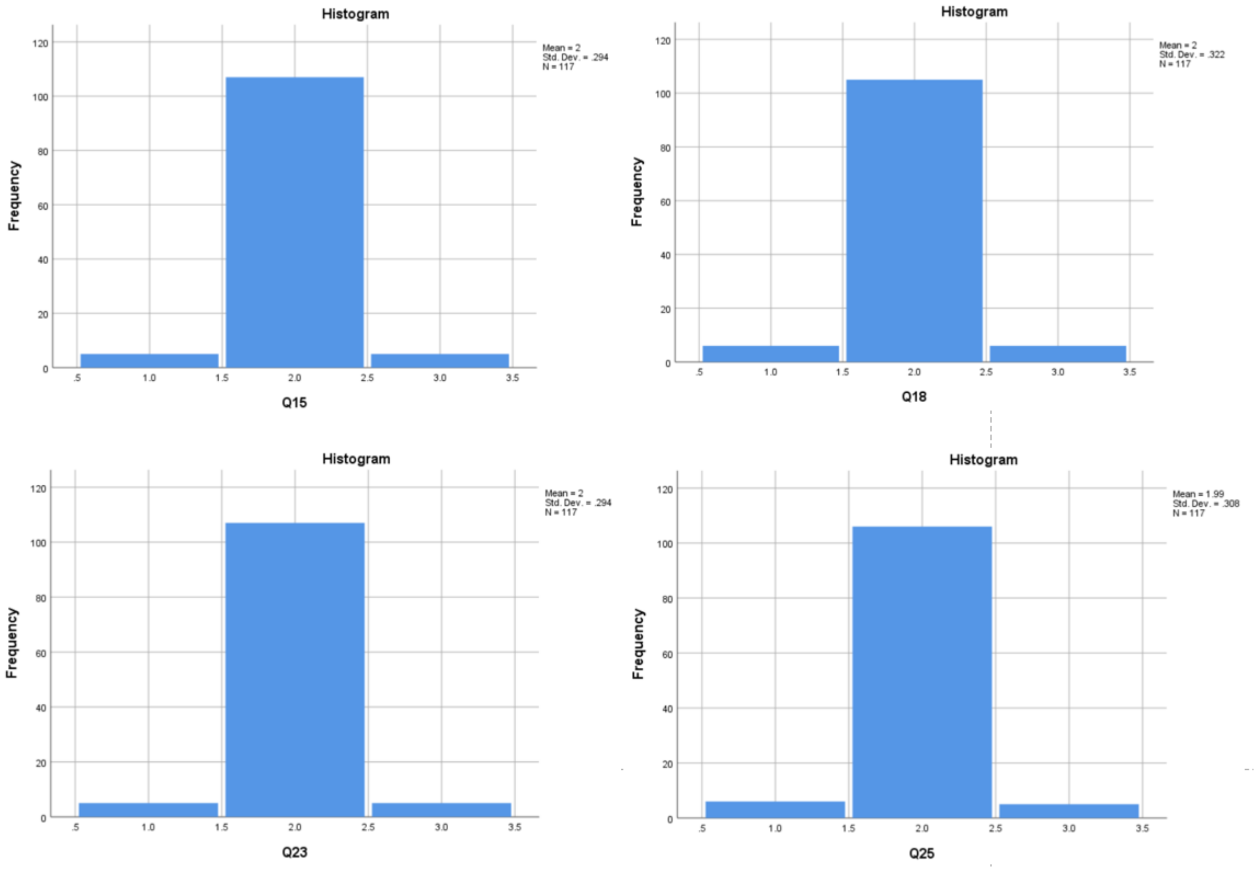


Figure 21: Normally distributed samples (Source: Researcher)

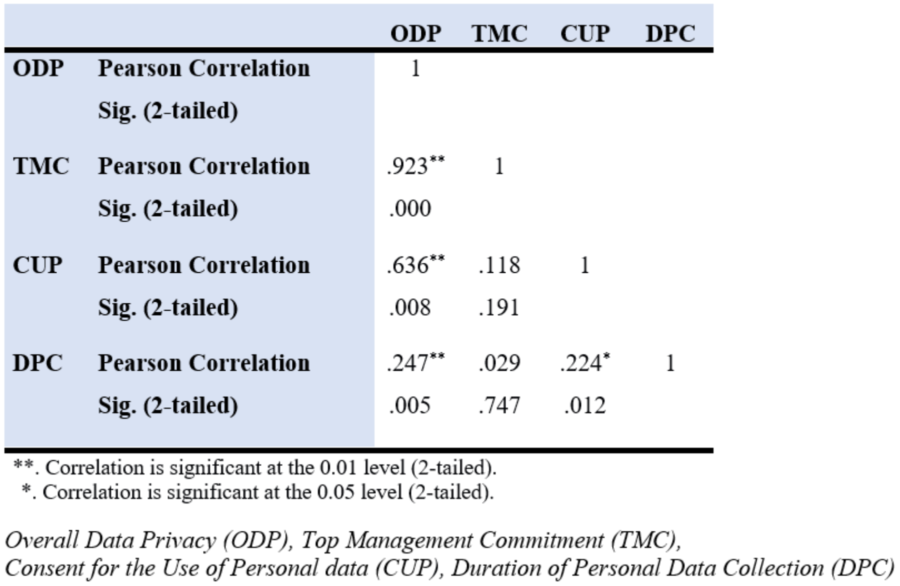
4.4. Analysis of the Questionnaire Data

The questionnaire data was analysed in such a way that the findings addresses the research objectives as described in Figure 16. The first research objective was to assess top management commitment to data safety as a factor that enables data privacy infringement. Exploratory factor analysis resulted in extracting two other factors alongside top management commitment to data safety as factors that enable data privacy infringement in the context of big data analytics. Afterwards a correlation analysis was carried out to assess the degree of correlation of the extracted data privacy factors and also to identify the most influential factors on data privacy.

4.4.1. Correlation Analysis of the Data Privacy (DP) factors

Using the SPSS, the Bivariate correlation analysis was carried out in order to assess the relationships amongst the three data privacy factors. The data used in this analysis were those collected from 125 seafarers represented in samples A, B, C, D, E and the result of the analysis is interpreted in terms of the value of Pearson correlation coefficient such that the closer the value is to ±1, the higher the correlation, the closer it is to 0, the lower the correlation of the factors (Field, 2005). Table 7 shows the result of the correlation analysis.

Table 7: Correlation analysis results (Source: Researcher)



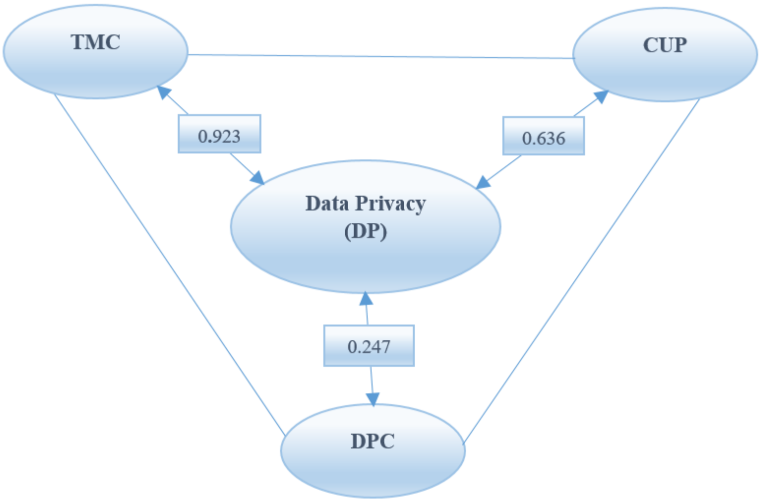
		ODP	TMC	CUP	DPC
ODP	Pearson Correlation	1			
	Sig. (2-tailed)				
TMC	Pearson Correlation	.923**	1		
	Sig. (2-tailed)	.000			
CUP	Pearson Correlation	.636**	.118	1	
	Sig. (2-tailed)	.008	.191		
DPC	Pearson Correlation	.247**	.029	.224*	1
	Sig. (2-tailed)	.005	.747	.012	

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Overall Data Privacy (ODP), Top Management Commitment (TMC),
Consent for the Use of Personal data (CUP), Duration of Personal Data Collection (DPC)

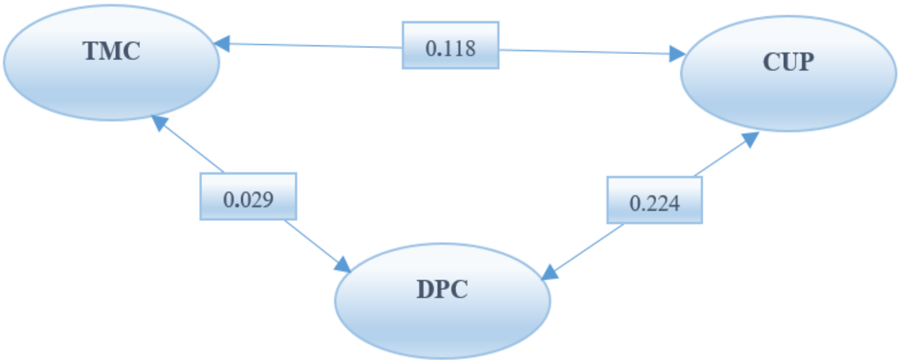
From the analysis, the correlation between the overall data privacy (ODP) and its three factors showed that the highest positive correlation was with the Top Management Commitment (TMC) at .923 and with consent for the Use of Personal data (CUP) at .636 while the lowest was Duration of Personal Data Collection (DPC) at .247.

Also, by comparing the correlation of the three data privacy factors amongst themselves, the analysis showed that there is no significant correlation between TMC and CUP as well as DPC while the CUP and DPC has a very low positive correlation at .224. The figure 22 and 23 summarizes the results.



