Aalborg University Copenhagen

Master of Science (MSc) in Engineering in Operations and Management Engineering

Master Thesis

Big Data development in the tanker industry - Drivers and Challenges

Anton Perepelita   Victor Ulrich Kromann

June 1, 2021

AALBORG UNIVERSITY
DENMARK

Faculty of Engineering and Science
Aalborg University

Title:
Big Data development in the tanker industry - Drivers and Challenges

Theme:
Big Data, Maritime Industry, Semi-Systematic Literature Review, Digitalization, IoT, Shipping, Tankers

Project Period:
Feb. 2021 - Jun. 2021

Participants:
Victor Ulrich Kromann
Anton Perepelita

Supervisors:
Niels Gorm Maly Rytter

Copies: 1

Page Numbers: 71

Date of Completion:
June 1, 2021

Abstract

The maritime tanker industry faces increasing pressure to discover new ways of reducing environmental impact, while also increasing value offered. The industry faces many challenges, most of which relate to the unique asset structure of the industry. An analysis of existing literature reviews has been conducted and potential for an improved research has been found. By incorporating best practices from the previous reviews, as well as taking into account found flaws, a semi-systematic literature review is conducted and compared the literature findings with findings from interviews with industry stakeholders. Several areas of misalignment between industry's interest and literature's practices are outlined. Additionally, literature's perceived drivers and challenges and those of the industry are identified and compared. Potential solutions to the mentioned challenges are suggested, research and industry focus is proposed.

Abstract (Danish)

Den maritime tankindustri står over for stigende pres for at opdage nye måder at reducere miljøpåvirkningen på, samtidig med at øge værdien. Branchen står over for mange udfordringer, hvoraf mange vedrører brancheens unikke aktivstruktur. En analyse af eksisterende literatur er blevet lavet og vist potentiale for forbedret forskning er blevet fundet. Ved at inkorporere best-practices fra tidligere review, samtidig med at tage forbehold for funde mangler, et semi-systematisk literatur er udført og sammenlignes med resultaterne fra interviews med personer i industrien. Der identificeres flere områder med misforhold mellem litteraturens opfattede drivere og udfordringer og branchens. Fremtidig forskning og potentielle løsninger til funde udfordringer foreslås.

Acknowledgement

We would like to thank our supervisor for his guidance during this project, and express our gratitude to the interview participants for taking the time to share their insight and experience, without whom, this project would not have been possible

Table of contents

1 Introduction

With the tanker industry being responsible for 29% of the worlds sea-born trade by tonnage, and high growth over a number of years (*Statista*), one would expect the tanker industry to have developed accordingly and be utilizing new technologies to create greater value and be prepared for future challenges. However, as people with industry insight can acknowledge, the industry is characterized by a rigid culture with a tradition of acting on instinct. The increasing focus on environmental efforts and pressure from regulatory bodies and costumers, the entire Maritime Industry is searching for innovative ways to solve the new challenges. Big Data is seen as potential solution to some of the challenges that industry faces, and Big Data technologies are already being utilized in a wide range of industries. The Maritime Industry finds itself in an interesting position, where much of the needed technology is already developed and tested in other industries. The challenges lie in the Maritime Industry's unique asset situation. The assets generating the data are floating around the sea, far from the users of the data. This creates a potential lag in decision making. Within the Maritime Industry, the tanker segment faces its own unique challenges separate from the rest of the industry. The homogeneity of the cargo, and the short asset ownership periods in the tanker segment, results in the need for innovative solutions to unique challenges.

Initial research of existing literature reviews indicates that Big Data in Maritime is a new topic with increasing interest. Few literature reviews exist, each having their own strong and weak sides.

2 Problem Definition

2.1 Existing Literature Reviews

To the best of our knowledge, five literature reviews have been done on topics covering different degrees of "Big Data" and "Maritime." The existing literature reviews' methodology and findings are analyzed to argue for the relevance of this research. Fernández et al. 2018

Munim et al. 2020	Big Data and AI
Fruth and Teuteberg 2017	Digitalization gap review
Aiello, Giallanza, and Mascarella 2020	Preliminary gap analysis
Yang et al. 2019a	AIS applications
Alan 2018	Future ICTs impact on maritime transport

Table 1: Existing relevant literature reviews

Munim et al. 2020 used bibliographical data to map current research topics within Big Data and Maritime Industry, identifying the four clusters (1) Digital transformation, (2) Application of Big Data from AIS, (3) Energy efficiency, and (4) Predictive Analytics. Their number of articles is higher than ours. The more extensive number of articles can be attributed in part to a broader keyword search, in which they included "Artificial Intelligence" and "Business Intelligence" and less strict inclusion validation. Their literature review focused on identifying areas of research and intensity of focus in these areas through quantitative methods only. They found that the topic of "Big Data and Maritime" is a stand-alone research cluster with exponential growth in the number of articles. They identify technological advancement as the most extensive area of interest within the literature, and that "softer" topics, such as Drivers and Challenges, should get more attention.

Aiello, Giallanza, and Mascarella 2020 describe the gap between the current digitalization level and a "future" state. The article describes a large variety of topics, i.e., shipbuilding and infrastructure. The authors do not describe their methodology, introducing possible bias. The authors do not go into details on any one of the topics. The authors do conclude that data is rarely used in the decision-making process, regardless if it is available or not.

Fruth and Teuteberg 2017 use a structured literature review to explore the current state and what challenges the maritime logistics sector will face in the future. The topic is similar to that of Aiello, Giallanza, and Mascarella 2020, but with a higher level of robustness. Their applied literature list is small compared to the scope of the search terms. The article's conclusion states that too little research exists at the time of writing to answer their research question correctly but that the lack of research in itself calls for further research.

Yang et al. 2019a explore the history and potential of AIS through a critical literature review. The article does describe search terms or inclusion methods, reducing reproducibility. The article describes the potential strength of using the existing available AIS data. AIS data could provide insight into vessel operations, improve navigation and reduce collision risks. They also describe challenges related to fragmentation, high mobility, and cross-border operations of the industry. They conclude that AIS can be a necessary tool for achieving zero-emission goals and potentially more research.

Alan 2018 use a systematic literature review to understand the impact of Advanced Information and Communication Technologies on maritime transport. The article concludes that existing business model approaches might not be applicable with the emergence of new disruptive ICTs.

With technologies developing fast and other literature reviews having their own flaws, it has been decided to conduct a research on the current Big Data applications, its Drivers and Challenges. A new research therefore should learn from the best practices of the former ones, whilst aim at improving the mentioned flaws.

3 Problem Statement

The maritime industry faces increasingly higher demands from lawmakers and customers regarding transparency, efficiency, and environmental impact. In addition to that, actors in the industry aim to decrease their operational costs and utilize unrealized potential, resulting from smarter decisions based on precise, up-to-date data. Therefore, the actors are constantly looking towards digitalization to find the answers to these challenges. Primarily the domain known as Big Data is seeing much interest from practitioners. It is a relatively new phenomenon that recently penetrated many industries, especially the Maritime Industry. Within the Maritime Industry, the tanker segment faces unique challenges related to the segments unique asset strategy. In existing literature reviews, a missed potential has been identified to explore these challenges.

This paper explores Big Data applications and challenges in the maritime tanker industry through a semi-systematic literature review, interviews with key actors, and analysis of current technology state. After the analysis, "challenges" and "drivers" will be outlined. "Challenges" will summarize the flaws found in the available literature about Big Data phenomena as well as common problems faced by the actors working with it. "Drivers" will list factors that motivate the actors to implement Big Data techniques, such as industry regulations, cost savings, potential competitive advantage. The research questions are the following:

3.1 Research Questions

1. What is the current level of Big Data utilization and its literature representation in the maritime tanker industry?
2. What drives the actors towards improvement realization through Big Data utilization?
3. What are the improvement areas complicating Big Data utilization and how can they be overcome?

4 Scope

Since Big Data is a new phenomenon in the Maritime Industry, and within this, the tanker segment faces unique and unexplored challenges, the research position is that Big Data usage to improve operational and economic efficiency and legal compliance is underrepresented in the literature. Therefore, this project's scope is to assess the current level of Big Data literature presence while building a bridge between the academic literature and industry's realities through several interviews and surveys.

The focus of the research questions in this paper is on research results, and tanker industry application as well as drivers and challenges related to Big Data. The focus is represented in the literature review by mainly investigating new conclusions on the issues and industry representation. The research question's position is that the focus of literature is misaligned with the interests of the industry.

The literature is covered based on selecting relevant tags and criteria input in scholarly search engines of online libraries of Aalborg University Copenhagen and Copenhagen Business School. While the research questions mainly address the Tanker Industry, this does not rule out literature related to the more broadly defined Maritime Industry. Literature that will be highly relevant is expected to be of low volume due to the newness and niche of the topic. If we assume the subject to be under-representation, a comprehensive review will have higher representation confidence.

The literature is grouped according to conceptual and methodical relevance to the topics addressed., allowing us to present the findings with a proper level of granularity.

This paper's audience is both active practitioners in the industry and researchers aiming to understand the current state of research and application of the topic.

5 Theory

5.1 Big Data

5.1.1 Conceptualizing Big Data

Before diving deeper into the research, it is important to understand what ”Big Data” stands for. Big Data is a term which was first used by the two NASA scientists, M. Cox and D. Ellsworth, who described vast amounts of data generated by supercomputers unable to be visualized or processed by technologies of that time (Cox and Ellsworth 1997). Application-controlled demand paging for out-of-core visualization. Nowadays this term is described in dictionaries as “sets of information that are too large or too complex to handle, analyze or use with standard methods”. Even though the newer definition mentions nothing about the lack of technologies to process and analyze Big Data, it is still a complex phenomena due to its different properties.

5.1.2 The 5Vs of Big Data

When people talk about Big Data and describe it from managerial or technical perspectives, they usually mention properties which fall under framework known as 5 Vs of Big Data. This framework tackles such properties of Big Data, as: Volume, Variety, Velocity, Veracity and Value.

Figure 1: The 5Vs

Volume

Even its name solely indicates that such data refers to large volumes of both unstructured and structured data. It can challenge hardware and software employed by Maritime companies not only because such volume must be stored somewhere, but it also must be processed, analyzed, displayed and evaluated. Nonetheless, as (Boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data. Information, Communication & Society, 15(5), 662–679) mention, *"Big Data is less about data that is big than it is about the capacity to search, aggregate, and cross-reference large data sets."* Therefore, such data comes from many different sources inside and outside of ships and must be merged to form an even bigger dataset for further analysis.

Variety
Coming from different sources, data enriches in its variety. To perform analytics towards optimization, a Maritime company would have to gather different kinds of data, such as: data coming from ships' vital components, satellite data about ships location, weather condition etc. This situation is even more complicated as such various data has to be captured differently and potential new data sources have to be smoothly integrated with existing data pipelines (Gupta 2016). An extra layer in this property is added by the fact that there could be many data sources which are not still visible or reachable. This leads to a Maritime company sitting on its data without a clue of its existence or any way to capture it.

Velocity
In order to be able to draw better conclusions on data, it has to be captured for a long period of time and at high frequency. High frequency allows to see even the slightest changes between different ships' components and weather conditions, which would be the reason a Maritime company is using Big Data. However, it also imposes a challenge. This data has to be verified, analyzed for insights, updated and maintained at the same speed, otherwise it will not bring the competitive advantage companies seek (Gupta 2016). This puts high responsibility on the systems in and outside ships, as well as people working with them.

Veracity
This property indicates that Big Data can impose a challenge in its validity and consistency. Are we sure our sensors are accurately capturing data of engine's performance? What about the next ship? Not only could the quality of sensors differ from ship to ship, but also the way it is captured. Such questions could lead a Maritime company to delays in decision-making or loss of time and frustration. If data is unreliable because of low quality, it can lead to weak analysis and misinterpretation which would make decision-making poor and ultimately result in a huge loss of money.

Value
The last traditional part of framework is Value, and it is a very important one. It is a key reason why Big Data has gathered so much interest in many industries (Hagen & Khan, 2014). It refers to the insights hidden behind terabytes of data, waiting to be discovered. However, these insights can only be achieved if all the other properties are in place. If a Maritime company is to gain com-

petitive advantage from its Big Data utilization, it is to have enough of data to analyze (Volume), different sources of data (Variety), high frequency or number of data points entering the system (Velocity), reliable validity as well as consistency (Veracity) and finally – right people who can find this value through advanced analytics and industry-specific knowledge.

In addition to the previously mentioned Vs, Gupta 2016 add an extra one and calls it Vision.

Vision
The last property points out to a very crucial idea that a company has to have a strategy of how it would like to use Big Data and where does it want to go with it eventually. Moreover, it requires a vision to start a Big Data integration project due to high investments and long payback period. Big Data requires significant investments in technology systems and personnel, commitment, and teamwork as well as enough optimism to persist despite the potential failures. "... a vision which is not a one-time, specific goal that can be met, then discarded, increases the prospect of improvements in organizational performance, because such an abstract vision suggests a longer-lasting organization that is desirable to followers, and encourages effective group formation to carry out the vision." (Kantabutra, 2009).

5.2 Literature Review

This section will explore the theory behind our choice of methods regarding the literature review, from purpose and choice of over-reaching approach to scoping, finding the literature, and evaluating the findings.

5.2.1 Understanding the Literature Review

The literature review is an ongoing process meant to provide the researchers the necessary knowledge of contemporary research. It provides a foundation to the statements and arguments in justifying the choice of research (Bryman 2016, Snyder 2019, Cooper 1985).

The literature review consists of an overreaching structure chosen by its relevance to the research purpose. A taxonomy of the literature review further identifies key elements present in all structures (Bryman 2016, Snyder 2019). If the review aims to converge knowledge in a complex field in a transparent and reproducible manner, one might consider the systematic approach (Bryman 2016, Davis, Mengersen, and Bennett 2014). The systematic approach aims to synthesize quantifiable and analyze the literature to produce a quantifiable contribution of measurable effect. This approach has stringent guidelines, making it possible to undertake the review within smaller time frames. The disadvantage of this approach is the assumption of there exists enough quantifiable material within the narrow scope to confidently conclude. It can also be that the material within the field or scope is too qualitative to create a quantifiable analysis confidently. In these cases, it might be more efficient to make a semi-structured (or narrative) approach. The semi-structured approach is criticized as less rigid and more likely to be subject to the researchers' bias (Tranfield, Denyer, and Smart 2003). The semi-structured literature review requires careful attention to the methods and transparency in the search criteria applied to avoid the issues above. This approach can also offer necessary flexibility, which might be useful as there could be a lack of available literature on the topic of research. Due to such consideration, this is the chosen approach for this research.

A Taxonomy of the literature review further identifies elements present in all proper reviews. The exact definitions of and borders between the elements tend to vary slightly (Snyder 2019, Bryman 2016), but in general they are:

1. Purpose
2. Scope
3. Search Strategy
4. Evaluation
5. Analysis

Some experts adds self-reflection as a crucial element often overlooked (Cooper 1985). In this context, self-reflection refers to the ability to reflect over how well the literature was covered, the inter textual coherence created and understanding of bias.

5.2.2 Creating the purpose

It is essential that the purpose is not formed based on a chosen approach but rather that the approach is chosen based on the purpose. The purpose often in the form of a research question or problem statement meant to guide the researchers in searching, selecting, and evaluating the literature.

The purpose should reflect the researcher's objective with conducting the literature review. If the review is meant to criticize current literature coverage or compare evidence, this should be apparent in the purpose. The purpose should also reflect how broadly the researcher aims to target with the review, if it will be convergent or divergent, and who the recipients are. This can be stated either directly or indirectly.