Data Privacy (ODP), Top Management Commitment (TMC),
Consent for the Use of Personal data (CUP), Duration of Personal Data Collection (DPC)

Figure 22: Correlation between the overall data privacy and Its factors (Source: Researcher)



Data Privacy (ODP), Top Management Commitment (TMC),
Consent for the Use of Personal data (CUP), Duration of Personal Data Collection (DPC)

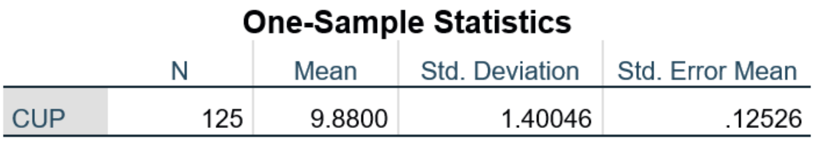
Figure 23: Correlation between the Data Privacy factors (Source: Researcher)

In conclusion, the correlation analysis gave the study a better comprehension of how much the three factors affects the data privacy score as a whole. Particularly, the analysis showed that two factors (TMC and CUP) are highly positively correlated with the overall data privacy implying that the Top management commitment (TMC) and the consent of use of Personal data (CUP) are believed to be the most important aspects that affects the Data Privacy of seafarers in their various shipping companies.

4.4.2. Main Hypothesis Testing- Hypothesis H1

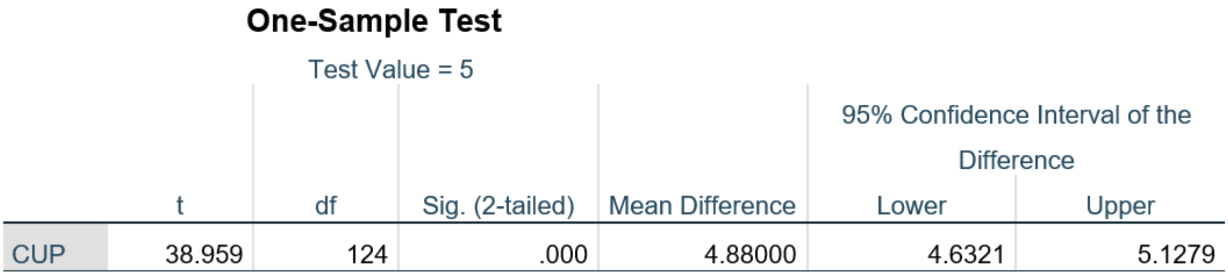
This section aimed to validate the hypothesis H1 developed in the data analysis section (section 4.1.3). To ascertain whether there is statistical evidence that the mean score of consent for use personal data (CUP) factor of the related samples are substantially different from the mean score of the respondents, a one sample t-test was conducted using SPSS. To support hypotheses, a "p-value" of less than 0.05 and a t-test absolute value greater than 1 were required as suggested by (Kahlouche, 2021). The result as shown in Table 8 and 9, showed that the requirement to support the hypotheses were met as p =.000 and t = 38.959. In addition, the result indicated the sample mean score of 9.88 which is higher than the test value of 5, this implies that all the respondents strongly agree that consent is not sought and obtained before the use of their personal data. Thus, supporting the main hypothesis H1- *“Big data analytics in the shipping companies negatively impacts on the privacy rights of seafarers.”*

Table 8: Result of One-sample test (Source: Researcher)



One-Sample Statistics

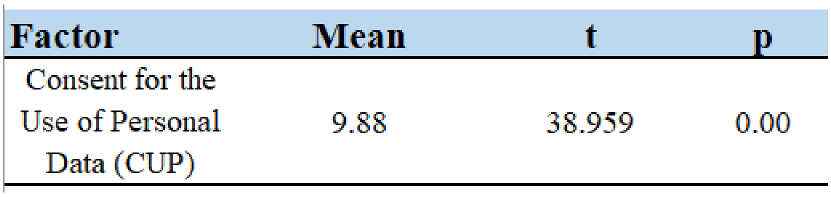
	N	Mean	Std. Deviation	Std. Error Mean
CUP	125	9.8800	1.40046	.12526



One-Sample Test

	Test Value = 5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
CUP	38.959	124	.000	4.88000	4.6321	5.1279

Table 9: Summary of test result (Source: Researcher)



Factor	Mean	t	p
Consent for the Use of Personal Data (CUP)	9.88	38.959	0.00

4.5. Discussion of Findings

The research evaluated the impact of big data analytics on the privacy rights of seafarers and how security and privacy infringement issues can be linked to the collection, storing, analysis, processing, destruction, reuse and sharing of data. The research further analysed the relation that exists between big data characteristics and privacy infringement from the standpoints of data collection, storage, processing, sharing and accessibility. The study focused on ships engaged in both international and domestic voyages in order to achieve this aim.

The results of the research findings are summarized as follows:

1. **Results from the Exploratory Research:**

An exploratory survey was conducted for this research and this survey includes 4 interviews with various maritime stakeholders i.e. ship-owner, a data controller and two seafarers with over 10 years working on board vessels sailing on both International and domestic waters. The interview was conducted in order to ascertain how seafarer's data are stored and processed as well as whether the consent of seafarers is sought and obtained before their personal data are used. In the course of these interviews, 3 categories were identified with 10 related themes to the topic of this research. These categories include: organizational data management, impact of big data analytics on the privacy rights and consent for the use of data along data continuum. Also, one hypothesis was formulated (tagged H1), this hypothesis was formulated to address the research objectives.

Hypothesis (H1): *"Big data analytics in the shipping companies negatively impacts on the privacy rights of seafarers"*

2. Result of the Survey Questionnaire:

The survey questionnaire was developed by the researcher in order to assess the impact of big data analytics on the privacy rights of seafarers and to determine whether the consent of these seafarers are sought and obtained before their personal data are used. In the quantitative data analysis, the survey questionnaire was tested for validity and reliability, then reconstructed using Exploratory Factor Analysis of the SPSS tool (SPSS 26).

This analysis resulted in the extraction of three (3) factors: top management commitment to data safety, consent for the use of personal data and duration of personal data. Furthermore, a correlation analysis was done and the result showed that the top management commitment to data safety and consent for the use of personal data are highly and positively correlated with the overall data privacy rights infringement, that is to say the failure of the top management to commit to the data safety of seafarers and no seeking of consent for the use of personal data are key element encouraging infringement of privacy rights. Therefore, this can be assumed to represent reliable predictors and indicators thereby supporting the main hypothesis.

Five research objectives were part of the study, as was already indicated. Answering research question one addresses the first research objective, while confirming a research hypothesis addresses the second research objective, the third and fourth Objectives were achieved through the review of existing literatures on the impact of big data analytics on privacy rights and the fifth objective was addressed through review of existing literature on privacy by design as a solution for privacy issues.

Research Objective One

"Assessing top management commitment to seafarer's data safety as a factor that enable data privacy rights infringement"

Research question (Q1):

Whether the top management of shipping companies prioritize profit making over data safety?

One of the premise of this research is that the adoption of data safety measure is a strategic decision of an organization (Brous et al., 2017), hence, it requires top management commitment and leadership to create a data safety culture in the organization which will promote the employees trust in the ability of the organization to ensure safety of their personal information. The essence of this question was to investigate the perception of seafarers on the commitment of the top management towards privacy rights protection and also the perception of seafarers on the existing data management policies in place aimed at protecting seafarer's privacy rights along the entire data continuum – from data sensing to its extraction, storage, aggregation, transcription, retention, analysis and retention

In order to answer this research question one, the study employed both qualitative and quantitative data analysis and a cursory look at the result of the analysis shows that the interviews conducted by the researcher yielded a classification of the factors related to the impact of big data analytics on seafarer's privacy rights into three main categories. While the category of organizational data management included three themes, impact of big data analytics on the privacy rights included eight themes and consent for the use of data along data continuum included five themes (See section 4.1.2).

Flowing from the details of this interview and the analysis of the themes, a questionnaire detailing questions related to the analyzed themes were developed, the questionnaire detailed all core questions related to the research topic and questions alike. The questionnaire was tested for validity and reliability, then reconstructed using Exploratory Factor Analysis of the SPSS tool (SPSS 26). This analysis resulted in the extraction of three (3) factors: top management commitment to data safety, consent for the use of personal data and duration of personal data.

The survey questionnaire was viewed by 163 respondents, for which 37 respondents partially completed the survey and a total number of 125 respondents fully completed the survey. For accuracy of the analysis, the researcher filtered the result using time spent in answering the survey as a yardstick to determine accuracy of the response, hence, surveys completed under 3 minutes were filtered out. After these invalid questionnaires were filtered out, a total number of 125 responses was

assumed as valid and therefore analysed. The valid data was then organized and prepared for analysis and coded into SPSS 26.0 software.
Firstly, a Bartlett test of sphericity and Kaiser Mayer Olkin (KMO) tests were performed to determine the data sustainability to conduct an exploratory factor analysis (EFA) and the obtained values for the test were satisfactorily as shown in Table 4. Then, the exploratory factor analysis was performed which resulted in the extraction of three factors (as shown in Table 6) on which a validity and reliability test was carried out to estimate the internal consistency of items combined in the factors using Cronbach Alpha value, the result shown in Table 5 and 6 showed that the scale is reliable as it satisfied the rule of thumb that says a scale is reliable if the Cronbach alpha value is greater than 0.70 (Hair et al., 2013) and the factor loading for every item must exceed 0.5 (Awang ,2014).
Furthermore, a correlation analysis was carried out to assess the relationship amongst the three factors related to data privacy right infringement in the context of big data analytics and to identify which factors contribute most to the privacy rights infringement, the result of the analysis shown in Table 7 interpreted in terms of the value of Pearson correlation coefficient such that the closer the value is to ±1, the higher the correlation, the closer it is to 0, the lower the correlation of the factors (Field, 2005) showed that the top management commitment to data safety (TMC) is highly and positively correlated to the overall data privacy rights infringement factors which implies that the failure of the top management to commit to the data safety of seafarers or the prioritization of profit making over data safety by the top management of shipping companies are key element encouraging infringement of seafarer’s privacy rights.

In addition, the feedback from the questionnaire generally shows that majority of the respondents representing 72.79% are of the perception that the top management of their companies somewhat prioritize profit making over their data safety. Similarly, majority of the respondents representing 74.27% agree that they do not get clear information from the top management on how their personal data are managed.
Also, 83 number of respondents representing 66.18% believes that there are no adequate data privacy policy in their company to protect them along the entire data

continuum, 66.91% of seafarers agree that there is no insurance policy in place in their company to protect them against privacy breach while 64.71% representing majority agree that they are not compensated for personal data privacy infringement. It is important to stress that this position was confirmed from the interviews conducted.