5.2.3 Scoping the review

The scoping process determines the boundaries of the review. The scope should enable the researchers to clearly complete the purpose in an efficient manner. An unclear or inefficient scope can result in lack of rigor and reliability. The enumerated characteristics below 2 presents elements the scope should address.

1.	Focus	What kind of literature the review is addressing
2.	Goals	Why are review being done and to what end?
3.	Perspective	The viewpoint of the review
4.	Coverage	The amount and area of literature to be reviewed
5.	Organization	How is the literature presented
6.	Audience	Who is the recipient of the review
7.	Integration	How does the review add to current knowledge

Table 2: A taxonomy of literature reviews (Cooper 1985, Wong et al. 2013 & Baumeister and Leary 1997)

5.3 Surveys

The project involves the collection of data through interviews and online self-completing questionnaires. The following section explains the theory behind the applied survey methods.

5.3.1 The semi-structure interviews

A semi-structured interview is a method within qualitative research in which an interview guide directs the interviews, but the respondents and the interviewers are permitted and expected to dwell outside the initial scope of the interview guide. In contrast to the qualitative method of structured interviews, rich-text responses are positive. Through exact transcription and coding, such responses can result in more profound and sometimes unexpected perspectives on the issues. Issues with the semi-structured approach can be the increased reliance on the skills of the interviewer. Careful attention is needed when listening to the respondent, asking new questions, and adapting depending on the answers. An interview guide establishes generalization aspects between the interviews by tying the discussion to the research question.

Ensuring Reliability and Validity

When using an approach with a high degree of freedom, such as the semi-structured approach, it is essential to pay extra attention to ensuring Reliability and Validity.

Classical methods of ensuring reliability entail repeatability and reproducibility. Ensuring this is difficult due to the nature of qualitative research, as it heavily relies on the data collection process's current setting. We instead achieve reliability through transparency of the data collection, selection, and synthesizing process.

Structuring the interviews

The paper aims to explore multiple areas with possibly highly differing and highly nuanced perspectives on the topic depending on the context of their work. The chosen structure reflects this by creating generalized questions which invites the interviewee to explain their position according to the problem.

The specific topics to be explored in the interviews are decided by the findings in the literature review and the problem statement.

Before asking interview questions directly related to the research questions, the interviewee's understanding and position regarding Big Data needs to be established. exploring the interviewee's understanding of Big Data is needed. Followed by their position regarding the topic. Then questions related to their usage of big Data can be asked. Issues identified in the literature review will be presented to the interviewee. This will assist in creating points of comparison between the surveys and review.

Selecting target group

The target group is selected using purposive sampling theory, specifically for use in grounded theory and theoretical sampling. This is essentially about selecting the unit of analysis within qualitative analysis. The purpose of this sampling is to discover categories and relationships between units of analysis through coding. Cases are established based on heterogeneity and homogeneity between themselves and the research questions. Within these cases, participants are selected based on their ability to saturate the research question.

Synthesizing

Categories will be formed based on the interview responses through coding. The responses will the be grouped by the categories and the setting of the interviewees. These groupings will then enable the identification of themes and elements in the responses which will create a model of Big Data interest, understanding and application in the industry. If the synthesis process highlights areas which is currently not saturated by the interviews, this will be explored through self-completing questionnaires or request for further interviews depending on allowed time.

6 Methodology

A mixed method inductive qualitative approach based on grounded theory is chosen for this project.

This mixed method consists of a semi-systematic literature review, semi-structured interviews, and self-completing questionnaires. The objectives of the methods are to develop an understanding of academic and industry perspective on the usage, challenges, and drivers involving Big Data in Maritime Tanker industry. The method focuses on the tanker section with sampling selection, but it expected the findings can be generalized over the maritime industry to some extend.

Diagram 2 illustrates the research methodology split into three section. The remainder of the methodology is as follows; (1) Grounded theory, (2) Literature Review, (3) Surveys. Survey encompasses both the interviews and the questionnaires.

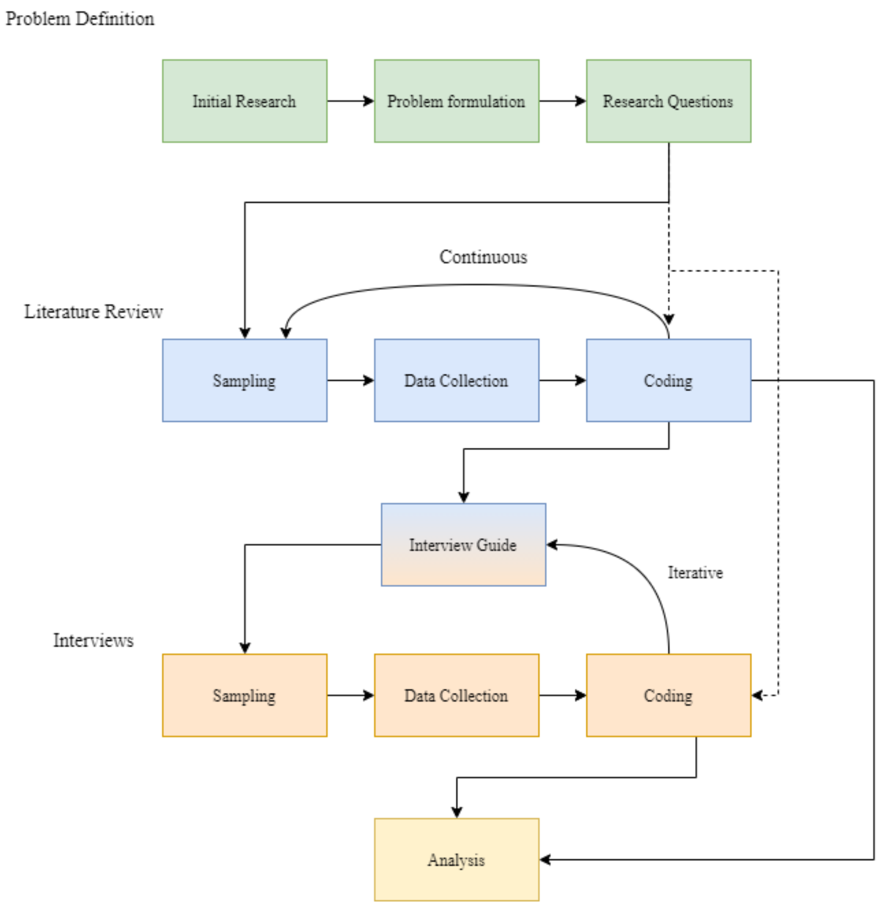


Figure 2: Methodology

6.1 Grounded Theory

Grounded Theory is a framework for working with qualitative data and it was chosen due to it's iterative design and well document processes (Bryman 2016). The framework involves treating the coding of the data as a continuous process, allowing the on-going coding and learning to influence the process of further sampling, data collection and coding. Grounded Theory uses the concept of saturation to describe the point at which coding has reached a sufficient level to apply early review findings in the next part of the project. In the case of this project, this means the development of the interview guide and interview sampling method. The interview guide and interview process is itself iterative, with the coding of interviews allowing changes to the interview guide and sampling depending on the findings. This flexibility of the literature review and interview guide results in a less rigid structure, but increases the researcher's ability adapt to new findings (Tranfield, Denyer, and Smart 2003). To accommodate the Grounded Theory framework, the literature review is semi-systematic, and the interviews semi-structured.

6.2 Literature Review

6.2.1 Purpose

The purpose of the literature review is to develop concepts of current literature coverage of Big Data in maritime through coding. The coding will also be used to generate interview guide and sampling.

6.2.2 Scope

The scoping method is based on a literature synthesis matrix defined by Cooper 1985 and further explored by Baumeister and Leary 1997 and Wong et al. 2013. The highlighted cells in the table below shows categories defining the literature review of this project.

Chraracteristics	Categories			
Focus	Research Results	Research Methods	Theories	Applications
Target	Integration	Criticism	Central Aspects	
Perspective	Neutral representation	Representation of a position		
Cover	Completely	Completely and selective	Representative	Fundamental
Organization	Historically	Conceptually	Methodically	
Target Group	Specialized Scientists	General Scientists	Practitioners	General Audience

Table 3: Scope

6.2.3 Sampling

It has been decided to start the research using the databases available to online libraries of Aalborg University and Copenhagen Business School. Should the need arise, the search will also incorporate databases of IEEE. Primo, Google Scholar, Taylor & Francis, as well as sources mentioned by the selected literature. Any papers will be included if they relate to the Maritime Industry and identify Big Data in the same way as this research. There is no geographical restriction to the origin of the papers, though the language will be limited to English only.

The discovered literature will be appraised according to their contextual relevance to the problem statement. The problem statement includes interest in both academic and industry, with the hypothesis of the academic coverage being insufficient.

Part of the design phase is to create search strings which will yield enough literature sources needed for the research. After a brainstorm session the following terms were identified of interest to Big Data in Maritime Industry and chosen research questions:

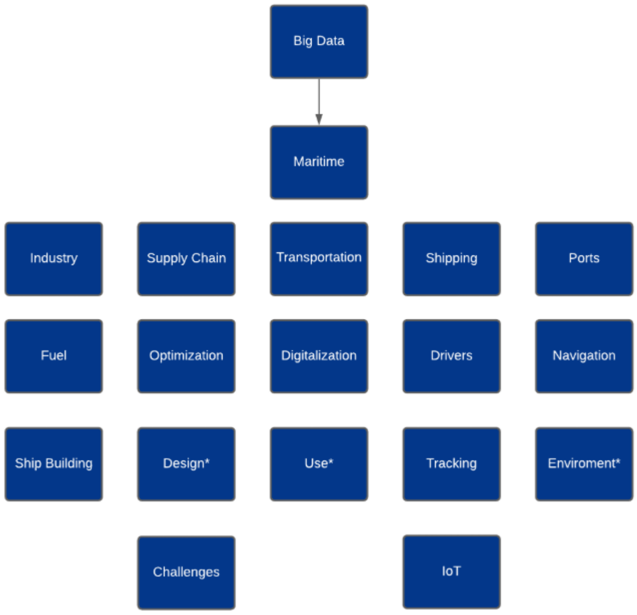


Figure 3: Terms used for search strings

Using "Big Data" alone would not be enough as the term is present in many other industries, therefore it has been decided to use it in conjunction with "Maritime", limiting the search to the use of Big Data analytics in the Maritime Industry. Excluding "Big Data" string lead to papers discussing general concepts about Maritime industry or in some cases to those, discussing other parts of digitalization in Maritime Industry, therefore it has been kept throughout all the strings. Some terms had a "*" sign next to it in order to incorporate its derivatives during the search, for example by searching for Design*, other forms as Designing, Designed etc. Search strings were obtained by connecting the above-mentioned terms together and used in the search. They are presented in the table below:

Strings	
"Big Data" AND "Maritime" AND "Industry"	
"Big Data" AND "Maritime" AND "Supply Chain"	
"Big Data" AND "Maritime" AND "Transportation"	
"Big Data" AND "Maritime" AND "Shipping"	
"Big Data" AND "Maritime" AND "Ports"	
"Big Data" AND "Maritime" AND "Fuel"	
"Big Data" AND "Maritime" AND "Optimization"	"Big Data" AND "Maritime" AND "Drivers"
"Big Data" AND "Maritime" AND "Digitalization"	"Big Data" AND "Maritime" AND "Navigation"
	"Big Data" AND "Maritime" AND "Environment*"
	"Big Data" AND "Maritime" AND "Use*"
	"Big Data" AND "Maritime" AND "IoT"
	"Big Data" AND "Maritime" AND "Design*"
	"Big Data" AND "Maritime" AND "Ship Building"
	"Big Data" AND "Maritime" AND "Tracking"
	"Big Data" AND "Maritime" AND "Challenges"

Figure 4: Designed search strings

Once the sampling is designed, the next phase initiates it. Snyder 2019 recommends launching a pilot test of the review process to assess how well it works as well as what output it yields and adjust it if necessary. The literature itself can be selected in three different ways, each varying in the degree of time it consumes as well as thoroughness of results they yield:

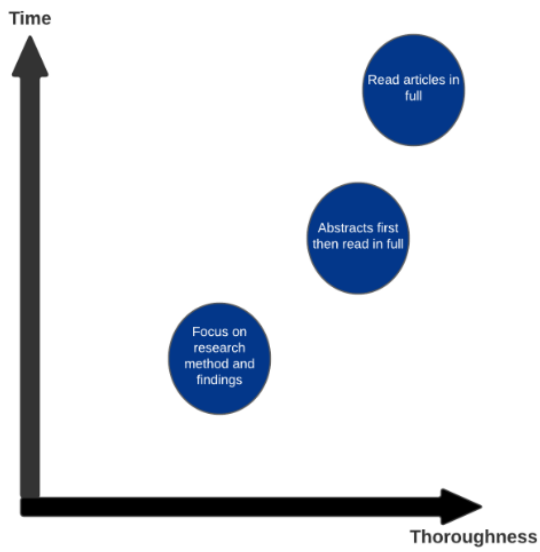


Figure 5: Ways to select the literature

For this research, the following process is adopted:

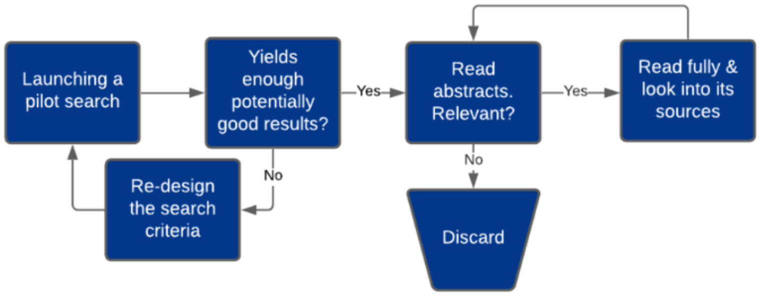


Figure 6: Literature review process

First, a pilot search will be launched at different online databases, using the strings mentioned in the previous step. If the number of good literature sources will not be sufficient, the strings and overall approach will be re-designed until it yields a good number of results suitable for the research. Then the abstracts of the sources will be read to evaluate their relevance to the study. The irrelevant studies will be disregarded, decreasing the total number of found literature. Whenever an article will lead to another article, which might be useful, its abstract will be read, and the process will repeat itself. Such process will help the study save time whenever possible whilst still not impacting the thoroughness quality of the study.