It is beyond doubt that the main corporate governance issue challenging organizations is that of accountability. Corporate governance obviously relates to the exercise of power over and responsibility for the organization and its employees (Blackwell Publishers, 2000). A corporate governance is expected at its barest minimum to provide an assurance of the organization's capacity to support accountability to its employees (Brooks, 1997). Accountability in this context is one of the 4 pillars of corporate governance which also include responsibility, transparency and fairness (King Report, 2001). However, one of the main challenges facing the convergence of corporate governance is that of information or data security. It has become an onerous task to convince the top management of organizations to be responsible and accountable and therefore prioritize the safety of employee's data and information over profit making and this is the same in the shipping industry where a majority of the respondents representing 72.79% agreed that their top management prioritize profit making over their data safety.

This finding is reinforced in the decade old information security breaches survey conducted by PriceWaterhouseCoopers (PWC) in 2002 wherein they found that *''... the root cause is that data security is treated as an overhead rather than an investment''* (PriceWaterhouseCoopers, 2002). Similarly, data security can be argued to begin and end with good corporate policies (Whitman & Mattord, 2003). Therefore, it is pertinent that the top management of organizations makes policies that protects the entities within an organization and further ensure that their privacy rights are protected. However, from the responses of the survey questionnaire it is crystal clear that an overwhelming majority representing 66.91% agree that their company does not have an adequate insurance policy to protect them along the entire data continuum. Furthermore, majority of the respondents representing 64.71% agree that there is no compensation mechanism in place to compensate them for privacy breaches. This shows that the organizations have failed two pillars of corporate governance (accountability and responsibility) and for fairness and transparency, it is shown from the result of the survey that 55.9% of respondents agree that they are not aware how their personal data is managed and processed. These findings regarding the failure of the top management to commit to the data safety of seafarers as a key element encouraging infringement of privacy rights is aligned and also replicated in previous studies.

Research Objective Two

"Ascertaining whether the consent of seafarers is sought and obtained before their data are used by their companies'"

Hypothesis (H1): *"Big data analysis in the shipping companies negatively impacts on the privacy rights of seafarers.*"

The above research objective of this study was mainly addressed through statistical analysis of the quantitative data although a part of it was also achieved through qualitative analysis. The researcher conducted interviews to ascertain whether the consent of seafarers is sought and obtained before their data are used by their companies. The analysis of the result of the interviews showed that majority of the shipping companies fail to seek the consent of seafarers before their personal data are used.

In order to properly address this research objective, the researcher formulated the aforementioned hypothesis (H1) for statistical analysis. To ascertain whether there is statistical evidence that the mean score of consent for use personal data (CUP) factor of the related samples are substantially different from the mean score of the respondents, a one sample t-test was conducted using SPSS. To support hypotheses, a "p-value" of less than 0.05 and a t-test absolute value greater than 1 were required as suggested by (Kahlouche, 2021). The result as shown in Table 7, showed that the requirement to support the hypotheses were met as p =.000 and t = 38.959. In addition, the result indicated the sample mean score of 9.88 which is higher than the test value of 5, this implies that all the respondents strongly agree that consent is not sought and obtained before the use of their personal data. Thus, supporting the main hypothesis H1- *“Big data analytics in the shipping companies negatively impacts on the privacy rights of seafarers”*

From the above analysis and the result of the interview, it can be shown that the results validates the formulated hypothesis and confirms that the consent of seafarers is not sought and obtained before their personal data are used. This finding tallies with the findings in existing literatures. These findings are embedded in recent high profile cases where the behaviours of some big companies were called into question regarding their use of datasets. Circa 2018, a revelation by Christopher Wylie revealed that British consulting firm Cambridge Analytica in collaboration with Facebook harvested the personal data of about 80 million (eighty million) of the app users without their consent (Isaac & Singer, 2019). Similarly, according to Singer & Conger (2019), YouTube which is a platform owned by Google were fined about $US170 million for the use of personal information of its users that are minors without parental consent for advertisements. Finally, in 2019, Bounty were fined about £400,000 for extracting users personal data and sharing them with a third parties for marketing purposes (Postelnicu, 2019).

The above cases have brought to the fore a number of questions regarding privacy protection and has shown that in most cases express consent of data subjects is not always sought and obtained before their personal data are used by the companies thereby raising serious privacy issues and this is the obvious case in the shipping industry as shown by the result of analysis.

Research Objective Three

“Highlighting the costs, benefits as well as the externalities which are associated with the use of big data by shipping companies”

“This is addressed in the literature review section 2.5.1”

Research Objective Four

“Investigating how the inherent characteristics of big data are linked to security and privacy infringements.”

“This is addressed in the literature review section 2.6”

Research Objective Five

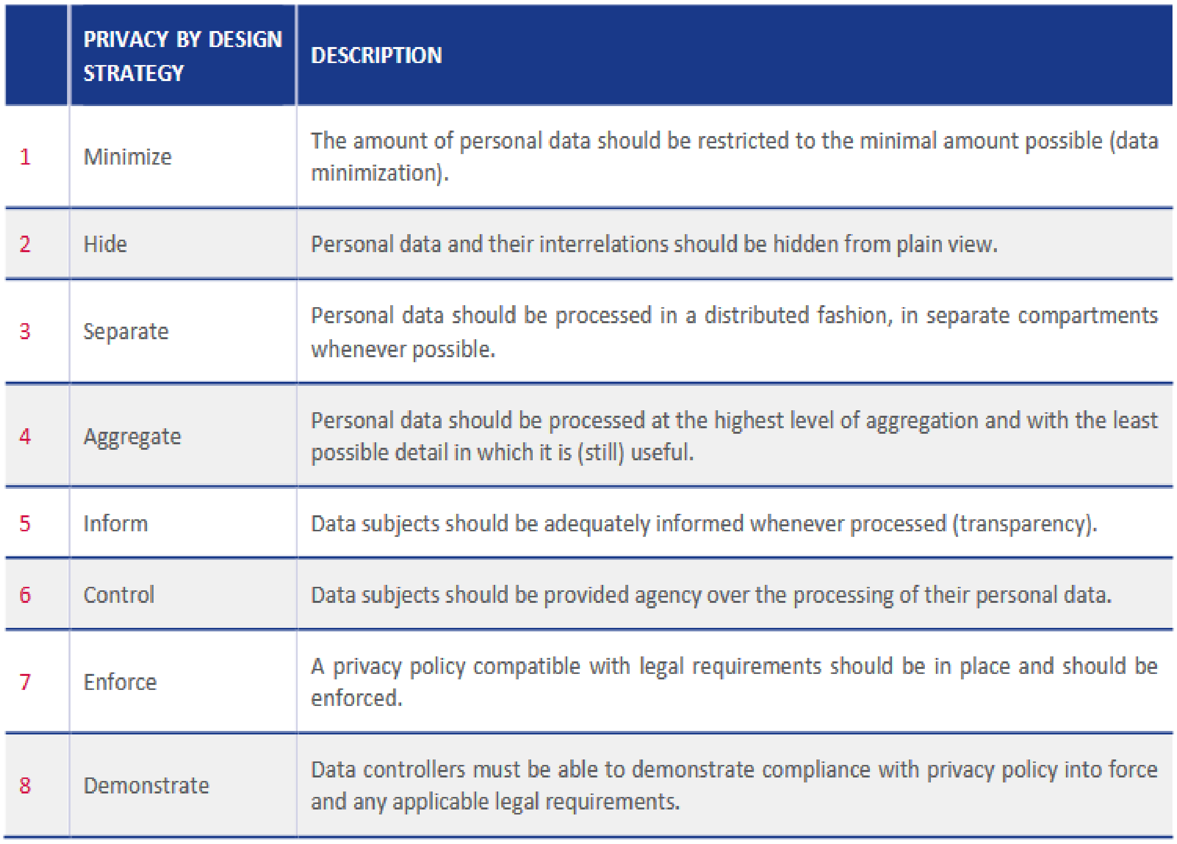
“Recommending the adoption Privacy by Design strategy as a solution data privacy issues in shipping companies”

“This is addressed through review of literature on Privacy by Design Strategies”

Big data analytics undoubtedly involves enormous data gathering and the subsequent processing of this personal data has obviously raised very serious privacy issues, particularly in relation to widespread electronic surveillance, profiling, and publication of sensitive information. It is therefore very crucial to set boundaries for big data processing and incorporate the proper data safety measures at the very centre of the analytics value chain in order to enable all the benefits of analytics without breaching people's privacy. In order to achieve this, this study will therefore examine why it is necessary to change the conversation from "big data vs privacy" to "big data with privacy," accepting the privacy and data protection principles as a fundamental value of big data. Understanding the concept of privacy by design is essential in order to identify the privacy requirements earlier in the big data analytics value chain and also in the subsequent implementation of the required technical and organizational safeguards in place.