6.2.4 Data collection

Once the first strings were able to yield sufficient number of results, it has been decided to proceed with the original plan. After searching the online libraries of AAU and CBS with the outlined strings, the following results were obtained:

Strings	№ of Results (AAU)	№ of Results (CBS)
"Big Data" AND "Maritime" AND "Industry"	3	21
"Big Data" AND "Maritime" AND "Supply Chain"	0	5
"Big Data" AND "Maritime" AND "Transportation"	7	7
"Big Data" AND "Maritime" AND "Shipping"	1	15
"Big Data" AND "Maritime" AND "Ports"	5	10
"Big Data" AND "Maritime" AND "Fuel"	1	8
"Big Data" AND "Maritime" AND "Optimization"	8	8
"Big Data" AND "Maritime" AND "Digitalization"	1	6
"Big Data" AND "Maritime" AND "Drivers"	2	0
"Big Data" AND "Maritime" AND "Navigation"	2	23
"Big Data" AND "Maritime" AND "Environment*"	7	3
"Big Data" AND "Maritime" AND "Use*"	0	0
"Big Data" AND "Maritime" AND "IoT"	2	4
"Big Data" AND "Maritime" AND "Design*"	6	0
"Big Data" AND "Maritime" AND "Ship Building"	0	2
"Big Data" AND "Maritime" AND "Tracking"	2	5
"Big Data" AND "Maritime" AND "Challenges"	7	5
	54	**122**
	176	

Figure 7: Search results

During the initial literature search, the relevance of one or another source was determined based on presence of key terms in its name or abstract. If a paper contained term "Big Data" in its name, it has been automatically included. Whenever a paper did not contain a term "Big Data" in its name, the abstract was read to assess its relevance to the research. Finally, if "Big Data" was missing in the abstracts, the whole paper was searched for it using string search commands. The total number of initial literature search resulted to be 176 papers. This list was then narrowed down when duplicates between the strings as well as two university libraries were eliminated, and abstracts were read in full. After reading the abstracts it has been observed that some of the papers related to Maritime area, but outside of the research scope, for example those targeting biological perspectives. As part of the continuous process of finding literature, the list was also expanded on based on references and sources of known literature. The final number of relevant sources resulted to be **100 papers**.

After the relevant literature was found using the designed approach, it is important to standardize the way these articles will be evaluated, such as identifying similar patterns between the papers and coming up with the right way for coding these similarities. Eventually, the standardized evaluation should lead to the answers for the research questions.

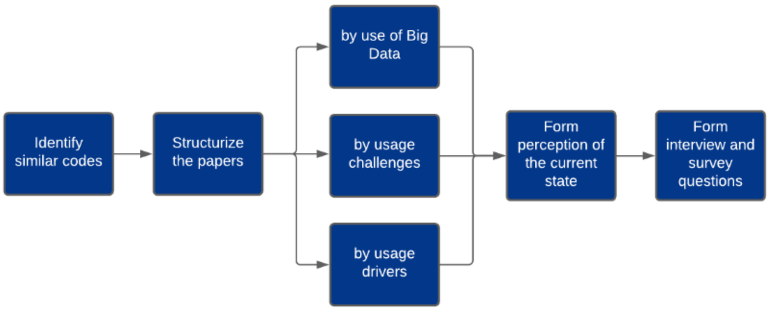


Figure 8: Evaluation approach

The evaluation process is showed in Figure 8. It starts with identifying the similarities between the found literature sources and structuring them by different areas of Big Data as well as challenges and drives associated with it. Eventually, literature review findings will help forming a perception about the current state of Big Data in Maritime Industry and its literature, which will lead to development of various questions used during interviews and surveys. The interviews and surveys will build a bridge between literature perception and current realities, which will be used to dig deeper into the situation.

6.3 Surveys

Two types of surveys, semi-structure interviews and self-completing questionnaires, were designed using the learning from the literature review with the purpose of answering the research questions from the perspective of actors in the maritime tanker industry.

The semi-structure interviews followed an iterative approach, in which feedback and coding of completed interviews were used to alter the interview guide, sampling and interview process.

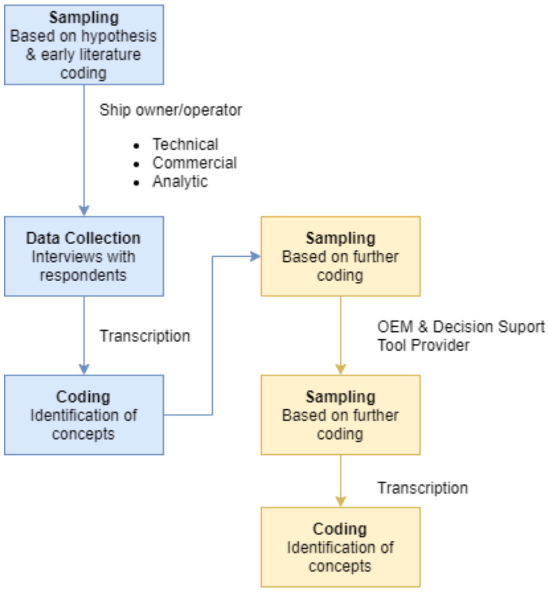


Figure 9: Interview process example

Semi-structured interviews are executed following an interview guide based on the research questions and preliminary findings from the literature review. Self-completing questionnaires are designed to amend the interviews. They are sent to a wide net of potential stakeholders. The following sections describe the methods applied in the surveys, in forming the interview guide, reaching out to potential stakeholders, and synthesizing results.

6.3.1 Interview Guide

The Questions first establishes a baseline with face-sheet information. Followed by an introduction of the research question and findings so far. The questions are then designed to invite different perspectives on the research questions, with different levels of abstraction and precision. As with the nature of semi-structured interviews, the interview guide is not followed exhaustively. Additional questions are likely to be asked depending on the participants responses.

Originally, the interview guide had the participants describe their understanding of Big Data before being introduced to the research, as to not introduce bias, but as described in interview findings, this introduced problems, and the order was switched.

Interview Guide
Request permission to record

Establishing a baseline

1. What are your responsibilities, roles and positions within the organization?
2. How long have you been working in the organization and in the industry?

   (If Relevant) Have you had other roles and/or responsibilities within the organization or industry?

   Present research Purpose

3. What is your understanding of the concept "Big Data"?

   Present Big Data in the context of our research

Big Data Application

4. What kind of data do you work with?

   (If Relevant) What is the volume of the data?

   (If Relevant) What is the frequency of the data?

5. What tools/methods do you use to process/evaluate data

   (If Relevant) What tools/methods for Big Data do you know of?

6. Do you see potential of existing Big Data improving your work?

   Present preliminary application findings from literature review

7. Do you recognize our findings in your experiences with Big Data?
8. What is your experience with this kind of data?

   (If no experience) Do you know any work being done with Big Data?

   (If any experience) Can you elaborate on this?

Challenges with Big Data

Present preliminary findings of challenges from literature review

9. Do you recognize any of these challenges in your experiences with Big Data?

   Are there any other challenges preventing you from using existing Big Data?

   Is there Big Data sources you do not currently have, which could be useful?

Drivers of Big Data

Present preliminary findings of Big Data drivers from literature review

10. Are above-mentioned drivers similar to the ones you experience?

    Anything else motivating you to use Big Data analytic?

6.3.2 Surveys

The surveys provides a qualitative perspective through semi-structured interviews, complemented by self-completing questionnaires.

The surveys are consists of an interview guide, target group, data collection (conducting interview, sending questionnaires), and coding. The continuous coding affect the interview guide such that questions are adapted, order is changed, and new participants are identified.

The interview guide is based on early coding results which shows a tendency towards tendency towards usage being AIS based, challenges being related to data volume and variety, solutions being AI and platforms, and drivers being increased safety, cost reduction, and business insight.

The target group has the homogeneity of being a part of the maritime tanker industry, and heterogeneity of different parts of said industry. That heterogeneity is split into three cases:

6.3.3 Target Group

Homogeneity: Maritime Industry

Heterogeneity: Different parts of the Big Data value chain

Case 1: Ship owners/managers

Case 2: Decision-support tool providers

Case 3: OEM's

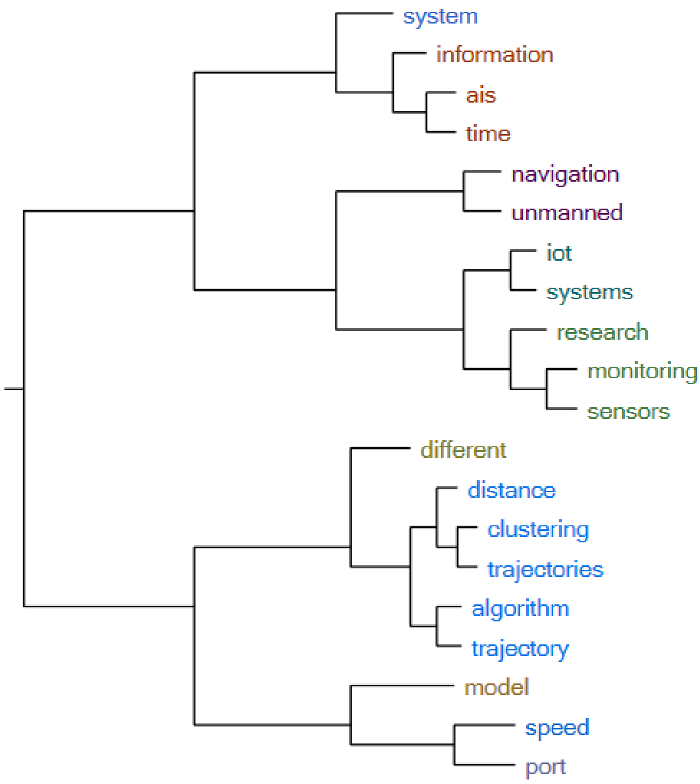


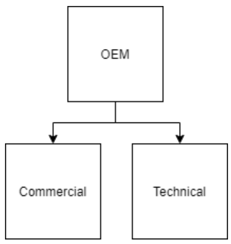
Figure 10: Caption

Early coding findings indicate focus on the vessel operators utilizing Big Data generated by the vessels for operational or commercial in the form of fuel reduction, safety or business insight. Original sampling only included this group, but the responses of the participant showed the need to expand the sampling group to include case 2 and 3.

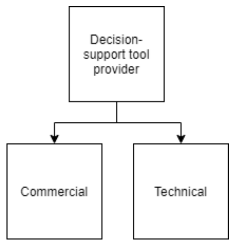
Decision-support tool providers are chosen due to articles presenting data handling and decision-support tool platforms in correlation with challenges regarding Big Data utilization.

While OEM's where not mentioned in literature, it was found relevant to explore the position of the Big Data providers.

Among the cases, a wide range of potential stakeholders within the cases is contacted. The list below shows the contacted stakeholders



(a) OEM



(b) Tool Provider

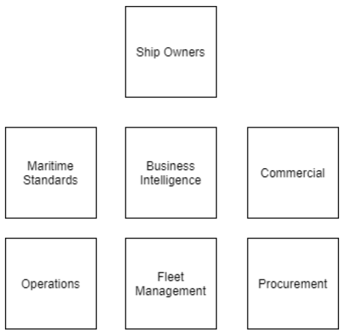


Figure 11: Ship Owners

Positive response was received from two ship owners, with perspectives from technical, commercial, business intelligence, and form one decision-support tool provider with commercial and technical, and from one OEM with commercial and technical perspective. Additional attempts were made to get response with questionnaires, but did not recieve useful responses.

6.4 Coding

Coding is a labeling and categorization process for working with qualitative data. Our process involves the use of the Computer-Aided Qualitative Data Analysis Software (CAQDAS), NVIVO. NVIVO is powerful file handling tool, with build in code analysis tools. The approach chosen for this research, is as mentioned above, based on a continuous process. An initial understanding is created by the researchers using existing knowledge and already known literature. The initial understanding takes the form of the research questions in this project. The nodes used to code are created based on the research question. The nodes chosen for the project are Use, Drivers, Challenges.

Nodes

- Challenges of Big Data
- Drivers of Big Data
- Use of Big Data

These nodes were expanded upon as the nature of the problem became clearer. The nodes "Definition" and "Current State" is added as individual nodes.

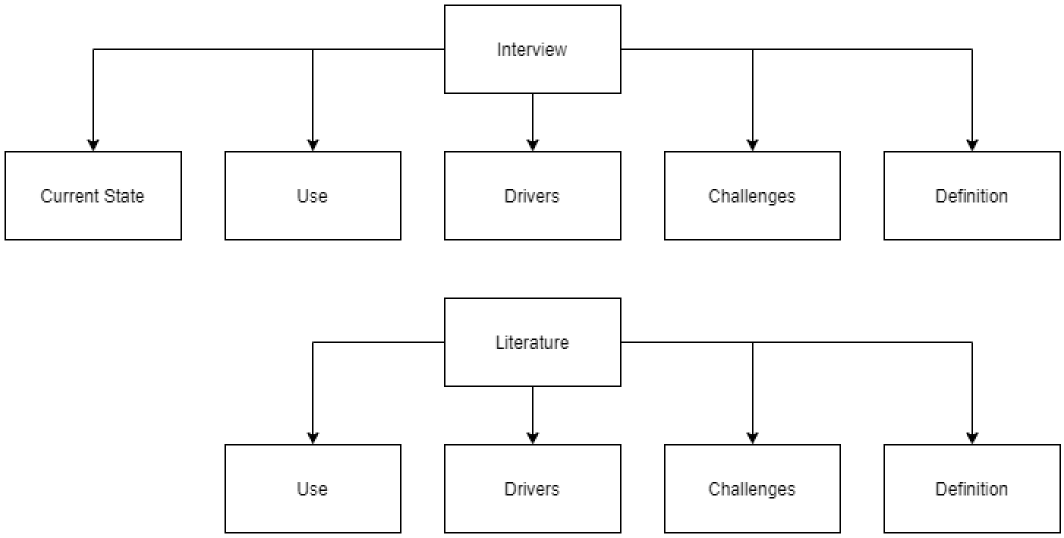


Figure 13: Coding Hierarchy

Current State node is for when participants mentions relevant general information about the industry or company. Use is for practical Big Data usage. Drivers incorporate all the reasons for working with Big Data. Such as imposed industry regulations or decreased costs. Definition covers different ways how people perceive Big Data, which can vary significantly in Interviews. Deriving definition of Big Data from literature would allow to form a baseline for the interviews whilst making sure that the term is discussed on the same premise.

A potential criticism of coding is loss of context, resulting in a misrepresentation of the paper's intentions. The project uses context mapping in the form of cluster analysis to prevent a misunderstanding. The cluster analysis can be used to see topic correlation. Using Pearson Correlation Coefficient and a cluster number where no noticeable changes occurred when increasing it, the following clusters emerge.

7 Literature Findings

This section section presents literature findings by introducing existing literature reviews and their findings with a comparison of methods. The remainder of the section is arranged as follows: 7.1 Literature list, 7.2 Article Statistics, 7.3 Definition, 7.4 Use, 7.6, and 7.5.