According to Ann Cavoukian (2013), the idea of privacy by design involves integrating privacy safeguards and privacy-enhancing technologies (PETs) directly into the architecture of information technologies and systems. Privacy by design cannot be limited to the use of PETs or to a list of merely broad rules. In actuality, it is a process that incorporates numerous organizational and technological elements that uphold privacy and data protection standards. The European Union Agency for Network and Information Security (ENISA) in its 2014 study, investigated the idea of privacy by design using pertinent research in the area and provided eight privacy by design techniques, both data- and process-oriented, with the intention of retaining certain privacy goals. A quick summary of the suggested design strategies is given in Table 10

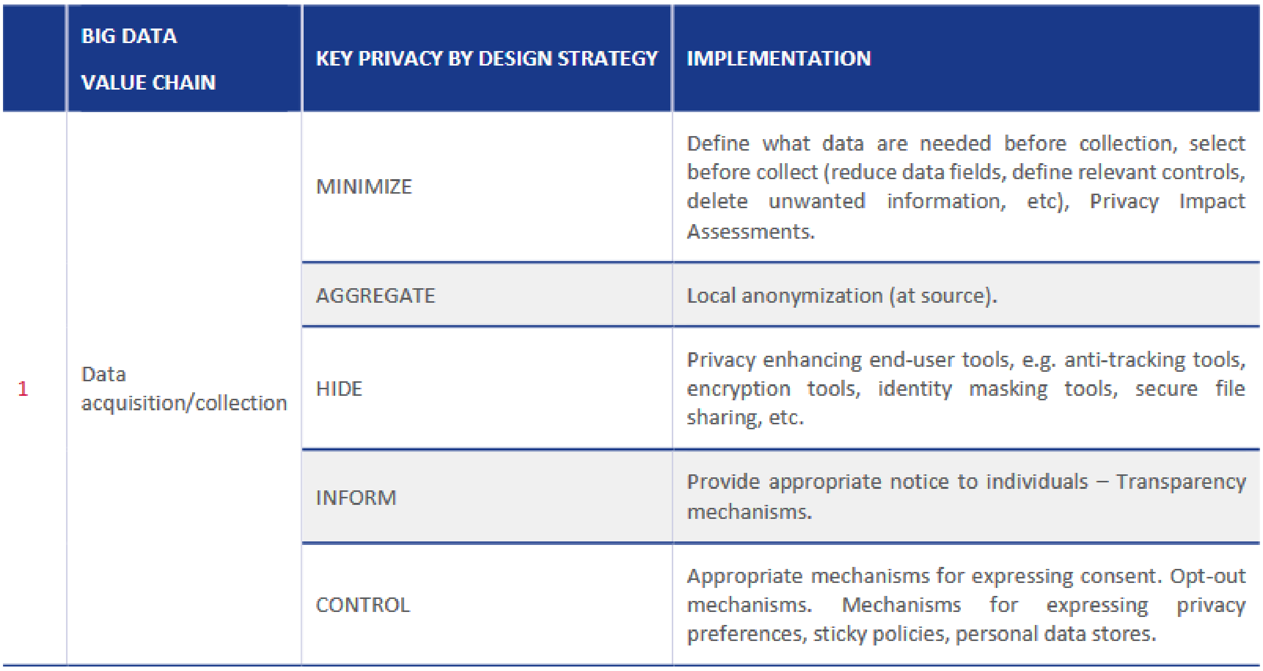
Table 10: Summary of design strategies (Cavoukian, 2013)



	PRIVACY BY DESIGN STRATEGY	DESCRIPTION
1	Minimize	The amount of personal data should be restricted to the minimal amount possible (data minimization).
2	Hide	Personal data and their interrelations should be hidden from plain view.
3	Separate	Personal data should be processed in a distributed fashion, in separate compartments whenever possible.
4	Aggregate	Personal data should be processed at the highest level of aggregation and with the least possible detail in which it is (still) useful.
5	Inform	Data subjects should be adequately informed whenever processed (transparency).
6	Control	Data subjects should be provided agency over the processing of their personal data.
7	Enforce	A privacy policy compatible with legal requirements should be in place and should be enforced.
8	Demonstrate	Data controllers must be able to demonstrate compliance with privacy policy into force and any applicable legal requirements.

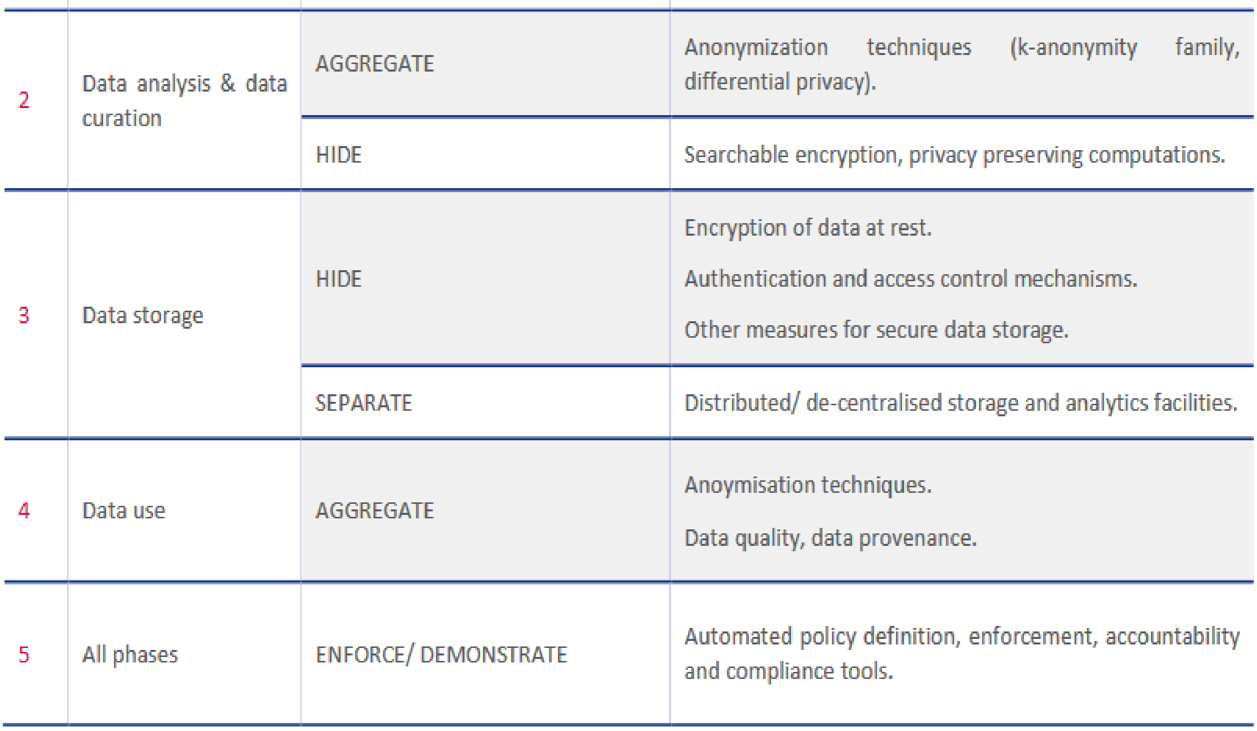
In the value chain of big data analytics, it is crucial to consider the goal of each stage of the big data analytics value chain as well as the perspectives of all parties involved (data subjects, various data controllers and processors, third parties). The exact privacy requirements as well as the necessary implementation steps for each phase may be different for each phase. Therefore, it is pertinent to note that in addition to the requirements of each specific phase, it is crucial to adopt a cogent strategy for protecting big data privacy, taking into account the entire lifecycle of the analytics. This means that the issue is not simply one technology or another, but rather the effective synthesis of numerous technologies to meet all of the needs of the various data processing nodes. Therefore, Table 11 below shows the privacy by design strategies in the big data value chain.

Table 11: Strategies for Data Acquisition



	BIG DATA VALUE CHAIN	KEY PRIVACY BY DESIGN STRATEGY	IMPLEMENTATION
1	Data acquisition/collection	MINIMIZE	Define what data are needed before collection, select before collect (reduce data fields, define relevant controls, delete unwanted information, etc), Privacy Impact Assessments.
		AGGREGATE	Local anonymization (at source).
		HIDE	Privacy enhancing end-user tools, e.g. anti-tracking tools, encryption tools, identity masking tools, secure file sharing, etc.
		INFORM	Provide appropriate notice to individuals – Transparency mechanisms.
		CONTROL	Appropriate mechanisms for expressing consent. Opt-out mechanisms. Mechanisms for expressing privacy preferences, sticky policies, personal data stores.

Table 12: Strategies for Data Analysis and Curation, Storage and Use (Cavoukian, 2013).



2	Data analysis & data curation	AGGREGATE	Anonymization techniques (k-anonymity family, differential privacy).
		HIDE	Searchable encryption, privacy preserving computations.
3	Data storage	HIDE	Encryption of data at rest. Authentication and access control mechanisms. Other measures for secure data storage.
		SEPARATE	Distributed/ de-centralised storage and analytics facilities.
4	Data use	AGGREGATE	Anoymisation techniques. Data quality, data provenance.
5	All phases	ENFORCE/ DEMONSTRATE	Automated policy definition, enforcement, accountability and compliance tools.

Building on the privacy by design strategies discussed above, this study will now take a step forward to present an overview of some identified technologies that can be implemented and/or developed by shipping companies in the specific context of big data. it is also important to stress that some of these technologies are readily available and used by some shipping companies in the traditional data processing (Carr et al, 2004). However, it all boils down to the scale for which it is in use, therefore, this study seeks to discuss the peculiarities and adoption of specific innovations in big analytics, taking into account the volume, velocity, diversity, and validity which is the hallmark of big data as discussed in previous chapters.
Firstly, this study will discuss the concept of anonymization which is obviously in use currently in most companies as a conventional approach in data analytics but facing huge challenge in this era of big data, this study will proceed to examine the new developments in encryption which ensures privacy preserving data analytics with no disclosure of private data of seafarers. The study will also discuss the concept of privacy by security which is an overall security framework for data safety particularly as it relates to access control policies and its effective enforcement. The study will conclude with a discussion on some transparency and control mechanisms such as consent and notice and other user privacy preferences.

- **Anonymization:** This basically modifying a data or removal of identifying information from a data so that the original source will remain unknown neither can they be re-identified. In anonymization, the privacy of seafarers can be compromised by either an identity disclosure or an attribute disclosure. In order to curb these challenges, there is a need to develop privacy measures like k-anonymity and its extensions (p-sensitive k-anonymity, l-diversity, t-closeness, (n,t)-closeness) and ε-differential privacy models (crowd-blending privacy or BlowFish) (D'Acquisto et al., 2015). The k-anonymity model classifies data attributes into several non-disjoint types such as identifiers, quasi-identifiers and confidential attributes while the differential privacy model seeks to ultimately limit the main impact of each individual subject's contribution to the analysis outcome by not anonymizing the data set but rather anonymizing answers to interactive queries sent to the database (Domingo-Ferrer et al.,2016).

- **Encryption:** This is basically a security technique used in transforming data so that only authorized persons can use or read the data set. This can be essential for big data so far it is performed with the appropriate encryption algorithms and key sizes which are properly secured. Usually in big data analytics, searches and other computations are allowed for stored data therefore the question often arise, how can this be done without decrypting the data which ultimately contradicts the whole notion of encryption?

  An encryption technique which can be used to solve the above challenge and also protect seafarer’s privacy is known as attribute-based encryption (ABE), this technique makes it possible to share data among various user groups while protecting users' privacy. In example, ABE integrates access control with public-key cryptography in a way that the cipher text and secret key used for encryption rely on specific characteristics (such as the person's location, occupation, or habit). In this method, the cipher text can only be decrypted if the supplied set of attributes coincides with the cipher text’s attributes. A simple example of ABE is Identity-based encryption (IBE) where both the cipher text and secret key are linked to identities (such as the user's email address), and decryption is only possible when the identities are equal.

  Another technique is the functional encryption which is an improvement on ABE where a user with particular characteristics will have access to a key that will allow them to perform a specific function on the encrypted dataset. This is particularly useful when a set of cipher text is publicly visible but can only be partially decrypted and used for processing that is specified in the secret key. An example of this will be where an encrypted seafarer’s data is only used to reveal merely the overall number of seafarers affected by a certain disease (e.g. covid 19) without needing to decrypt their personal information. As a result, functional encryption can be quite intriguing for big data cloud storage and a big data security solution.

Encrypted search can also serve as a powerful tool in big data analytics for shipping companies as this will enable a full search functionality without the disclosure of the user's personal information. This will be particularly important for the shipping companies when dealing with query-answer systems. The company can conveniently extract relevant information from the dataset without accessing the original data.

There are other encryption techniques such as the property preserving encryption, Symmetric Searchable Encryption, Public Key Searchable Encryption, structured encryption, and, Oblivious RAM, Secure multi-party computation etc.

- **Consent, ownership and control:** In big data, user control is a key objective that may be attained by using a multichannel strategy. One conceivable solution is consent (obviously the most important one). A good example of this technique is "tagging" each unit of personal data with "metadata" outlining the necessary conditions for data protection.

  Given that consent is a fundamental tool for data protection, therefore, consent in big data should go beyond the current models and offer more automation, both in the collecting of consent and its withdrawal. It could be interesting to investigate the use of software agents to obtain consent from users based on the characteristics of particular applications. Other useable and practical user affirmative activities that could constitute consent may include the use of gesture, spatial patterns, behavioural patterns, and motions in granting permission. taking into consideration the sensors and smart devices in big data.

  Another good approach is known as user control where seafarers and controllers will be given an opportunity to indicate their privacy policies and requirements before the data processing stage. This strategy is also referred to as "privacy preferences." Sticky policies can offer a way to link privacy choices to particular data sets and subsequently influence decisions on how data is processed. Seafarers can convey to each data controller their preferences and inclinations regarding the processing of personal data, particularly through the formal statement of acceptable privacy choices. For instance, seafarers can choose the permitted recipients, deletion period, and permissible reasons for the processing of personal data. The commitments in this statement can then serve as a privacy contract between the seafarers and the data controllers which details the agreed situations for data collection and processing.

CHAPTER FIVE: CONCLUSION, LIMITATION AND RECOMMENDATION FOR FURTHER RESEARCH

5.1. Conclusion

This research evaluated the impact of big data analytics on the privacy rights of seafarers and how security and privacy infringement issues can be linked to the collection, storing, analysis, processing, destruction, reuse and sharing of data. The research further analysed the relation that exists between big data characteristics and privacy infringement from the standpoints of data collection, storage, processing, sharing and accessibility. The study focused on ships engaged in both international and domestic voyages. In order to achieve this aim, five objectives were outlined as follows:

a. Assessing top management commitment to seafarer's data safety as a factor that enable data privacy rights infringement.

b. Ascertaining whether the consent of seafarers is sought and obtained before their data are used by their companies.

c. Highlighting the costs, benefits as well as the externalities which are associated with the use of big data by shipping companies.

d. Investigating how the inherent characteristics of big data are linked to security and privacy infringements.

e. Recommending the adoption Privacy by Design strategy as a solution data privacy issues in shipping companies.

The research employed an exploratory mixed method design. This consisted of semi-structured interviews conducted to investigate the impact of big data analytics on the privacy rights of seafarers as well as if the consent of seafarers is sought and obtained before their data are used by their companies. The choice of participants for these interviews consisted of a ship-owner, a data controller and two seafarers working on board ships engaged in both international and domestic voyages. Responses from the interviews assisted in the design of the questionnaire.

The aim of the survey questionnaire was to assess the factors that enable data privacy rights infringement in shipping companies and to ascertain whether the consent of seafarers is sought and obtained before their data are used by their companies.

The first and second research objectives were attained through the analysis of the content of the interviews, the conduct of EFA and correlation analysis of the qualitative data as well hypothesis testing. While, the third and fourth objectives were achieved through the review of existing literatures on the impact of big data analytics on privacy rights and the fifth objective was addressed through review of existing literature on privacy by design strategy as a solution for data privacy issues.

The study found that three factors enables data privacy rights infringement which are: Top management commitment to data safety (TMC), Consent for the Use of Personal Data (CPU) and the Duration of Personal Data Collection (DPC). It was found that top management of shipping companies prioritizes profit making over data safety thereby encouraging privacy rights infringement of seafarers. It was also found that seafarers consent is not sought for before the collection and use of their personal data by their companies which violates the provisions of the general data protection regulation (GDPR) and as such contributes to the infringement of seafarer’s privacy rights.

5.2. Limitation

There is a paucity of literature in this field of research, the researcher could find little or no literature on the impact of big data on seafarer’s privacy rights. This could be attributed to the fact that this is a new field of research and the concept of big data analytics is undeveloped in the shipping industry. Secondly, there was difficulty in obtaining real-time data from shipping companies for analytics. If real-time data was obtained, this study would have been able to conduct analysis and show how private data of seafarers are revealed without their consent most especially identification data, location data and data from a ship’s VDR.

5.3. Recommendation for Further research

The role of data in the maritime insurance business is expected to undergo a fundamental transformation as a result of developments in big data analytics in the shipping industry. Big data analytics is expected to evolve the role of insurance from the current concept of" understand and protect" to a concept of 'Predict and prevent'. Although this evolution in the maritime insurance industry holds considerable promise and economic advantages, however, it raises challenges such as data protection and privacy, insurance individualization, and competition.

Maritime insurance stakeholders like Individuals, organizations and policymakers will obviously be confronted with intricate trade-offs when striking a balance between the advantages of sharing personal data and the risks involved in the use of these personal data. The difficulty in striking balance between these trade-offs is as a result of the fact that the trade-offs are often subjective, context-specific as well as non-measurable.

It is recommended that a further research is conducted to contribute to an informed and fact-based regulatory debate on these trade-offs. In order to achieve this, a discussion should be made on the societal and economic advantages of using big data analytics in maritime insurance as well as the major issues that have come up in the public and regulatory discourse. The study is recommended to identify the main trade-offs resulting from the increased usage of personal data in maritime insurance while evaluating the consequences of the different policy choices. The recommended research will employ a mixed method of quantitative and qualitative analyses of the bibliometric data and content of publications while following a 4 step mixed-methods approach as shown in the diagram below.

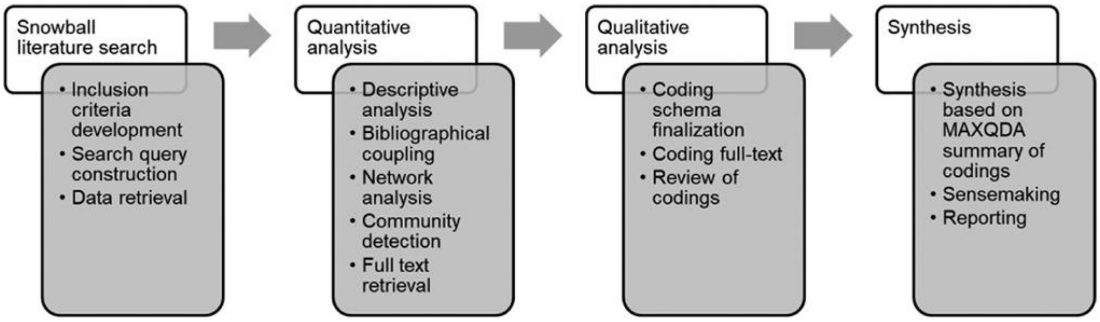


Figure 24: Recommended 4 step mixed- methods approach (Suominen & Hajikhani, 2021)

REFERENCES

Ajibade, S. S., & Adediran, A. (2016). An overview of big data visualization techniques in data mining. *International Journal of Computer Science and Information Technology Research*, *4*(3), 105-113.

Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. *Surveillance & Society*, *12*(2), 185-196.

Atas, A. H., & Çelik, B. (2019). Smartphone use of university students: Patterns, purposes, and situations. *Malaysian Online Journal of Educational Technology*, *7*(2), 59-70.

Barth, S., Ionita, D., & Hartel, P. (2022). Understanding Online Privacy—A Systematic Review of Privacy Visualizations and Privacy by Design Guidelines. *ACM Computing Surveys (CSUR)*, *55*(3), 1-37.

Bayne, S., Evans, P., Ewins, R., Knox, J., & Lamb, J. (2020). *The manifesto for teaching online*. MIT Press.

Benato, B. C., Gomes, J. F., Telea, A. C., & Falcão, A. X. (2021). Semi-automatic data annotation guided by feature space projection. *Pattern Recognition*, *109*, 107612.

Berman, S., & Stern, H. (2011). Sensors for gesture recognition systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(3), 277-290.

Bhatia, A., & Mittal, P. (2019). Big data driven healthcare supply chain: understanding potentials and capabilities. In *Proceedings of International Conference on Advancements in Computing & Management (ICACM).*

Burmeister, H. C., Bruhn, W., Rødseth, Ø. J., & Porathe, T. (2014). Autonomous unmanned merchant vessel and its contribution towards the e-Navigation implementation: The MUNIN perspective. *International Journal of e-Navigation and Maritime Economy*, *1*, 1-13.

Borgi, T., Zoghlami, N., & Abed, M. (2017). Big data for transport and logistics: A review. In *2017 International Conference on Advanced Systems and Electric Technologies (IC_ASET)* (pp. 44-49). IEEE.

Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: a survey. *Future generation computer systems*, *56*, 684-700.

Boyd and Crawford (2012) " Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon," Information, Communication & Society 15, no. 5 (2012): 662-679.

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing*, *93*(1), 79-95.