In order to assess the literature representation of Big Data in the Maritime Industry, there were reviewed 100 papers. The initial number of papers mentioned in Section** increased after sources were analyzed and relevant papers were added to the final list. All of the papers were read and coded to get a literature representation of Big Data. Afterward, there were added tags to segment the papers based on discussed topics and a statistical overview. The tags were designed after the literature review process was concluded and all papers were coded, allowing for a generalization of mentioned topics. One paper could have different tags based on the discussed topics in it. The table below describes these tags:

7.1 Literature list

Citation	Year	Tags	Segment
Fernández et al.	2015	Platform, Data handling	Port
Lytra et al.	2020	About Big Data, Platform	General
Lepore et al.	2016	Fuel (Environment), Predictions, Navigation	General
Lee et al.	2018	Fuel (Environment), Speed	Liner
Li et al.	2017	Navigation	General
Dhar	2016	Navigation	General
Kontopoulos, Varlamis, and Tserpes	2021	Navigation , Safety	General
Tsou	2019	Logistics, Data handling	Port
Zhang and Lam	2019	About Big Data, Management	General
Tsou	2017	Safety, Navigation	Port
Farah Al Kaderi and Rida	2020	Logistics, Management	Port
Munim et al.	2020	Literature review, About Big Data	General
Wang et al.	2020	Navigation	General
Wen et al.	2021	Predictions,	General
Ullo and Sinha	2020	Literature review,	General
Pezzani and Heller	2019	About Big Data, Navigation	General
Kamolov and Park	2019	Navigation , Logistics, Platform	Port
Lin and Xiao	2020	Predictions,	General
Adland, Jia, and Strandenes	2017	Commercial,	General
Kokkinakos et al.	2017	Data handling, Management	General
Mirović, Miličević, and Obradović	2018	Energy, Logistics, Safety	General
Jović et al.	2019	About Big Data,	General
Kacprzyk	2016	Challenges, Predictions, About Big Data	Liner
Han and Yang	2020	Navigation , Safety	General
Zaman et al.	2017	Challenges, About Big Data, Predictions	Liner
Bo and Meifang	2020	Logistics, Operations	Port
Hu	2020	About Big Data, Platform	General
Zhang et al.	2018	Navigation	General
Fiskin, Cakir, and Sevgili	2020	Navigation , Safety	Tugboat
Ryazanov et al.	2021	Predictions,	General
Li et al.	2016	Platform,	General
Wang, Ma, and Chen	2017	Platform,	General
Kim, Jung, and Park	2021	Fuel (Environment), Energy, Predictions	Liner
Xiao, Chen, and Mcneil	2021	Operations, Logistics	General

Table 4 continued from previous page

Citation	Year	Tags	Segment
Fruth and Teuteberg	2017	Literature review,	General
Cui	2018	Data handling, Platform	General
Moreira, Vettor, and Soares	2021	Navigation , Fuel (Environment), Speed	General
Waluyo et al.	2017	Navigation , Platform	General
Yan et al.	2018	Navigation , Speed, Fuel (Environment)	General
Cepeda et al.	2020	Fuel (Environment), Navigation	General
Jia, Prakash, and Smith	2019	Commercial, Predictions	Bulker
Zhang, Huisingh, and Song	2019	Navigation , Safety	General
Fujino, Claramunt, and Boudraa	2018	Navigation	General
Wang, Claramunt, and Wang	2019	Navigation	General
Sheng and Yin	2018	Navigation	General
Man et al.	2020	Energy, Fuel (Environment)	General
Yang et al.	2019	Literature review, Navigation	General
Kenyon et al.	2017	Operations, Data handling	Port
Wang et al.	2021	Safety, Logistics	Port
Lin	2020	Logistics, Operations	General
Arifin et al.	2017	Safety,	General
Qiao et al.	2019	Safety, Logistics	General
Perera and Mo	2016	Operations, Navigation	General
Lei	2017	Safety, Navigation	General
Fu et al.	2020	Safety,	Port
Peng et al.	2018	Operations, Logistics, Commercial	Port
Isenor et al.	2016	Data handling, About Big Data	General
Yoo	2018	Navigation , Safety	General
Jia et al.	2017	Navigation , Logistics, Commercial	Port
Xiao et al.	2020	Logistics, , Commercial	Tanker
Thombre et al.	2016	Safety, Navigation , Fuel (Environment)	General
Brouer, Karsten, and Pisinger	2018	Logistics,	Liner
Zhang et al.	2020	Safety,	General
Karvelis et al.	2020	Predictions,	General
Jeon et al.	2018	Fuel (Environment), Operations	General
Venskus et al.	2019	Navigation , Safety	General
Zhang and Zheng	2020	Logistics, Data handling, Platform	General
Yang et al.	2016	Data handling,	General
Sarabia-Jacome et al.	2019	Data handling, Logistics	Port
Perera and Mo	2018	Operations,	General
Gao and Shi	2019	Navigation	General
Perera and Mo	2018	Navigation , Fuel (Environment)	General
Luaces and Karimpour	2018	Platform, Data handling	General
Fernández et al.	2016	Platform, Data handling	Port
Coraddu et al.	2017	Fuel (Environment), Predictions	General
Yuen, Xu, and Lam	2020	About Big Data,	General
Kallimani	2018	About Big Data, Challenges	General
Lee et al.	2020	Navigation , Predictions	General
Wu, Chen, and Tsau	2016	Logistics, Operations	General
Alan	2018	Literature review,	General
Zhou et al.	2020	Navigation , Predictions	General
Alop	2019	About Big Data, Challenges	General
Cheng et al.	2018	Navigation , Predictions, Fuel (Environment)	General
wLei	2019	Navigation , Predictions	General
Aiello, Giallanza, and Mascarella	2019	Literature review, About Big Data, Challenges	General
Hu and Yu	2020	About Big Data,	General
Shaw and Tzu	2019	Fuel (Environment), Operations	Tanker
Zerbino et al.	2019	Logistics, Data handling	Port
Jeong, Woo, and Park	2020	Shilpbuilding, Predictions	General
Li et al.	2018	About Big Data, Energy, Challenges	General
Sanchez-Gonzalez et al.	2019	Shilpbuilding, About Big Data, Literature review	General
Schoening	2018	Navigation	General
Yang et al.	2019	Predictions,	General

Table 4 continued from previous page

Citation	Year	Tags	Segment
Li, Yang, and Han	2020	Data handling,	General
Lambrou, Watanabe, and Iida	2019	About Big Data,	General
Huang et al.	2014	About Big Data, Challenges	General
Mosavi et al.	2019	About Big Data, Predictions, Energy	General

Table 4: Literature List

The table below explains the meaning of the given tags.

Tag	Description	Count
Navigation	Improving navigation and trajectories using AIS data	33
Fuel (Environment)	Optimizing fuel consumption to decrease environmental impact	13
Management	Discuss human aspect of working with Big Data	3
Operations	Improving operations using Big Data	10
Logistics	Logistical improvements (scheduling, port entry)	17
Platform	Development of a software or a system for Big Data usage	11
Data handling	Discuss working with data (validity, transmission etc.)	13
Literature review	Review present literature about a certain digitalization topic	7
About Big Data	Discuss topics around Big Data (challenges, definition etc.)	19
Predictions	Using Big Data to predict a certain phenomenon	17
Safety	Adress safety improvements using Big Data	15
Speed	Adress speed optimization using Big Data	3
Commercial	Discuss commercial use of Big Data	5
Energy	Adress energy optimization using Big Data	5
Challenges	Discuss challenges surrounding Big Data use	7
Shipbuilding	Discuss how Big Data can be used to improve shipbuilding	2

Table 5: Tags

The graph below represents percentage distribution of the mentioned topics. Group "Other" represents tags which had a count of less than 13.

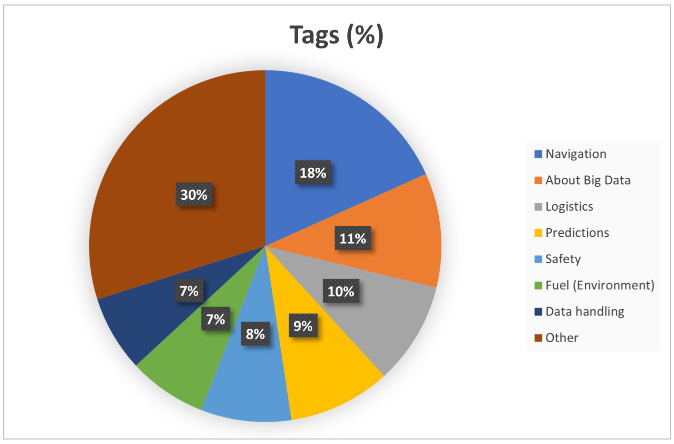


Figure 14: Percentage of tags

It can be observed that most of the present literature address the topic of navigation and improving vessels' trajectories, earning the biggest percentage of 18%. It has been quite often discussed in conjunction with "Safety" improvements which accounts for 8% and of available literature. Sometimes "Navigation" was also seen with "Fuel (Environment)". Nonetheless, the latter was also seen on its own and constitutes 7% of found literature. "Fuel (Environment)" ties with "Data handling". 11% of papers were devoted to discussing a phenomenon know as Big Data as well as its meaning to the Maritime Industry. It is closely followed by "Predictions" and "Logistics".

Papers were also segmented by its publication year and relevance to a certain Maritime Segment. The graph below represents the number of papers published in the recent years:

Figure 15: Year of publication

Even though there were no restrictions applied to published year, there were no papers found earlier than 2014. Another reason could be the fact that Big Data is a recent topic and therefore only recently made its way to the Maritime Industry. First two years share roughly the same amount of published papers, whereas the upcoming years only increased in the total number, indicating at increasing interest towards to and practicies of Big Data.

As can be seen below, most of the papers are not referring to a specific Industry actor and are therefore marked as "General". Many papers refer to "Ports" and logistical optimizations happening there. "Liner" and "Tanker" share the third and fourth places respectively, whilst the "Other" refer to tied "Bulker" and "Tugboat".

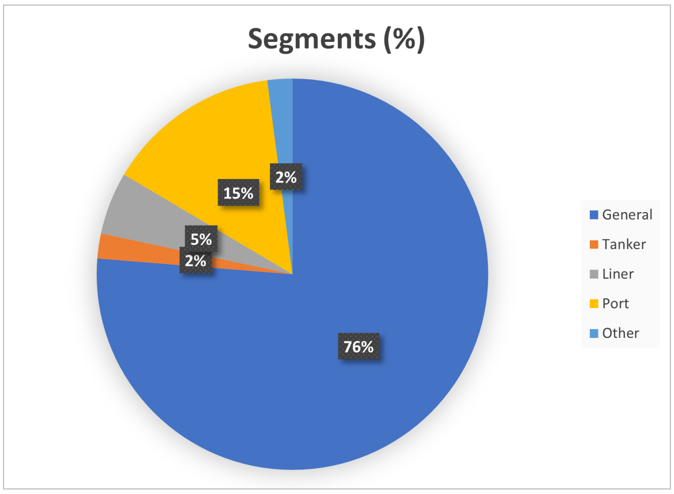
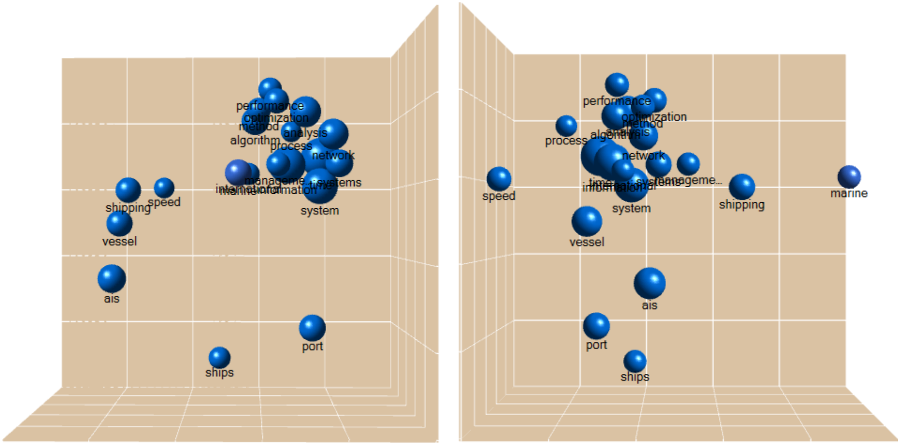


Figure 16: Industry's Segments

7.2 Cluster diagram

Figures 17a and 17b visualizes a 3D cluster analysis using Pearson's Correlation Coefficient based on a text frequency query over the included literature. The distance between two points shows the correlation value between said points, and the size of the point is the word frequency. The strength of such a model is to showcase common connections between topics. An issue can be that there is an underrepresentation of important topics if often abbreviated or shortened. Figure 18 represents the same data-set, with the "nearness" of branches illustrating the Correlation between to topics. The CAQDAS software NVIVO used to compute the values represented on the graphs.



(a) 3d Cluster Map *(b) 3D Cluster Map*

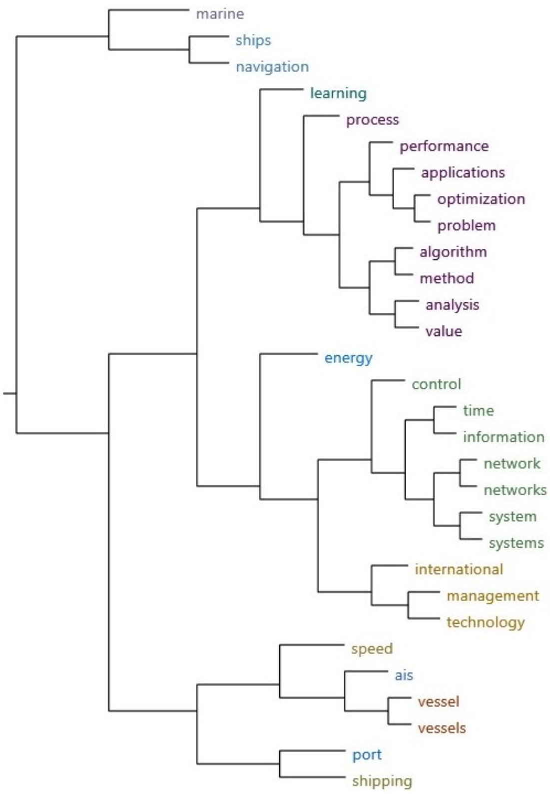


Figure 18: Dendrogram of cluster analysis

The diagrams are used in forming the interview guide, specifically when presenting the focus of literature to the participants.

7.3 Definition of Big Data

Many literature sources started with discussing what Big Data is and some refer to it as an important advancement within the Maritime Industry (Yang et al. 2019a). It has been observed that they mention different properties relating to the traditional 5 Vs of Big Data. Munim et al. refers to Big Data as "large amounts of data volume" whilst Mirović, Miličević, and Obradović and Snijders, Matzat, and Reips add to that that such data sets are "so large and complex that traditional software is unable to process them". Gantz and Reinsel estimated that nowadays the total amount of Big Data in the world should reach 40 trillion gigabytes. Lohr points out that "Big Data" term is often used not only when referring to the data itself, but also to the advancing trends in technology which exploit the opportunities such data offers. Such trends could be the increased use of Artificial Intelligence or Machine Learning techniques (Munim et al. 2020). De Mauro, Greco, and Grimaldi adds that such data is of a high volume, velocity and variety. Mirović, Miličević, and Obradović further elaborates that variety refers to "heterogeneity" of data types and sources it is coming from. Gandomi and Haider refers to velocity as the speed rate at which data is generated and analyzed. Jeon et al. when discussing prediction of fuel consumption, mentions that such data is updated and gathered in real-time from multiple sources. An example of such mix is presented by Venskus et al. when discussing traffic anomaly detection. They mention that a maritime trajectory analysis includes vessel identification data, meteorological data, traffic parameters (like speed and rotation) and form a "large-scale complex data structure" which is "challenging for human-based analysis and anomaly detection"

7.4 Use of Big Data

This section addresses different uses of Big Data mentioned in the literature.

After analyzing and coding above-mentioned papers, clear patterns and distinctions between them were formed. Figure 19 represents a word cloud of code "Use", presenting most frequent words. According to the cloud and conducted analyzes, Big Data use can be generally divided into several categories. There were also smaller groups, partly using Big Data, such as Shipbuilding (also including system design for unmanned vessels) or Commercial, where there were papers discussing business profitability of utilizing Big Data. However, such groups were not big enough to form a specific.

Figure 19: Use Word Cloud

Navigation
Navigation plays one of the biggest parts in Big Data use in Maritime Industry according to the literature. The main interest towards navigation arises from the possibility to avoid collision between the vessels, which will increase safety of both unmanned and traditional vessels as well as route optimization, which reduces fuel consumption hence improving environmental situation and decreasing operational costs. This would not be possible without another important technology – Automatic Identification System.