Brous, P., Janssen, M., Schraven, D., Spiegeler, J., & Duzgun, B. C. (2017, April). Factors Influencing Adoption of IoT for Data-driven Decision Making in Asset Management Organizations. In *IoTBDS* (pp. 70-79).

Burrough, P. A., McDonnell, R. A., & Lloyd, C. D. (2015). *Principles of geographical information systems*. Oxford university press.

Buza, K., Nagy, G. I., & Nanopoulos, A. (2014). Storage-optimizing clustering algorithms for high-dimensional tick data. *Expert Systems with Applications*, *41*(9), 4148-4157.

Bygrave, L. A. (2017). Data protection by design and by default: deciphering the EU's legislative requirements. *Oslo Law Review*, *4*(2), 105-120.

Carr, N. G., & Carr, N. G. (2004). *Does IT matter?: information technology and the corrosion of competitive advantage*. Harvard Business Press.

Cavoukian, A. (2013). Privacy by design: leadership, methods, and results. In *European Data Protection: Coming of Age* (pp. 175-202). Springer, Dordrecht.

Churi, P., Pawar, A., & Moreno-Guerrero, A. J. (2021). A comprehensive survey on data utility and privacy: Taking Indian healthcare system as a potential case study. *Inventions*, *6*(3), 45.

Cimino, G., Ancieri, G., Horn, S., & Bryan, K. (2014). Sensor data management to achieve information superiority in maritime situational awareness. *CMRE Formal Report, NATO Unclassified*.

Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of big data: How it's changing the way we think about the world. *The Best Writing on Mathematics 2014*, 20-32.

Dabiri, S., & Heaslip, K. (2018). Transport-domain applications of widely used data sources in the smart transportation: A survey. *arXiv preprint arXiv:1803.10902*.

Data, B. (2015). Transport: Understanding and assessing options. In *International Transport Forum http://www. internationaltransportforum. org OECD/ITF*.

D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y. A., & Bourka, A. (2015). Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics. *arXiv preprint arXiv:1512.06000*.

Domingo-Ferrer, J., & Soria-Comas, J. (2016, September). Anonymization in the time of big data. In *International conference on privacy in statistical databases* (pp. 57-68). Springer, Cham.

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. In *AIP conference proceedings* (Vol. 1644, No. 1, pp. 97-104). American Institute of Physics.

Du, Y., Meng, Q., Wang, S., & Kuang, H. (2019). Two-phase optimal solutions for ship speed and trim optimization over a voyage using voyage report data. *Transportation Research Part B: Methodological*, *122*, 88-114.

Dyson, G., 2013. No Time Is There - The Digital Universe and Why Things Appear To Be Speeding Up. [Online] Available at: http://longnow.org/seminars/02013/mar/19/no-time-there-digital-universe-and-why-things- appear-be-speeding/[Accessed 7 August 2014]

Einav, L., & Levin, J. (2014). Economics in the Age of Big Data. Science, 346, Article ID: 1243089.

Egger, R., & Yu, J. (2022). Data science and interdisciplinarity. *Applied Data Science in Tourism*, 35-49.

EMSA (2017). Vessel traffic monitoring in EU waters (SafeSeaNet). http://www.emsa.europa.eu/ssn-main.html

Equasis (2015). The world merchant fleet in 2015.

Ervural, B. C., & Ervural, B. (2018). Overview of cyber security in the industry 4.0 era. In *Industry 4.0: managing the digital transformation* (pp. 267-284). Springer, Cham.

European Commission (2016). Exploiting maritime Big Data, the Blue Hub. https://ec.europa.eu/eurostat/cros/system/files/exploiting_maritime_big_data_a.pdf pp8

Faiz, T. (2019, November). Multi-approaches on scrubbing data for medium-sized enterprises. In *2019 International Conference on Digitization (ICD)* (pp. 75-86). IEEE.

Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PloS one*, *15*(2), e0228987.

Fluit, A. (2011). AIS information quality report of static AIS messages:‘’AIS Information Quality Report’’ Region: HELCOM.

Futurenautics (2016) Maritime Satellite Communications & Applications 2016. Pp.23

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72, pp. 59-139). Cham, Switzerland: Springer International Publishing.

Graupner, S., Bartolini, C., Motahari, H., & Erbes, J. (2012, July). Evolving social media into productivity platforms. In *2012 Annual SRII Global Conference* (pp. 183-190). IEEE.

Gruschka, N., Mavroeidis, V., Vishi, K., & Jensen, M. (2018, December). Privacy issues and data protection in big data: a case study analysis under GDPR. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 5027-5033). IEEE.

Grejner-Brzezinska, D. A., Toth, C. K., Moore, T., Raquet, J. F., Miller, M. M., & Kealy, A. (2016). Multisensor navigation systems: A remedy for GNSS vulnerabilities?. *Proceedings of the IEEE*, *104*(6), 1339-1353.

GSMA, 2017. Connected society. Unlocking rural coverage: Enablers for commercially sustainable mobile network expansion. GSMA, London.

Gudivada, V., Apon, A., & Ding, J. (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software*, *10*(1), 1-20.

Hajli, N., Shirazi, F., Tajvidi, M., & Huda, N. (2021). Towards an understanding of privacy management architecture in big data: an experimental research. *British Journal of Management*, *32*(2), 548-565.

Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, *6*(1), 1-16.

Hassan, Q., (2014). Demystifying cloud computing. The Journal of Defense Software Engineering.

Hoofnagle, C. J., van der Sloot, B., & Borgesius, F. Z. (2019). The European Union general data protection regulation: what it is and what it means. *Information & Communications Technology Law*, *28*(1), 65-98.

Ibekwe-SanJuan, F., & Geoffrey, B. (2017). Implications of big data for knowledge organization. *Knowledge organization*, *44*(3), 187-198.

IMO. AIS Transponders. http://www.imo.org/en/OurWork/Safety/Navigation/Pages/AIS.aspx

IMO. Carriage of chemicals by ship. MARPOL Annex II- Carriage of noxious liquid substances in bulk. http://www.imo.org/en/OurWork/Environment/PollutionPrevention/ChemicalPollution/Pages/Default.aspx

International Telecommunications Union (ITU) (2011) Working Group 5B-3 Maritime pp4.

ISACA, I. (2014). A Professional Practices Framework for IS Audit/Assurance. *Rolling Meadows, IL*, *60008*, 181.

Jan, B., Farman, H., Khan, M., Imran, M., Islam, I. U., Ahmad, A., ... & Jeon, G. (2019). Deep learning in big data analytics: a comparative study. *Computers & Electrical Engineering*, *75*, 275-287.

Kahlouche, N. (2021). Organizational subcultures and safety culture in shipping: case study of Algeria.

Keeso, A. (2014). Big data and environmental sustainability: a conversation starter. *Smith School of Enterprise and the Environment. Working Paper Series*, (14-04).

Kessler, D. J., Ross, S., & Hickok, E. (2014). A Comparative Analysis of Indian Privacy Law and the Asia-Pacific Economic Cooperation Cross-Border Privacy Rules. *Nat'l L. Sch. India Rev.*, *26*, 31.

Khan, L. U., Yaqoob, I., Tran, N. H., Kazmi, S. A., Dang, T. N., & Hong, C. S. (2020). Edge-computing-enabled smart cities: A comprehensive survey. *IEEE Internet of Things Journal*, *7*(10), 10200-10232.

Khan, S., Nazir, S., García-Magariño, I., & Hussain, A. (2021). Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion. *Computers & Electrical Engineering*, *89*, 106906.

Koga, S. (2015). Major challenges and solutions for utilizing big data in the maritime industry.

Kshetri, N. (2014). Big data′ s impact on privacy, security and consumer welfare. *Telecommunications Policy*, *38*(11), 1134-1145.

Lampe, Ove Daae (2011). Interactive Visual Analysis of Process Data. http://bora.uib.no/handle/1956/5302 Pp39

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, *6*(70), 1.

Latifov, K. (2019). A critical evaluation of potential outcomes of using modern artificial intelligence and big data analysis technology in the maritime industry.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Matekenya, W., & Ncwadi, R. (2022). The impact of maritime transport financing on total trade in South Africa. *Journal of Shipping and Trade*, *7*(1), 1-17.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think* . Houghton Mifflin Harcourt.

Mendel, J. M., & Korjani, M. M. (2014). On establishing nonlinear combinations of variables from small to big data for use in later processing. *Information Sciences*, *280*, 98-110.

Monohakobi Technology Insitute, NYK Group (2016). Utilizing Big Data and the Internet of Things in Shipping http://www.mlit.go.jp/common/001127982.pdf pp.5

Mosquera, I. (2017). Privacy and confidentiality in exchange of information procedures: some uncertainties, many issues, but few solutions. *Intertax*, *45*(5).

Neff, G., & Nafus, D. (2016). *Self-tracking*. MIT Press.

Oatley, G. C. (2021). Themes in data mining, big data, and crime analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1432.

Osekowska, E., Johnson, H., & Carlsson, B. (2017). Maritime vessel traffic modeling in the context of concept drift. *Transportation research procedia*, *25*, 1457-1476.

PCAST, 2014. Big Data and Privacy: A Technological Perspective. s.l.:President's Council of Advisors onScience and Technology: Executive Office of the President.

Roy, C., Rautaray, S. S., & Pandey, M. (2018). Big Data Optimization Techniques: A Survey. *International Journal of Information Engineering & Electronic Business*, *10*(4).

Russom, P. (2013). Managing big data. *TDWI Best Practices Report, TDWI Research*, 1-40.

Rustad, M. L., & Koenig, T. H. (2019). Towards a global data privacy standard. *Fla. L. Rev.*, *71*, 365.

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, *54*(5), 758-790.

Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, *2*(5), 1-22.

Schneier, B. (2015). *Secrets and lies: digital security in a networked world*. John Wiley & Sons.

Sekkesaeter, O. (2017). *Shipping in the digital age: how feasible is the application of big data to the maritime shipping industry, and under what conditions can it be developed to become an integral part of its future?* (Doctoral dissertation, University of Geneva).

Shi, L., Shi, D., Zhang, X., Meunier, B., Zhang, H., Wang, Z., ... & Song, J. (2020). 5G Internet of radio light positioning system for indoor broadcasting service. *IEEE Transactions on Broadcasting*, *66*(2), 534-544.

Sharma, S. (2019). *Data privacy and GDPR handbook*. John Wiley & Sons.