AIS
Perhaps the biggest publicly available source of Big Data is Automatic Identification System (AIS), developed in the 1990s with the main goal of preventing ship collisions and improving navigation safety (Yang et al. 2019a). It enables ships equipped with it and coastal authorities communicate between each other over a long distance (Yang et al. 2019a). AIS operates by capturing and transmitting vessel's information in the background at regular intervals whilst also receiving information from other vessels equipped with AIS (Dhar 2016). AIS therefore offers a possibility to work with large volume and high velocity data, already publicly available for those interested. And there are many of such examples.

Sheng and Yin 2018 propose ship trajectory clustering model to extract shipping route behavior based on AIS data, converting complex data in reliable information whilst helping Tianjin port authorities to improve overall safety in the Maritime environment. The model is able to detect outliers in AIS data transmitted by a vessel, indicating at potential abnormal behavior, which could both notify port decision makers and send a warning to other vessels.

Li et al. 2017 implement Big Data analytics with a similar purpose. They overcome some limitations of the previous method by implementing a multi-step trajectory clustering to perform trajectory analysis to identify abnormal patterns and mine customary route data to increase transportation safety.

In addition to the previous findings, Dhar 2016 developed an algorithm which would use Spatio-Temporal variables to narrow down data to a manageable amount without losing important information needed to derive anomalies in vessels' trajectories.

In (Yoo 2018), this practice was also expanded to incorporate vessels without installed AIS system, such as fisher boats. Near-miss density map was created around the southern coastal sea area of Korea using RGB colors with the purpose of increasing shipping safety using AIS data and fishing boat location database.

Fujino, Claramunt, and Boudraa Oct 27, 2018 contributes with yet another attempt to improve maritime safety with a way to extract course patterns from AIS data and send real-time warnings against off-course using Machine Learning and LDA algorithm.

Tsou 2019b conducts safety assessment for a Keelung Harbor entry using AIS data combined with weather data. The model provides decisions makers at ports with a possibility to detect any abnormalities in vessel's behavior at any given time with different weather conditions. In addition to that, this model can be used to establish an operation model for harbor's traffic flow.

Fuel consumption
Environmental regulations play a very important role in motivating companies explore Big Data opportunities. Therefore, another important topic mentioned by many papers relates to fuel consumption and CO2 emissions. Generally, it has been observed that fuel consumption is decreased by either improving navigation strategy or finding an appropriate speed level for a route. Some of the examples follow.

Lepore et al. 2017 compares different regression techniques to predict CO2 emissions using navigational data. Data gathered over one year was used for this analysis resulting into smarter predictions and outlined relationship between route length and vessel's speed.

Jeon et al. 2018 predicts fuel consumption of the vessel's main engine using different data sources like navigation, weather, ship operation etc. As in many other examples, the process started with data handling and incorporated different approaches such as building regression models and artificial neural networks with the latter yielding better results.

Lee et al. 2018 use weather archive data to assess the impact of weather conditions, in particular wind and current, on a given voyage route and its fuel consumption.

The proposed algorithm considers trade-off between minimizing fuel consumption due to decreased speed and maximizing service level agreements.

Kim, Jung, and Park 2021 utilized in-service data from a 13.000 twenty-foot equivalent unit (TEU) ship to predict fuel consumption. After making a comparison between multiple linear regression and an artificial neural network models, the latter showed better results. They were able to outline different operating conditions yielding to more optimized fuel consumption, such as a draught level. Using this model, ship's navigator would be able to compare different routes on consumed fuel and choose the most optimal one.

Coraddu et al. 2017 uses three approaches of predicting fuel consumption: White, Black and Gray Box models. After testing the models on the historical data gathered during two recent years, the authors developed a trim optimization technique for reducing the fuel consumption which depicted promising results.

Shaw and Tzu 2019 provides a different perspective to a traditional understanding of the relationship between ship's draft and consumed fuel. Their results indicated that deep draft resulted in a better fuel consumption whilst allowing the ship to carry more cargo.

Port
Ports are one of the most important parts of Maritime supply chain. As can be seen from the Figure 19, it is respectively represented in the Big Data literature when discussing its use. There are several major subcategories of Big Data usage at ports, such as Information flow & Efficiency (Sarabia-Jacome et al. 2020, Fernández et al. 2018, Zerbino et al. 2019) Port State Control (Wang et al. 2021, Fu et al. 2020, Tsou 2019a) inspections and traffic (Sheng and Yin 2018, Tsou 2019b). Since, traffic was already partly explained in the navigation section, the first two categories will be further described.

Information flow & Efficiency
Handling hundreds of ships daily, ports always require optimization from the perspectives of logistics and operations whilst generating big volumes of data. This data needs to be efficiently shared between the port decision-makers, which sometimes poses a problem. Therefore, a big part of literature related to ports focuses on designing and improving data flow platforms.

Sarabia-Jacome et al. 2020 discusses limitations of Electronic Data Interchange (EDI) and Port Community Systems (PCS) whilst introducing Seaport Data Space (SDS) with an integrated Big Data architecture allowing for a better operations planning.

Fernández et al. 2018 builds an online platform called SmartPort for visualization and management of a seaport data, preparing it to be implemented at a port. In Fernández et al. 2016 they dive deeper into how the platform operates and implement it at Las Palmas de Gran Canaria port.

Zerbino et al. 2019 applied process mining to a seven-month dataset of freight export at one mid-sized port, identifying several process inefficiencies and proposing ways to mitigate them.

Port State Control (PSC)
Another identified sub-category of Big Data use relates to PSC inspections. Improving ship detention predictions improves the overall maritime safety whilst increasing ports' operation efficiency.

Wang et al. 2021 introduces a model to analyze dependencies among different risk factors which have influence on PSC inspections based on a data gathered for three years. Their results allow for improvements in model efficiency used for ship detention predictions.

Fu et al. 2020 discovers correlations between parent ship deficiencies PSC inspection dataset, identifying major property indicators, like ship type, age, deadweight etc. Their findings could provide basic guidelines which would improve overall ship inspection efficiency.

Tsou 2019a tries to reduce overall detention rate of ships by analyzing PSC data from 2000 to 2016 and improve efficiency of the inspection by "eliminating subjective bias of human experience to some extent".

7.5 Drivers of Big Data

In the literature, Drivers refer to motivators for working with Big Data utilization. Figure 20 illustrates a word frequency query on the drivers coding.

Figure 20: Drivers Word Cloud

From the Driver queries, we identify the following drivers.

Environment
In the literature, emission reduction is a very large driver, coming from different regulations. Some articles refer to regulatory bodies and international agreements, such as International Maritime Organization (IMO) and Kyoto Protocol (Kim, Jung, and Park 2021, Shaw and Tzu 2019, Dhar 2016). This is explored through many fuel reduction practices (Fu et al. 2020, Jeon et al. 2018, Shaw and Tzu 2019, Lee et al. 2018, Kim, Jung, and Park 2021, Xiao, Chen, and Mcneil 2021), and to some extend through the ones focusing on route optimization (Wang et al. 2020). Some articles relate the need for Big Data to voyage and vessel specific information for research institutes and governmental bodies to better understand environmental impact of oil transport and sale (Jia, Prakash, and Smith 2019), Zerbino et al. 2019, Aiello, Giallanza, and Mascarella 2020).

Cost
As with environment, fuel reduction is a large cost method in Big Data literature. Some articles combine cost and environment with fuel reduction (Lee et al. 2018, Kim, Jung, and Park 2021, Xiao, Chen, and Mcneil 2021), while others focus solely on cost (Jeong, Woo, and Park 2020, Coraddu et al. 2017). Route and operational optimization is also related to cost savings (Jeong, Woo, and Park 2020, Xiao, Chen, and Mcneil 2021, Jeon et al. 2018). Also, increased information for decision making and operations process improvement (Lytra et al. Jun 2017, Fujino, Claramunt, and Boudraa Oct 27, 2018).

Safety
Many articles argue for the need of increased safety through improved navigation (Wang et al. 2020 Kokkinakos et al. Jun 2017, Dhar 2016, Lee et al. 2020, Luaces and Karimpour 2018) and some argue that safety can be improved through route optimization (Cheng et al. 2019, Lee et al. 2020). Safety is also used in arguments for increased information access and sharing (Isenor et al. 2017 Pezzani and Heller 2019)

Business insight
Two articles touch on possible business opportunities generated by Big Data. Xiao et al. 2020 uses Big data to analyse oil transport, and Aiello, Giallanza, and Mascarella 2020 mentions that Big Data is a key factor to generate value in the increasingly digital world.

7.6 Challenges of Big Data

Figure 21 illustrates a word frequency query on the Challenges code.

Figure 21: Challenges Word Cloud

Volume

The large amount of data can pose challenges in handling, storing and processing it. Some of the examples were found in the literature.

Lytra et al. Jun 2017 describe the large data-sets used for Big Data applications, such as NASA environmental reports and AIS data. It mentions the lack of capabilities within existing data management and analytics techniques.

Dhar 2016 explore navigation using AIS, and describe the issue of working with and storing the large amount of data generated by AIS. The describe the need for distilling methods for AIS data.

Isenor et al. 2017 state that the large amount of data generated by AIS can put considerable burden on existing systems and maintainers.

When discussing volume, some articles suggest specialized platforms and cloud solutions for handling the data Fernández et al. 2018, Lytra et al. Jun 2017, Decision Support System (Lee et al. 2018, Venskus et al. 2019), and AI and Neural Networks have been suggested as ways to solve the problem of data variety (Gao and Shi 2019).

Variety

There were found different instances where data variety was mentioned as a challenge due to existence of different systems, sensors and protocols as well as sources.

Lytra et al. Jun 2017 describes the challenge of working with data collected from many different systems, which can require different protocols, different compression and decompression methods. Data can also have different fundamental structures, with some data being in the form of weather and voyage data, and some being rich-text incident and inspection reports. Some data is generated internally, while other data is collected from buoy and third-party sources.

Lee et al. 2018 focus on developing a model for fuel consumption. The article describe the challenge of creating such a model, due to the variety of factors and data needed to accurately describe the conditions of a vessel. Different vessel routes have different weather, similar vessels can have different responses to the weather. Different data sources will have different data formats, and data-handling systems need the capability to parse any number of formats and protocols.

Wang et al. 2020 explores technologies needed for unmanned vessels, and describes the challenge in working with different types of sensor/systems which can have redundant, conflicting, or missing information.

Venskus et al. 2019 describes the need for data generalization when gathering information from several systems needed to create Big Data applications.

Veracity
Low veracity can influence the users ability to analyse the data and create models accurately depicting reality.

Adland, Jia, and Strandenes 2017 states that AIS is difficult to use in Big Data application due to time-varying coverage and quality issues. The article suggest comparing AIS data with port and cargo data to verify the data.

Lytra et al. Jun 2017 describes how different degrees of accuracy and uncertainty in sensors results in missing, fractured, inaccurate data. Proposes data filtering and validation to be build into data handling platforms.

Value
The value in Big Data does not emerge from the collection of data, but from the information extracted and successful utilization of the data.

Aiello, Giallanza, and Mascarella 2020 mentions Big Data is the future for many companies who want to stay competitive, but there are potential difficulties in utilizing existing business plan methods and considerations when working with Big Data. Big Data is the future for many companies who want to stay competitive

Zhang and Lam 2019 mentions the top barrier for adopting Big Data being managerial- and culture-based. There is a lack of managerial sponsorship and understanding towards Big Data implementation.

8 Interview Findings

Many participants' involvement in the interviews depended on confidentiality. Instead of making some responses confidential and some not, the choice is made to address all responses as equally confidential. No identifying names or companies are used to describe the findings. The findings are aggregated over multiple responses. This might lower the validity of the findings from a reader's perspective, but this confidentiality was necessary to ensure a broader range of participants.

Iterative interviews

As with the nature of qualitative data collection through interviews, and per the project methodology, there needs to be an iterative process of adapting the interview questions depending on feedback or responses from participants. This iterative process sometimes includes restructuring the interview guide from one interview to another. The original interview guide can be seen in section 6.3.1.

Sampling

The original sample group consisted of participants representing different parts of the maritime tanker industry, see section 6.3.3. The focus was put on achieving the perspectives of different business segments within the ship owners/operators area. This results from early category development through literature coding, identifying the following list of business segments to be most relevant.

Participants

- Commercial
- Operations
- Business Intelligence

Several interview requests were sent out to people within these categories.

During coding and evaluation of interview responses, it became apparent that the perspectives of OEMs and decision support tools were needed to fully understand the state of Big Data utilization within the maritime tanker industry. As such, they were contacted for interviews or surveys. The list of cases and participants was expanded to the one seen in the methodology chapter 6.3.3.

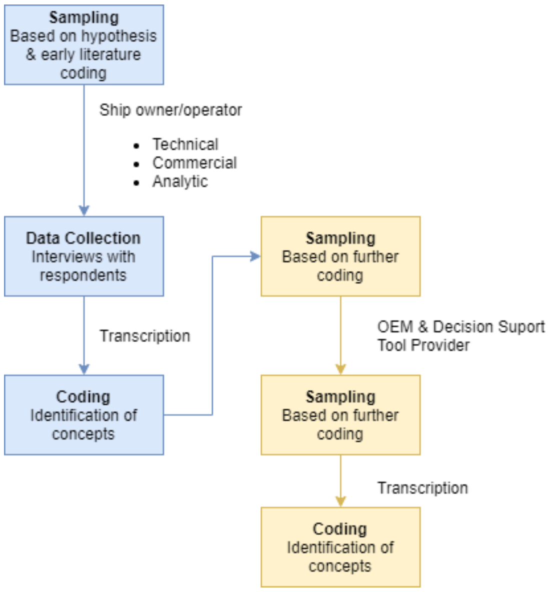


Figure 22: Iterative Interview process

Data collection

The data collection process evolved between each interview depending on participants' ability to answer questions, feedback, quality of the answers, and knowledge gaps identified. Initially, the interview guide had the participants state their understanding of the Big Data concept before introducing the problem statement and research. Expectations were that this approach helped reduce bias, but instead, it has been found that the participants had difficulty getting into the mindset required for the questions. By changing the interview guide first to introduce the research being done, the participant had an easier time articulating their later answers. While this might have introduced a degree of bias into the later interviews, the quality of answers went up as well. For example, questions regarding the current state use of Big Data platforms and storage were added to the guide, and the wording of questions changed depending on participants' understanding or knowledge of the topics.

The findings were divided into several cases. Case 1.1 represents a tanker owner/-operator who has not yet gathered or utilized Big Data. Case 1.2 is a significant tanker company that has begun using operational Big Data. Case 2 is a Decision Support Tool developer. Case 3 is a leading original equipment manufacturer (OEM).

8.1 Definition of Big Data

Depending on the role an interviewee had, the definition varied significantly, mainly focusing on one or two of the 5Vs of Big Data. Generally, interviewees identify

Big Data as vast amounts of data, which can be used in decision making and business planning with a combination of different advanced analytical techniques like Machine Learning and Artificial Intelligence. Such data is combined with multiple data sources for finding correlations between them, which cannot be found by working with data with the help of traditional analytical tools such as Power BI or Excel while looking at the industry from a high enough perspective.

	Definition
Case 1.1	Data is only good if used Combination of data sets identifying new relationships Used for decision making and strategy forming Data beyond BI capabilities
Case 1.2	Automated data processing Data velocity beyond human capabilities
Case 2	Data used in combination with ML/AI Data of a certain quality and purpose
Case 3	Data used for customer and equipment insight[c]@l@

Table 6: Definition

8.2 Use of Big Data

According to the interviews, Big Data is currently underutilized by major maritime actors. There are some attempts to do predictive maintenance instead of maintenance on run hours; however, it is yet not a fully developed practice. Some practices used hindcast methods on AIS data for future weather predictions to better plan voyages and tried to understand the impact of weather on ship's performance while validating noon reports. This also transitioned in some actors being able to improve the time charter equivalent (TCE) of the vessels using AIS data combined with commercial and weather datasets. Finally, some platforms are being developed, which utilize Big Data simulating voyages in the pool to choose the most appropriate vessel for a specific task. An OEM actively utilizes Big Data to get insight into customer purchasing patterns and equipment operational analysis from event logs.

8.3 Drivers of Big Data

	Drivers
Case 1.1	The next saving method Reduce human errors when report data Increase running hour granularity Identify fuel consumption causes Predictive maintenance Improve procurement Establishing a circular economy Environmental regulations Energy and route optimization Market changes means companies needs to adapt Improved new building Improved vessel management CO2 reduction New problems only solvable with Big Data
Case 1.2	Fuel optimization Regulations Compare different energy sources Predictive maintenance Insight into resource expenditure under different conditions
Case 2	Through data sharing platforms, it will be easier for other organizations to try Big Data Saving money and time through right maintenance scheduling CO2 reduction Improved vessel-harbor logistics
Case 3	Improving service and product Predictive maintenance Service improvement or software options increases sales potential

Table 7: Drivers

Most interviews recognized the benefits offered by Big Data analytics if implemented successfully. They would like to improve data granularity and quality and reduce the amount of human error when reporting the data. This would increase trust in reports and documentation, which is highly relevant when undertaking essential decisions.

Another critical driver for the participants was potential cost reduction concerning energy optimization and fuel consumption. Usage of Big Data analytic would provide an insight into energy expenditure under different weather conditions and fuel consumption causes. Additionally, this could provide a meaningful comparison between different energy sources (fuel types, for example) and their performance on the overall ship efficiency and expenditure under different conditions. Many also mentioned that Big Data could help design the "ultimate" new-build in the future by avoiding over-engineering of critical components. Additionally, participants would like to achieve predictive maintenance with the help of Big Data, which would help to target components for maintenance before the breakdown, saving them money and time.

Fuel consumption reduction also brings another critical mentioned driver - the industry's environmental regulations. Industry imposes regulations targeting CO2 emissions from vessels. Interviewees believe that Big Data could play an important role by improving navigation and choosing optimal speed for voyages.

Another driver was devoted to market development and competitors' actions. There might be some insights overseen currently which can be found only with the help of advanced analytics. Some current services might go extinct, hindering current business models and making it more attractive to invest in Big Data. Availability of different platforms from companies outside the Maritime industry, where AI can be applied to datasets, also makes it easier for companies to dive into Big Data. Further increase in fuel price can motivate Big Data projects even more. Some interviewees also mentioned that their companies might pursue Big Data analytics if the competitors will first initiate this "race."

According to the participants, Big Data can also enhance voyage optimization. This would allow optimizing logistics by improving vessel and port selection, and the choice of the most optimal route would be eased with more accurate navigation.

An OEM participant described the sales potential of offering more software-related products and services. The production cost of an already developed software solution is much cheaper than the production cost of physical equipment.

Finally, interviewers are interested in Big Data from the perspective of economics. Some believe Big Data can play an essential role in predicting cargo traceability or in another important concept - Circular Economic.

8.4 Challenges of Big Data

The interview guide tried to touch on all of the 5Vs of Big Data if they did not delve into the topics themselves. As a result, all interviews touched on all five main challenges as defined in the 5Vs, while also expanding it with Vision. However, there is a difference in the importance, urgency, and difficulty as evaluated by the participants. Below listing is an outtake from the coding matrix, showcasing the coding results concerning challenges. Many participants mention the desire and interest to work with Big Data, but many difficulties need to be solved.

	Challenges
Case 1.1	Data uncertainty due to maintenance history, environment, and wear Lack of data sharing options between organizations Sensor errors Long and uncertain realization time for Big Data project Uncertain business plan Lack of talent Risk of losing investments at vessel sale Tanker vessels rarely maintain steady ownership/management >3years Lack of industry standard for sensors and protocols Low data availability
Case 1.2	Difficult to measure the effect of predictive maintenance work Crew can make pro-active maintenance or changes to systems, impacting measurements Data uncertainty due to sensor drifting or other malfunctions Requires collaboration from OEM's, which can be difficult, time-consuming and expensive Investments reliant on market interests Risks loss of investments when selling/scrapping vessels
Case 2	High reliance on equipment quality Cooperation between platforms and different hardware vendors Requires clear vision and strategy before collecting data Need industry standards or certification body for Big Data Need backward compatibility between sensors and systems
Case 3	Difficult to predict human behaviour Getting data to shore Many start ups offer different systems People add bias to their analysis

Table 8: Challenges

The Volume and Velocity of the data were seen as the least potential constraints to working with Big Data. Especially regarding data transmission. Ship connectivity, while still occasionally lacking depending on area of operation and coverage. Smart data transmission systems has the capability to buffer messages until a connection is established. Regarding handling of the amount and frequency of data characterizing Big Data, some participants which were already utilizing Big Data, said they relied on third-party specialists, and were not worried. Another participant comments on the need for scalable solutions. But with the advent of cloud solutions such as Microsoft Azure, and Amazon Web Services it is also becoming less of a constraint. One company not yet utilizing or collecting Big Data did express their concerns about handling the data.

Variety is a constraint as expressed by one company not yet utilizing Big Data and one Decision Support tool provider. It is needed to communicate with many different systems supplied by different manufactures in order to get a thorough understanding of the vessels operational performance. These manufactures will have different attitudes towards giving access to the data, and can have a high variety of protocols for reaching the data. Different vessels will have different equipment and systems, and it will vary what data is available and in what format. Even if the vessels have identical systems, the available data channels might vary due to changes in requirements. This can make mapping data availability difficult. It requires the collaboration of multiple different companies. One respondent from a system provider describes the work involved with identifying and tagging available data, accessing said data, and enabling the back-end to work with this data seamlessly to be a big and urgent issue. Some respondents with a background in analytic and technical expressed the desire for standards and certification for sensors, as to ensure backwards compatibility and ease of implementation. The data collection systems and platforms need the capabilities to work with data from many different sources seamlessly. The variety of available tools on the market complicates it more to standardize the IT architecture from the industry perspective.

A large concern was the participants ability to **create Value** from the data collected. They mentioned the large volume, velocity, and variety needed over a large enough time-series to create useful models. Additionally, it can be difficult to measure the effect of predictive maintenance. They argued the need for data sharing options between organizations, as to both aid in developing accurate models, and to lower the hurdle for working with Big Data, or offer their services in analytic. Some argued that data can be used with a bias or a wrong intent to present the findings in a better light. A lack of talent within the industry increases the difficulty. There need to be a mix of maritime and data science knowledge together to properly work with the vessels and the extracted data. Finally, sometimes the data and applied analytics cannot explain the human behaviour which drives their commercial activities.

The Veracity of the the data is seen as one of the biggest challenges for working with Big Data. Uncertainty in the data can arrive in multiple forms. Vessels with similar equipment or baseline performance can have large variations in performance due to maintenance history, environment, wear on equipment and hull. Crews might make changes to the equipment as part of maintenance, but not log or communicate the work properly, and thereby introduce changes in the data which can be difficult to track down. Sensors might drift, machinery might malfunction and computers can make errors. Sometimes even small updates to used software can prevent the data from being collected successfully. In many other settings, this can be fixed by having specialists available to troubleshoot the issue, but for vessels, access to the equipment can be both difficult and unpredictable depending on voyage schedule and the markets.

The Vision of working with Big Data is the biggest issue as described by the participants. It is difficult to create business plans and commit resources for vessels which might be sold at any time. This creates a constant risk for loss of investments in both time, assets and money, making it difficult to convince management of such projects. According to some responses, there needs to be a clear vision and strategy for how the Big Data is going to be utilized. This is difficult due to the longitudinal nature of Big Data projects. It is difficult to create value from the data, as it the models are hard to test in an environment where the assets might have left the management of the organization before enough data is gathered to create an image of the operational performance.

8.5 Current State

Currently, big maritime players express awareness and interest in Big Data but are not utilizing it or are at lower levels of utilization maturity. Performance analysis is done with the help of noon reports filled out manually and then validated. The noon reports themselves at the current state cannot be classified as a Big Data source. The reason is that the results are presented as an average of the whole day, with significant quality issues, which limits the velocity and veracity part of Big Data. There are attempts to replace noon reports with the help of data loggers, allowing for higher data validity and velocity. However, right now, such projects are experiencing challenges mentioned previously, making successful implementation impossible.

The maintenance period is decided based on running hours or whenever a component is breaking down. It is the easiest way to do maintenance, which results in the loss of time and money. Systems are being optimized, and buffers are placed based on a feeling and experience rather than historical data. There are different technologies available on the market developed by such giants as Amazon, Google which offer the possibility to apply AI to the company's data.

Current more prominent Maritime players are developing, and re-adjusting internal applications meant to provide better planning of and improved overview over operations, which proves to be not an easy task. In some instances, interest is mainly directed to the commercial aspect of predictions rather than technical. All the system data is in the cloud. Different departments are recognizing the number of required investments as well as such investments have to be made beforehand. Consequently, they recognize the importance of transmitting this message to the upper management. However, few departments were found where extra resources were dedicated to it, which allowed hiring a dedicated team to do such analytics. These departments are the ones that managed to convey the importance of Big Data analytics to the upper management. Those who have not, whenever such analytics are required, outsource it from third parties. In addition to that, many interviewees recognize the importance of different industry actors collaborating. They believe many challenges could be solved, and a lower amount of resources would be required to get enough data.

Some participants who work with BI believe that Big Data has not much relevance in the Tanker segment instead of the container one. Data is mainly used to act as a ”window into business” for their customers.

9 Comparison of Findings

The following section compares the findings as described in section 7, Literature Findings and section 8, Interview Findings. By comparing the findings from the two methods, an understanding of the alignments and miss-alignments between industry and academics is formed. First (1), industry actors’ definitions of Big Data are compared with literature definitions to establish a baseline. The remainder of the section is as follows: (2) the Big Data usage described in the literature is matched against the current usage of Big Data as found through interviews. (3) drivers are compared on the same premise as the use, (4) and challenges.

9.1 Definition of Big Data

It can be concluded that respondents have a good representation of what Big Data is compared to how literature defines it. All of them mentioned that Big Data is manifested in big volumes of data that needs to be stored. Depending on the role in an organization, respondents focused on different Vs of Big Data. Some of them were aligned with literature and identified that Big Data is about merging different data together in order to find correlations between variables leading to important business value discovered. Others with a more technical background added that advanced analytical techniques like Machine Learning or Artificial Intelligence are needed to respond accordingly to high velocity and volume of data. There were few respondents that addressed Power BI and Excel to tools which they use when working with Big Data. However, the majority of other respondents as well as the literature identifies such data not as Big Data, because Big Data is “which traditional processing techniques or algorithms are unable to operate on Mirović, Miličević, and Obradović 2018”. Instead of traditional BI tools and Excel, literature describes many new platforms being developed to help companies operate with Big Data Fernández et al. 2018, Fernández et al. 2016, Luaces and Karimpour 2018. None also addressed the requirement of suitable hardware or computing power to perform such analytics. Therefore, it can be observed that sometimes a line between Big Data and data of a big volume may seem blurry to some participants.

9.2 Use of Big Data

This section compares how Big Data is used in the literature with its use in the industry as well as the current state findings.

It has been confirmed previously that Big Data is underutilized in interviewed organizations. Currently, performance analysis is done with the help of noon reports and inspection reports, which themselves have a low granularity. There were different attempts to implement dataloggers, but it proved to be difficult on a larger scale. Maintenance is therefore done when a component is broken down, causing increased lead times and costs. Therefore, Big Data can offer a great opportunity to implement predictive maintenance practices. However, there were found no cases in the literature of how such can be done. The closest to this could be considered PSC cases, where resources where devoted to predicting detentions at ports and increasing inspection efficiencies.

Some respondents expressed their believes that Big Data is currently used to improve designs of the future ships. However, this exact perspective was not found in the literature. Sanchez-Gonzalez et al. 2019 describes attempts of designing controlling systems for unmanned vessels, which would improve navigation by enhancing transportation safety. And Jeong, Woo, and Park 2020 uses Machine Learning to predict lead time required during fabrication of different ship components.

Currently, some actors managed to predict weather using AIS data and plan voyages better, which matches literature state of Big Data. In this example, industry and literature match in designing platforms for simulating voyages using Big Data.

Another mismatch lies in usage of commercial perspective, which comes with Big Data. As opposed to the industry's interest, there were found few papers focusing on utilizing Big Data to improve or investigate commercial side. Adland, Jia, and Strandenes 2017 uses AIS-derived data to estimate accuracy of official customs data, whilst Jia, Prakash, and Smith 2019 estimates vessels' payload based on AIS data to assess productivity and demand conditions in the shipping markets. Respondents expressed interest and are currently attempting to use Big Data to predict or explain certain buying patterns of their clients. However, examples of such attempts where limited to Power BI utilization and dashboard creation, which is not as advanced as Big Data utilization could be according to the literature.

9.3 Drivers of Big Data

There were many instances literature mentioned different drivers which was previously summarized into three categories: environment, cost, safety and business insight. It can be observed that current maritime actors are interested in reducing fuel consumption, which constitutes of around 50-60% of total operating costs (morethanshipping.com 2021). Reducing the fuel consumption would also help fulfilling industry's regulations regarding environment. They were also interested in identifying fuel consumption causes and comparing different energy sources against ship's performance. However, such comparison between different fuels or other energy sources has not been found in literature. Another practice observed in literature was improved navigation which would lead to route optimization and consequently to decreased fuel consumption. Additionally, improved navigation can also lead to increased safety of the maritime environment. This was also mentioned to be an interesting topic for industry actors as they expressed the desire to optimize energy and fuel required for different routes, when they referred to Big Data as "the next saving method".

Another indirect driver mentioned by respondents is improved data veracity through introducing data capturing means when implementing Big Data at an organization. However, this was not depicted as direct driver according to literature and in fact, veracity is considered to be one of the Big Data challenges. Therefore, it is considered that this driver is relating more to introducing data-capturing techniques where they are not already present, rather than driver for implementing Big Data exactly. Implementing such techniques would increase data granularity which is preferable according to industry actors.

One of the biggest drivers for industry actors mentioned in most interviews was predictive maintenance – ability to predict when a certain component will break down and tailor it before it ever happened. Additionally, similar data could be used to design newbuilds. However, such drivers were not mentioned in the found literature.

Participants also mentioned that Big Data is intriguing because it can help companies find hidden business information which would help gaining a competitive advantage or optimizing overall business. Such drivers were found in the literature, showing that Big Data indeed can help unveil something hidden to human eye.

During the interviews, there were also mentioned several drivers for implementing Big Data which relate to potential future benefits, such as use of Big Data during Circular Economy process or creation of an anonymous data sharing platform which would allow third parties run analytics on datasets. Even though many papers discussed creation of platforms, they have not mentioned how creation of such platforms would motivate analysts outside the company or industry to sell their services.