Shehab, N., Badawy, M., & Arafat, H. (2021). Big data analytics and preprocessing. In *Machine learning and big data analytics paradigms: analysis, applications and challenges* (pp. 25-43). Springer, Cham.

Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications*, *71*, 151-166.

Silva, B. N., Diyan, M., & Han, K. (2019). Big data analytics. In *Deep Learning: Convergence to Big Data Analytics* (pp. 13-30). Springer, Singapore.

Smestad, B. B. (2015). *A study of satellite ais data and the global ship traffic through the singapore strait* (Master's thesis, NTNU).

Sowmya, R., & Suneetha, K. R. (2017, January). Data mining with big data. In *2017 11th International Conference on Intelligent Systems and Control (ISCO)* (pp. 246-250). IEEE.

Steenbruggen, J., Borzacchiello, M. T., Nijkamp, P., & Scholten, H. (2013). Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. *GeoJournal*, *78*(2), 223-243.

Suominen, A., & Hajikhani, A. (2021). Research themes in big data analytics for policymaking: Insights from a mixed-methods systematic literature review. *Policy & Internet*, *13*(4), 464-484.

Tawalbeh, L. A., Muheidat, F., Tawalbeh, M., & Quwaider, M. (2020). IoT Privacy and security: Challenges and solutions. *Applied Sciences*, *10*(12), 4102.

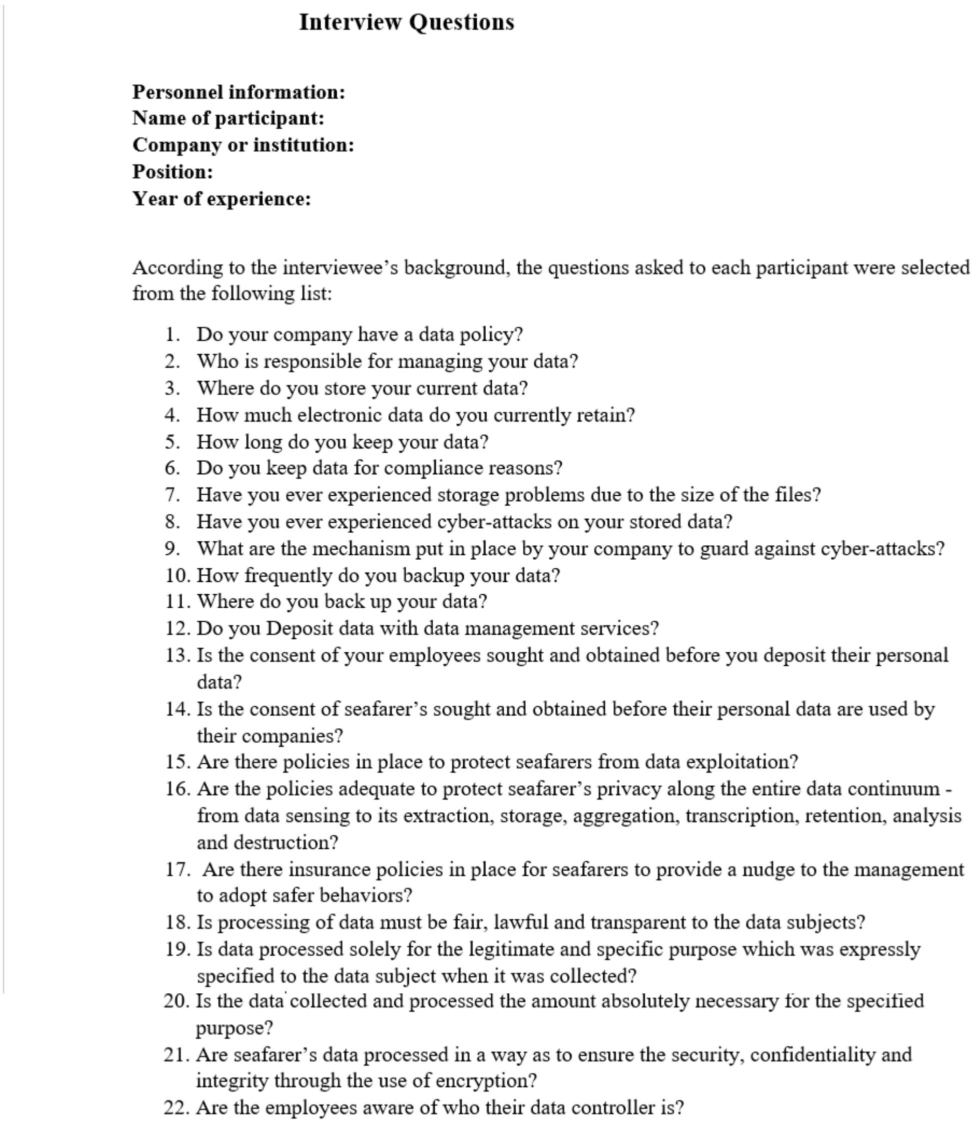
Tsai, C. W., Lai, C. F., Chao, H. C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big data*, *2*(1), 1-32.

Wijaya, W. M., & Nakamura, Y. (2013). Predicting ship behavior navigating through heavily trafficked fairways by analyzing AIS data on apache HBase. In *2013 First International Symposium on Computing and Networking* (pp. 220-226). IEEE.

Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2018). Big data analytics in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *20*(1), 383-398

Appendices

Appendix A: Personal Interview



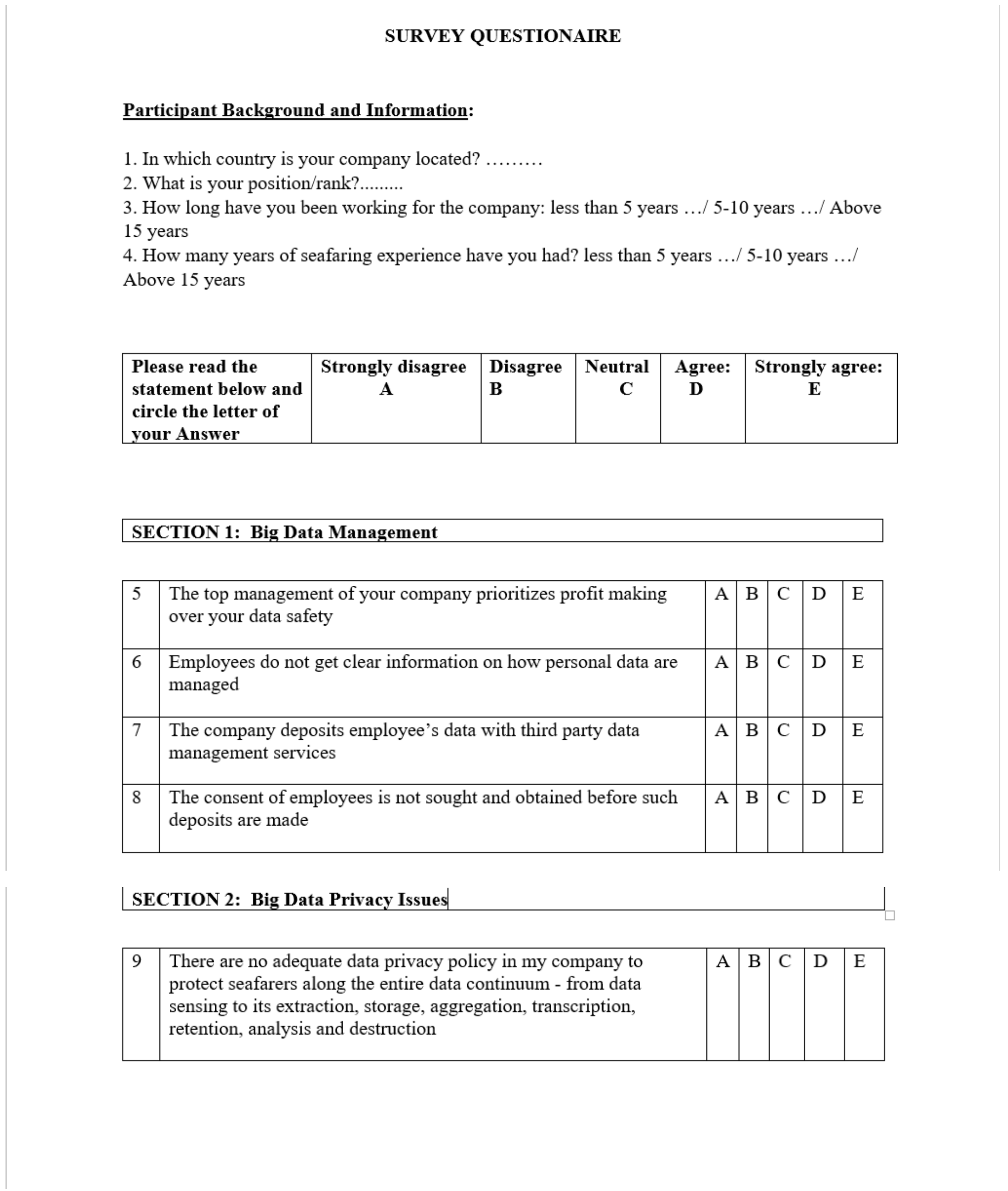
Interview Questions

Personnel information:
Name of participant:
Company or institution:
Position:
Year of experience:

According to the interviewee's background, the questions asked to each participant were selected from the following list:

1. Do your company have a data policy?
2. Who is responsible for managing your data?
3. Where do you store your current data?
4. How much electronic data do you currently retain?
5. How long do you keep your data?
6. Do you keep data for compliance reasons?
7. Have you ever experienced storage problems due to the size of the files?
8. Have you ever experienced cyber-attacks on your stored data?
9. What are the mechanism put in place by your company to guard against cyber-attacks?
10. How frequently do you backup your data?
11. Where do you back up your data?
12. Do you Deposit data with data management services?
13. Is the consent of your employees sought and obtained before you deposit their personal data?
14. Is the consent of seafarer's sought and obtained before their personal data are used by their companies?
15. Are there policies in place to protect seafarers from data exploitation?
16. Are the policies adequate to protect seafarer's privacy along the entire data continuum - from data sensing to its extraction, storage, aggregation, transcription, retention, analysis and destruction?
17. Are there insurance policies in place for seafarers to provide a nudge to the management to adopt safer behaviors?
18. Is processing of data must be fair, lawful and transparent to the data subjects?
19. Is data processed solely for the legitimate and specific purpose which was expressly specified to the data subject when it was collected?
20. Is the data collected and processed the amount absolutely necessary for the specified purpose?
21. Are seafarer's data processed in a way as to ensure the security, confidentiality and integrity through the use of encryption?
22. Are the employees aware of who their data controller is?