9.4 Challenges of Big Data

For ease of comparison, the findings will continue to be grouped according to the 5Vs. The 5Vs (including Vision) are meant to encompass all the potential challenges and are the primary method for categorization. The categorization of a challenge is not meant to be a definitive label but rather a suggestion for comparison. A challenge can also be multi-faceted and have aspects of multiple of the 5Vs.

Volume

As to reiterate, volume refers to the vast amount of data associated with Big Data. In literature, volume is given considerable focus, especially regarding AIS data and weather data. The volume of data is described as challenging to handle with current technology. Different distilling algorithms were applied to narrow down the datasets when analyzing the data without impacting their value. Additionally, different third-party platforms ease the work experience by handling data processing and visualization. Interviews showed data volume to be of little concern when it comes to storing the data. This is attributed to the widespread adoption of scalable cloud solutions.

Velocity
This is the speed at which new data is generated and needs to be processed and analyzed. Velocity can sometimes be related to volume. The speed at which the data is generated (AIS, weather) influences the volume of the data. Despite this relation, velocity is not mentioned in literature as a relevant challenge. Some articles describe velocity only as part of the Vs. when establishing Big Data definition, without it being a significant challenge.

Interviews showed high data velocity to be less of a concern than too low data velocity. While on-board connectivity and satellite coverage are still an issue, existing technology already uses smart connectivity and buffers, and data will be sent when a connection is re-established. Industry concern is on generating enough valid data to create models and use for AI and ML. The time it takes to generate enough data to create usable model results in a higher barrier of entry.

Variety
Variety is the differences between data sources and the data itself, resulting in different data handling and processing methods being needed. Variety is a challenge given much attention in the literature. Variety is most often described as caused by data originating from different systems, which transmits data in different formats over different protocols with different frequencies. Variety is also described as originating from environmental conditions by introducing additional data which needs to be collected and analyzed. The variety of data needed to evaluate a vessel can make it an infeasible challenge. The variety causes a challenge when creating systems for handling the data or when creating decision-support tools for using models.

Interview responses regarding data variety described it as a significant challenge. The variety is described as coming from differences in systems, manufacturers, specifications, and maintenance. This results in a large amount of work needed to collect, process, verify this data. It also requires collaboration with OEMs who might not be interested in sharing the data. This dependency on OEMs and data handling platforms raises the challenge of data ownership. The vessel owner and user of the data want to collect data from multiple systems, but the manufacturers of these systems might believe the data generated by their equipment to be their property and might not want the data to be shared or used by competitors. Sharing the data can also reduce business opportunities for the manufacturer to capitalize on the data themselves. There is a lack of standards and collaboration between organizations regarding Big Data in the industry, resulting in fragmented, case-specific designs and solutions.

Veracity
The veracity of data can be described as the quality and reliability of said data. In literature, veracity is related to the time-varying coverage and quality of AIS data. This creates difficulty with missing or incorrect data impacting models and vessel mapping tools. Concerning autonomous vessels, a large amount of different data is needed to monitor the vessels. Varying sensor quality and accuracy can cause low veracity of said data.

In Interviews, data veracity is approached differently depending on the case. Shipowners described data and equipment verification as a significant challenge due to the assets moving around at sea without qualified personnel on board. The difference between vessel response to weather and other environmental factors also makes it challenging to notice sensor drifting. As new equipment is developed and new features are added are changed on equipment, data collection systems may become obsolete due to new data formats or types. In the case of an OEM, they found no challenge in ensuring data veracity. They attribute this to having designed the tolerances and controlling the data naming schema.

Value
Value is using the created Big Data to generated business advantages. In literature, value can be an abstract concept, as the purpose of the research can be the value-adding aspect. The challenge addressed in the research can be said to aim to solve a value-adding challenge. Few articles address the difficulty in realizing value from Big Data, but the ones that do, attribute the most significant barriers to be managerial and cultural. New ways of forming business plans are needed, as traditional methods are not applicable in the changing digital climate.

Creating value from Big Data is a repeating challenge across the interviews. One of the concerns regards the asset uncertainty in the Tanker industry. Tanker vessels are rarely under the management or ownership of one operator for more than three years. This reduces the time available to gather data from a vessel and create models. By the time a vessel is equipped with data collection equipment and a performance model is created, the vessel may have been sold, resulting in an investment loss. The realization time for Big Data projects is long and uncertain, and no guidance exists for how to conduct such projects effectively. Another concern is the lack of talent for working with Big Data. It is described that particular insight into the maritime industry is needed to make sense of the data. To effectively use the collected data, a plan is needed beforehand to ensure a working data structure, as traditional business plan formats do not work with Big Data.

Vision
Vision is the challenge of communicating the need for Big Data, collecting talented

people, and commitment towards solving the challenges. A few articles mention the need for substantial organizational restructuring to create value-form Big Data but do not explore the potential solution. There is a lack of managerial understanding and sponsorship towards Big Data utilization within the industry.

The interviews revealed a deep interest in Big Data among the cases, but investment into such projects is also described as fickle and depending on the market state. Industry stakeholders appear to agree, Big Data is the following way to create value for customers, but the rigid nature of tanker companies and OEMs makes it challenging to change strategies. Currently, Big Data does not appear to be used in decision-making processes.

10 Discussion

This sections brings the knowledge generated from the previous parts together to answer the research questions and introduce relevant discussion points to the topics established.

There has been conducted an analysis of recent literature reviews to justify a need of a new one. It has been concluded that another review should try to improve the flaws of the previous ones. Munim et al. 2020 pointed out at the lack of discussion around "softer" topics of drivers and challenges associated whilst working with Big Data. To address this issue, there were found several articles focusing heavily on challenges (Dhar 2016, Kacprzyk 2016, Zaman et al. 2017, Kallimani 2018, Alop 2019, Li et al. 2018, Huang et al. 2015) as well as papers discussing opportunities or Big Data drivers (Zaman et al. 2017, Kenyon et al. 2018). In addition to that, all the other mentioned papers were analyzed and coded by "Challenges" and "Drivers" and the industry's perspective was introduced through semi-structured interviews.

Opposed to Aiello, Giallanza, and Mascarella 2020, a more structured and transparent methodology was developed to guide the reader throughout the used approach. In addition to that, the current review tries to identify the most dominant applications of Big Data, whilst also identifying the underrepresented ones and test if Big Data is currently utilized in industry's decision making process.

By conducting this research now, four years has passed since Fruth and Teuteberg 2017 conducted their structured literature review. Throughout this time Big Data has achieved more attention from the Maritime actors and consequently was utilized more, allowing for a richer literature representation. This made it possible for current research to assess better Big Data applications as well as continue the work of Yang et al. 2019a in exploring AIS usage in conjunction with emission reduction and navigation improvements.

Finally, by conducting this research through the means of literature review and industry interviews, we further elaborate on the findings of Alan 2018 related to the challenges of Big Data imposed on the current business models in the Industry.

Whilst being inspired by previous well-implemented practices and trying to improve made flaws, current use of Big Data, drivers and associated challenges were outlined.

10.1 Use of Big Data

There has been identified a clear mismatch between the uses of Big Data at current major maritime organizations and its literature representation. From the literature perspective, a lot of attention is addressed towards technical optimizations, concerning such topics as Navigation, Fuel consumption, Optimization of operations and information flow at Ports. There were also identified underrepresented topics as uses of Big Data in shipbuilding and commercial directions. However, there were found no sources discussing a very interesting topic for current maritime actors - predictive maintenance. Being able to switch from traditional maintenance-by-running-hours approach would save time and money, which could be invested elsewhere. Overall level of Big Data use in Maritime organizations is considered to be low. This is mostly because actors are afraid of performing a "leap of faith" and dedicating resources into such projects. This results in the lack of qualified personnel, data capturing technologies, computing hardware etc. Consequently, Big Data techniques remain at the initial level of implementation and actors rely on traditional performance evaluation techniques and decision making tools.

As was discussed during the interviews, this is mostly due to the lack of evidence presented in the literature. Stakeholders are reluctant to invest in something without a clear evidence of how much time it would take to generate profit. The reason for this is that many found studies do not take industry's perspective into consideration. Instead, they try to investigate something which has already been touched upon, and does not generate new knowledge areas. This limits the spectrum of investigation and generation of new perspectives in the literature.

10.2 Drivers of Big Data

As with the nature of any business and industry, all actions of the company should aim towards generating value. The essence of the identified drivers also relates to a companies ability to create value. There has been identified a driver in increasing safety and meeting industry's requirements. The safety of a vessel is paramount to its continued operations, and the avoidance of accidents is needed to not only reduce potential downtime, but also to maintain trust from customers. Additionally, with the increasing pressure from regulatory bodies towards environmental efficiency, such as IMO's EEXI, DCS, sulfur caps, and zero-emission goal, there is a pressure to find new and innovative ways to decrease pollution. Through AIS data, increased navigational safety has been proven to be a feasible objective as well as found its practices in reducing CO2 emissions.

From interviews and literature, we learn that emission reduction is often considered a by-product of fuel expenditure reduction. Big Data is described as the next cost-saving method to be explored due to the large different elements impacting a vessels fuel expenditure. Big Data, ML, and AI is seen as a promising combination for achieving these goals. Though many found sources do not go beyond theory crafting and limit themselves by implementing said theory on a smaller scale.

Another driver is manifested in increased business insight. Albeit a scarcely explored topic in literature, research have proven that Big Data can be used to gain insight into trading patterns and cargo of vessels globally. One business opportunity that has not been explored in literature , but seen in interview findings, is the development of new marketable services in the form of equipment, data, market insight. The cost of scalability in software compared to physical assets makes it a very attractive option.

Another driver seen in interview findings, but not in literature, is predictive maintenance. Through predictive maintenance, large amount of value could be gained by accurately timing maintenance and ordering of spares. Breakdowns could be easier avoided and untimely stops in operation due to planned maintenance by running hours could be made smarter.

10.3 Challenges & Solutions

As described in interview findings, the biggest challenge is the difficulty of creating and communicating value from Big Data projects. One other article that focuses on identifying maritime Big Data challenges also mentions managerial understanding and investment as the biggest obstacle. No other article explores the concept in any format and there is one article that only identifies the issue. Traditional business case formats do not apply for Big Data projects, and the rigid, risk-averse nature of the Maritime Industry is in opposition to the long and uncertain realization periods of Big Data projects. This results in the lack of dedicated teams exploring Big Data opportunities in the interviewed major Maritime organizations. In literature, there is a severe lack of case studies exploring the process of seeing monetary rewards from Big Data projects. The tanker segment receives very little consideration within the maritime industry, despite facing several unique and complex challenges. Due to the homogeneous nature of the tanker cargo, less commercial data is naturally generated, limiting the applicability of Big Data analytics. Additionally, due to the nature of the industry, frequency, and severity of inspections, it is rare for assets to stay under the same management for more than three years. The relatively short and uncertain asset ownership periods limit the time frame available for Big Data projects. The sale of a vessel can mean a loss of investment in data collection equipment and work hours. Some stakeholders have expressed the need for open data sharing platforms to make it easier to create the necessary amount of information for models, AI, and ML. A new issue with the idea of data sharing is data variety.

Literature findings and interview findings emphasize the challenge of data variety, and both attribute the source of variety to the need to communicate with different systems. These systems encompass all sources of data, from weather and wave conditions to AIS, to machinery, to sensors of many types. A vessel and its related systems are so complex and condition-dependent that no single data source can be used to establish an accurate model of overall vessel performance. Some of the reviewed articles suggest smart data platforms for congregating the data from different sources, and such platforms are also being developed and tested actively in the industry. But such solutions still have the problem of getting the data into the platform, interpreting the data, and even accessing the data from the equipment/systems. The issue of accessing the data creates a dependency on the OEMs in relation to the potential users of the data. This is an issue that is not seen in literature, despite being an active complication for those trying to access the data.

A related issue discussed in interview findings is the lack of standardization and lack of guaranteed compatibility in the industry for using Big Data. Other areas of the maritime industry are already heavily defined by standardization, i.e., IEC 61162, which standardizes digital interfaces, including protocols such as NMEA 0183 and NMEA 2000. Standardization of naming conventions could make it simpler for all parties to design scalable and versatile solutions for accessing and collecting data. These observations are in line with statements from the classification society ClassNK who has published a roadmap for utilizing Big Data. Part of this roadmap touches on ISO 19847, which is a proposed ISO standard for shipboard data servers to share field data at sea. This shows an awareness of the challenge across the industry and is a significantly unsaturated area within the literature. Future research should focus on the viability of possible data structures applicable from on-board systems to on-shore utilization.

When collecting, storing, sharing, and using data from different systems, through the different platforms and over different organizations, the question of data ownership arises. The vessel manager might infer that they own the data generated by their asset, but the OEM might infer that the data would not have been generated and captured without their equipment. The question only becomes more complicated when adding third-party data analysis platforms and solutions. We see multiple solutions appearing, advertising insight into specific systems on vessels, capitalizing on the data generated by their equipment. From the perspective of the individual stakeholders, it makes sense not letting competitors to access their data and lose potential earnings from customers paying to access the data. The concern is, this approach might further fragment data availability, with data only available within closed eco-systems.

The proposed solutions to the mentioned obstacles are multi-faceted.

1. Future research needs to examine ways of creating monetary value from Big Data projects, and new ways of forming business plans. This would bring facts used in arguing for implementing Big Data projects.

2. Big Data needs to be a consideration as early as the ship-design process, with the OEMs involved from the beginning. The OEMs have the capability and insight into the monitoring equipment to ensure data veracity. This would also allow enough data to be generated before the vessel changes its ownership as well as speed up the ROI process.

3. Data ownership needs to be incorporated into asset ownership. Methods and standards for describing vessels in the form of digital twins needs be formed at the ISO and IMO level. We suggest digital twins to be created and verified as part of the new-building process, and control of said digital twin and corresponding historical data can be transferred seamlessly at physical asset transfer. This would also increase effectiveness of efforts into achieving a circular economy and cradle-to-grave analyses if the life of the asset is available through digital twins.

11 Conclusion

This research aimed to explore the application, drivers and challenges related to Big Data in the maritime tanker industry. This was done through a semi-structured literature review, combined with interviews with industry stakeholders. The research was inspired by successful existing practices of other researches about Big Data in Maritime industry, whilst improving identified flaws and bringing its own industry's perspective. The addition of the stakeholder perspective allowed the project to approach the problem differently than existing researches, and create a context to the challenges not yet explored.

The research questions identified in the beginning were answered. All three answers are based on the findings from the mixed method of the literature review and the interviews.

Q1
The current level of Big Data utilization in the industry is not found to be at a level where the industry can properly take advantage of findings mentioned in its literature counterpart. Literature placed a high focus on realizing Big Data improvements from the technological aspects, like improving navigation, optimizing fuel and energy consumption, improving operations and information flow at ports. Besides these biggest groups, there were also identified smaller ones. However, Big Data is not currently utilized efficiently at the interviewed organizations. There were several attempts to implement advanced analytics, such as predictive maintenance, however, these are not fully utilized.

Q2
Several drivers are identified, both in literature and in practice. These industry drivers were found to be under-explored in literature, inviting future research into these topics. In particular, the literature does not place enough of emphasis on economic profitability of such projects. This discourages major actors from exploring advanced analytics associated with Big Data and prevents further resource dedication.

Q3
Practical challenges are identified and considered. It is found that while some of the challenges described in literature are relevant in practice, the majority of research focuses on solving challenges that the industry find either trivial or unrelated to their real concerns. A solution is suggested for both future research and industry efforts to strive towards.

11.1 Limitations

While attempts were made to gain the perspective of a broad range of industry stakeholders, the range could be further increased to gain a more accurate understanding of the interconnected problems across different sectors.