Appendix B: Big Data Survey Questionnaire



SURVEY QUESTIONAIRE

Participant Background and Information:

1. In which country is your company located? ………
2. What is your position/rank?.........
3. How long have you been working for the company: less than 5 years …/ 5-10 years …/ Above 15 years
4. How many years of seafaring experience have you had? less than 5 years …/ 5-10 years …/ Above 15 years

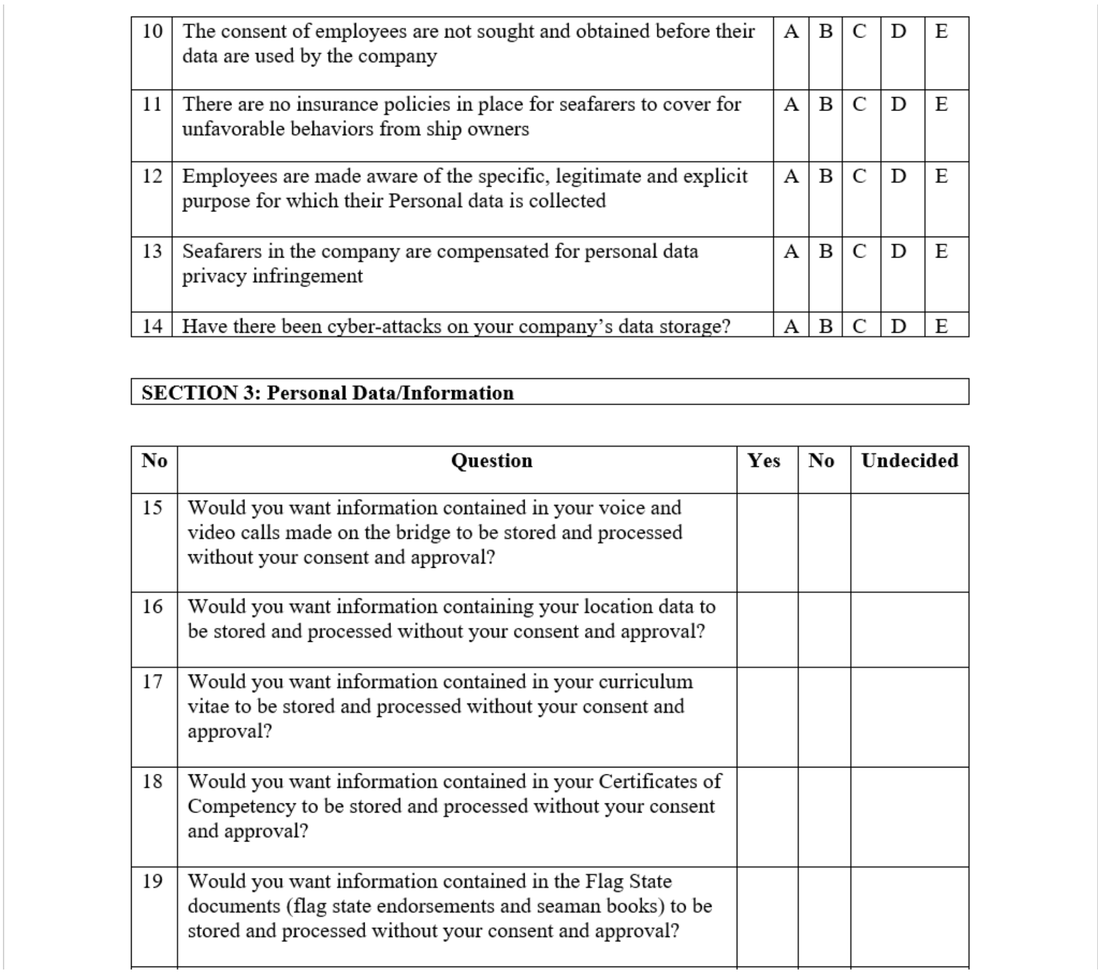
Please read the statement below and circle the letter of your Answer	Strongly disagree A	Disagree B	Neutral C	Agree: D	Strongly agree: E

SECTION 1: Big Data Management

5	The top management of your company prioritizes profit making over your data safety	A	B	C	D	E
6	Employees do not get clear information on how personal data are managed	A	B	C	D	E
7	The company deposits employee's data with third party data management services	A	B	C	D	E
8	The consent of employees is not sought and obtained before such deposits are made	A	B	C	D	E

SECTION 2: Big Data Privacy Issues

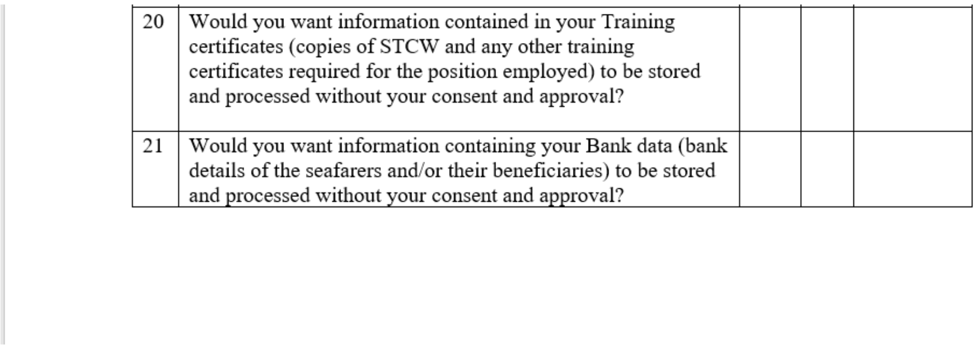
9	There are no adequate data privacy policy in my company to protect seafarers along the entire data continuum - from data sensing to its extraction, storage, aggregation, transcription, retention, analysis and destruction	A	B	C	D	E



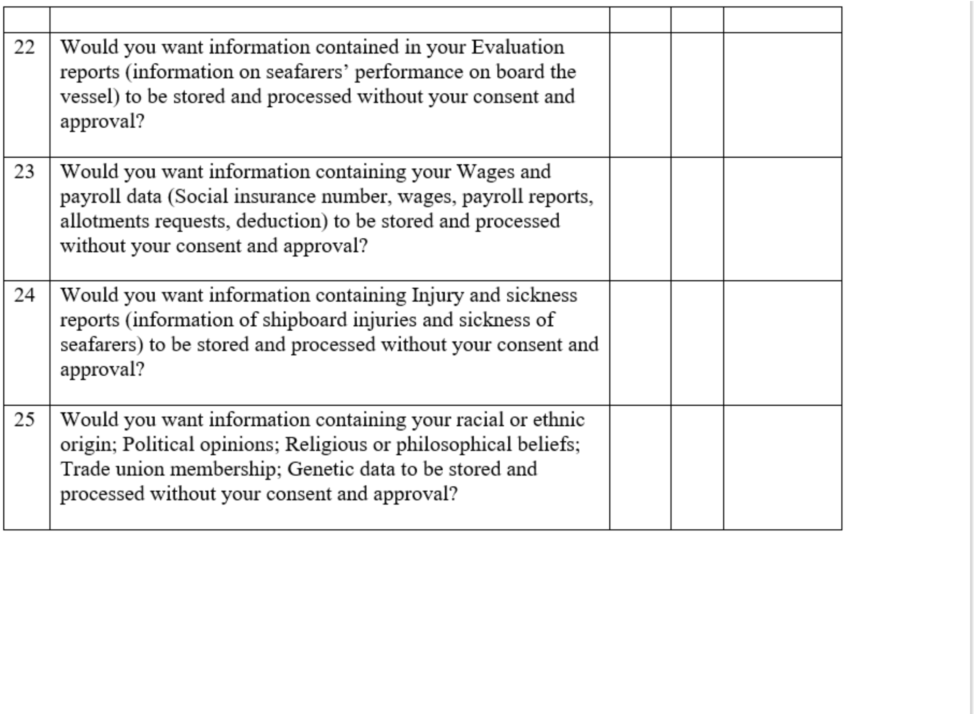
10	The consent of employees are not sought and obtained before their data are used by the company	A	B	C	D	E
11	There are no insurance policies in place for seafarers to cover for unfavorable behaviors from ship owners	A	B	C	D	E
12	Employees are made aware of the specific, legitimate and explicit purpose for which their Personal data is collected	A	B	C	D	E
13	Seafarers in the company are compensated for personal data privacy infringement	A	B	C	D	E
14	Have there been cyber-attacks on your company's data storage?	A	B	C	D	E

SECTION 3: Personal Data/Information

No	Question	Yes	No	Undecided
15	Would you want information contained in your voice and video calls made on the bridge to be stored and processed without your consent and approval?			
16	Would you want information containing your location data to be stored and processed without your consent and approval?			
17	Would you want information contained in your curriculum vitae to be stored and processed without your consent and approval?			
18	Would you want information contained in your Certificates of Competency to be stored and processed without your consent and approval?			
19	Would you want information contained in the Flag State documents (flag state endorsements and seaman books) to be stored and processed without your consent and approval?			



20	Would you want information contained in your Training certificates (copies of STCW and any other training certificates required for the position employed) to be stored and processed without your consent and approval?			
21	Would you want information containing your Bank data (bank details of the seafarers and/or their beneficiaries) to be stored and processed without your consent and approval?			



22	Would you want information contained in your Evaluation reports (information on seafarers' performance on board the vessel) to be stored and processed without your consent and approval?			
23	Would you want information containing your Wages and payroll data (Social insurance number, wages, payroll reports, allotments requests, deduction) to be stored and processed without your consent and approval?			
24	Would you want information containing Injury and sickness reports (information of shipboard injuries and sickness of seafarers) to be stored and processed without your consent and approval?			
25	Would you want information containing your racial or ethnic origin; Political opinions; Religious or philosophical beliefs; Trade union membership; Genetic data to be stored and processed without your consent and approval?			