We were also unable to gain usable responses to the distributed questionnaires, which could have added an additional quantitative aspect the analysis.

References

Adland, Roar, Haiying Jia, and Siri P. Strandenes (2017). "Are AIS-based trade volume estimates reliable? The case of crude oil exports". English. In: *Maritime policy and management* 44.5, pp. 657–665. DOI: 10.1080/03088839.2017.1309470. URL: http://www.tandfonline.com/doi/abs/10.1080/03088839.2017.1309470.

Aiello, Giuseppe, Antonio Giallanza, and Giuseppe Mascarella (2020). "Towards Shipping 4.0. A preliminary gap analysis". English. In: *Procedia manufacturing* 42, pp. 24–29. DOI: 10.1016/j.promfg.2020.02.019. URL: http://dx.doi.org/10.1016/j.promfg.2020.02.019.

Alan, Carmen B. (2018). *The disruptive impact of future advanced ICTs on maritime transport: a systematic review*. DOI: 10.1108/SCM-03-2018-0133].

Alop, Anatoli (2019). "The Main Challenges and Barriers to the Successful "Smart Shipping"". English. In: *TransNav (Gdynia, Poland)* 13.3, pp. 521–528. DOI: 10.12716/1001.13.03.05. URL: https://explore.openaire.eu/search/publication?articleId=doajarticles::b275fd79ec1ce14bf4500c6de481151e.

Arifin, Ajib et al. (Dec 20, 2017). "IoT-Based Maritime Application". English. In: BDIOT2017. ACM, pp. 191–194. DOI: 10.1145/3175684.3175729. URL: http://dl.acm.org/citation.cfm?id=3175729.

Baumeister, Roy F. and Mark R. Leary (1997). "Writing Narrative Literature Reviews". In: *Review of General Psychology* 1.3, pp. 311–320. DOI: 10.1037/1089-2680.1.3.311. eprint: https://doi.org/10.1037/1089-2680.1.3.311. URL: https://doi.org/10.1037/1089-2680.1.3.311.

Bo, Yang and Yang Meifang (2021). "Construction of the knowledge service model of a port supply chain enterprise in a big data environment". English. In: *Neural computing& applications* 33.5, pp. 1699–1710. DOI: 10.1007/s00521-020-05044-w. URL: https://search.proquest.com/docview/2493885059.

Brouer, Berit, Christian Karsten, and David Pisinger (2018). "Optimization in liner shipping". English. In: *Annals of operations research* 271.1, pp. 205–236. DOI: 10.1007/s10479-018-3023-8. URL: https://search.proquest.com/docview/2092727136.

Bryman, A. (2016). *Social Research Methods*. Oxford University Press. ISBN: 9780199689453. URL: https://books.google.dk/books?id=N2zQCgAAQBAJ.

Cepeda, Maricruz A. et al. (2020). *Environmental Impact of Ship Emissions Based on AIS Big Data for the Port of Rio de Janeiro*.

Cheng, Liang et al. (2019). *Using big data to track marine oil transportation along the 21st-century Maritime Silk Road*. DOI: 10.1007/s11431-018-9335-1.

Cooper, H.M. (1985). *A Taxonomy of Literature Reviews*. National Institute of Education. URL: https://books.google.dk/books?id=zVT6nwEACAAJ.

Coraddu, Andrea et al. (2017). "Vessels fuel consumption forecast and trim optimisation: A data analytics perspective". English. In: *Ocean engineering* 130, pp. 351–370. DOI: 10.1016/j.oceaneng.2016.11.058. URL: http://dx.doi.org/10.1016/j.oceaneng.2016.11.058.

Cox, Michael and David Ellsworth (Jan. 1997). "Managing big data for scientific visualization". In:

Cui, Jianfeng (2018). "Dynamic Migration Algorithm of Marine Big Data in Cloud Computing Environment". English. In: *Journal of coastal research* 83.83, pp. 706–712. DOI: 10.2112/SI83-117.1. URL: https://www.jstor.org/stable/26543041.

Davis, J., K. Mengersen, and S. Bennett (2014). "Viewing systematic reviews and meta-analysis in social research through different lenses". In: DOI: https://doi.org/10.1186/2193-1801-3-511.

De Mauro, A., M. Greco, and M. Grimaldi (2014). "What is Big Data? A Consensual Definition and a Review of Key Research Topics". English. In: *International Conference on Integrated Information.*

Dhar, Samir K. (2016). *A Dissertation entitled Addressing Challenges with Big Data for Maritime Navigation: AIS Data within the Great Lakes System.*

Farah Al Kaderi, Rim Koulalim and Mohamed Rida (2020). "A Multi-layer System for Maritime Container Terminal Management Using Internet of Things and Big Data Technologies". In:

Fernández, Pablo et al. (2016). "SmartPort: A Platform for Sensor Data Monitoring in a Seaport Based on FIWARE". English. In: *Sensors (Basel, Switzerland)* 16.3, p. 417. DOI: 10.3390/s16030417. URL: https://www.ncbi.nlm.nih.gov/pubmed/27011192.

Fernández, Pablo et al. (2018). "3D-Monitoring Big Geo Data on a seaport infrastructure based on FIWARE". English. In: *Journal of geographical systems* 20.2, pp. 139–157. DOI: 10.1007/s10109-018-0269-2. URL: https://search.proquest.com/docview/2016874000.

Fiskin, Remzi, Erkan Cakir, and Coşkan Sevgili (2020). "Decision Tree and Logistic Regression Analysis to Explore Factors Contributing to Harbour Tugboat Accidents". English. In: *Journal of navigation* 74.1, pp. 1–26. DOI: 10.1017/S0373463320000363.

Fruth, Markus and Frank Teuteberg (2017). "Digitization in maritime logistics—What is there and what is missing?" In: *Cogent Business & Management* 4.1. Ed. by Shaofeng Liu, p. 1411066. DOI: 10.1080/23311975.2017.1411066. eprint: https://doi.org/10.1080/23311975.2017.1411066. URL: https://doi.org/10.1080/23311975.2017.1411066.

Fu, Junjie et al. (2020). "Mining ship deficiency correlations from historical port state control (PSC) inspection data". English. In: *PloS one* 15.2, e0229211. DOI: 10.1371/journal.pone.0229211. URL: https://www.ncbi.nlm.nih.gov/pubmed/32084200.

Fujino, Iwao, Christophe Claramunt, and Abdel-Ouahab Boudraa (Oct 27, 2018). "Extracting Courses of Vessels from AIS Data and Real-Time Warning Against Off-Course". English. In: ICBDR 2018. ACM, pp. 62–69. DOI: 10.1145/3291801.3291823. URL: http://dl.acm.org/citation.cfm?id=3291823.

Gandomi, A. and M. Haider (2015). ""Beyond the hype: Big data concepts, methods, and analytics"". English. In: *International Journal of Information Management.* URL: https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

Gantz, J. and D. Reinsel (2012). ""The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East"". English. In: *IDC IView.* URL: https://www.emc.com/collateral/analyst-reports/idc-the-digital-universein-2020.pdf.

Gao, Miao and Guo-You Shi (2019). *Ship Spatiotemporal Key Feature Point Online Extraction Based on AIS Multi-Sensor Data Using an Improved Sliding Window Algorithm.* DOI: 10.3390/s19122706.

Gupta, Dr. Uma G. (2016). "Vision: A Missing Key Dimension in the 5V Big Data Framework". English. In:

Han, Peng and Xiaoxia Yang (2020). "Big data-driven automatic generation of ship route planning in complex maritime environments". English. In: *Acta oceanologica Sinica* 39.8, pp. 113–120. DOI: 10.1007/s13131-020-1638-5.

Hu, Shiyan and Bei Yu (2020). *Big Data Analytics for Cyber-Physical Systems.* English. Cham: Springer International Publishing AG. ISBN: 9783030434939. URL: https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=6237876.

Hu, Zhi-Hua (2020). "Data-Driven Analytics for China's Overseas Construction Projects in the Contexts of the Maritime Silk Road and Global Maritime Network". English. In: *Complexity*

(New York, N.Y.) 2020, p. 1. DOI: 10.1155/2020/8129172. URL: https://dx.doi.org/10.1155/2020/8129172.

Huang, Dongmei et al. (2015). "Modeling and Analysis in Marine Big Data: Advances and Challenges". English. In: *Mathematical problems in engineering* 2015, pp. 1–13. DOI: 10.1155/2015/384742. URL: https://dx.doi.org/10.1155/2015/384742.

Isenor, Anthony W. et al. (2017). "MSARI: A Database for Large Volume Storage and Utilisation of Maritime Data". English. In: *Journal of navigation* 70.2, pp. 276–290. DOI: 10.1017/S0373463316000540. URL: https://dx.doi.org/10.1017/S0373463316000540.

Jeon, Miyeon et al. (2018). "Prediction of ship fuel consumption by using an artificial neural network". English. In: *Journal of mechanical science and technology* 32.12, pp. 5785–5796. DOI: 10.1007/s12206-018-1126-4. URL: https://search.proquest.com/docview/2480996585.

Jeong, Ju Hyeon, Jong Hun Woo, and JungGoo Park (2020). "Machine Learning Methodology for Management of Shipbuilding Master Data". English. In: *International journal of naval architecture and ocean engineering* 12, pp. 428–439. DOI: 10.1016/j.ijnaoe.2020.03.005. URL: http://dx.doi.org/10.1016/j.ijnaoe.2020.03.005.

Jia, Haiying, Vishnu Prakash, and Tristan Smith (2019). "Estimating vessel payloads in bulk shipping using AIS data". English. In: *International journal of shipping and transport logistics* 11.1, pp. 25–40. DOI: 10.1504/IJSTL.2019.096864. URL: https://www.inderscienceonline.com/doi/10.1504/IJSTL.2019.096864.

Jia, Haiying et al. (2017). "Norwegian port connectivity and its policy implications". English. In: *Maritime policy and management* 44.8, pp. 956–966. DOI: 10.1080/03088839.2017.1366080. URL: http://www.tandfonline.com/doi/abs/10.1080/03088839.2017.1366080.

Jović, Marija et al. (2019). "Big Data Management in Maritime Transport". English. In: *Journal of Maritime& Transportation Science* 57.1, pp. 123–141. DOI: 10.18048/2019.57.09.. URL: https://search.proquest.com/docview/2359321274.

Kacprzyk, Janus (2016). *Big Data Optimization: Recent Developments and Challenges.*

Kallimani, James G. (2018). "The Challenges of Digitisation and Data Analysis in the Maritime Domain". English. In: *Maritime affairs (New Delhi, India)* 14.1, pp. 36–50. DOI: 10.1080/09733159.2018.1478433. URL: http://www.tandfonline.com/doi/abs/10.1080/09733159.2018.1478433.

Kamolov, Ahmadhon and Suhyun Park (2019). "An IoT-Based Ship Berthing Method Using a Set of Ultrasonic Sensors". English. In: *Sensors (Basel, Switzerland)* 19.23, p. 5181. DOI: 10.3390/s19235181. URL: https://www.ncbi.nlm.nih.gov/pubmed/31779227.

Karvelis, Petros et al. (2020). "PortWeather: A Lightweight Onboard Solution for Real-Time Weather Prediction". English. In: *Sensors (Basel, Switzerland)* 20.11, p. 3181. DOI: 10.3390/s20113181. URL: https://search.proquest.com/docview/2410053172.

Kenyon, G. N. et al. (2018). "Improving the return on investment in ports: opportunities in data management". English. In: *Maritime economics& logistics* 20.4, pp. 514–530. DOI: 10.1057/s41278-017-0078-4. URL: https://search.proquest.com/docview/2125521918.

Kim, Young-Rong, Min Jung, and Jun-Bum Park (2021). "Development of a Fuel Consumption Prediction Model Based on Machine Learning Using Ship In-Service Data". English. In: *Journal of marine science and engineering* 9.2, p. 137. DOI: 10.3390/jmse9020137. URL: https://search.proquest.com/docview/2486206310.

Kokkinakos, Panagiotis et al. (Jun 2017). "Big data exploitation for maritime applications a multi-segment platform to enable maritime big data scenarios". English. In: IEEE, pp. 1131–1136. DOI: 10.1109/ICE.2017.8280008. URL: https://ieeexplore.ieee.org/document/8280008.

Kontopoulos, Ioannis, Iraklis Varlamis, and Konstantinos Tserpes (2021). "A distributed framework for extracting maritime traffic patterns". English. In: *International journal of geographical* *information science : IJGIS* 35.4, pp. 767–792. DOI: 10.1080/13658816.2020.1792914. URL: http://www.tandfonline.com/doi/abs/10.1080/13658816.2020.1792914.

Lambrou, Maria, Daisuke Watanabe, and Junya Iida (2019). "Shipping digitalization management: conceptualization, typology and antecedents". English. In: *Journal of shipping and trade* 4.1, pp. 1–17. DOI: 10.1186/s41072-019-0052-7. URL: http://www.econis.eu/PPNSET?PPN=168691072X.

Lee, Habin et al. (2018). "A decision support system for vessel speed decision in maritime logistics using weather archive big data". English. In: *Computers & operations research* 98, pp. 330–342. DOI: 10.1016/j.cor.2017.06.005. URL: http://dx.doi.org/10.1016/j.cor.2017.06.005.

Lee, Jeong-Seok et al. (2020). "Verification of Novel Maritime Route Extraction Using Kernel Density Estimation Analysis with Automatic Identification System Data". English. In: *Journal of marine science and engineering* 8.5, p. 375. DOI: 10.3390/jmse8050375. URL: https://search.proquest.com/docview/2407666120.

Lei, Po-Ruey (2020). "Mining maritime traffic conflict trajectories from a massive AIS data". English. In: *Knowledge and information systems* 62.1, pp. 259–285. DOI: 10.1007/s10115-019-01355-0. URL: https://search.proquest.com/docview/2194975762.

Lepore, Antonio et al. (2017). "A comparison of advanced regression techniques for predicting ship CO2 emissions". English. In: *Quality and reliability engineering international* 33.6, pp. 1281–1292. DOI: 10.1002/qre.2171. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.2171.

Li, Huanhuan et al. (2017). *A Dimensionality Reduction-Based Multi-Step Clustering Method for Robust Vessel Trajectory Analysis.* DOI: 10.3390/s17081792.

Li, Xin et al. (Aug 2016). "Design and Planning of Heterogeneous Marine Sensor Networks for Marine Intelligent Transportation". English. In: IEEE, pp. 2052–2057. DOI: 10.1109/TrustCom.2016.0314. URL: https://ieeexplore.ieee.org/document/7847197.

Li, Yongyi, Zhongqiang Yang, and Kaixu Han (2020). "Research on the clustering algorithm of ocean big data based on self-organizing neural network". English. In: *Computational intelligence* 36.4, pp. 1609–1620. DOI: 10.1111/coin.12299. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/coin.12299.

Li, Yuzhou et al. (2018). "Marine Wireless Big Data: Efficient Transmission, Related Applications, and Challenges". English. In: *IEEE wireless communications* 25.1, pp. 19–25. DOI: 10.1109/MWC.2018.1700192. URL: https://ieeexplore.ieee.org/document/8304386.

Lin, Bo (2020). "Innovation Report of Big Data on Marine Forecast on the North Line of Maritime Silk Road". English. In: vol. 213. Les Ulis: EDP Sciences, p. 03023. ISBN: 2555-0403. DOI: 10.1051/e3sconf/202021303023. URL: https://search.proquest.com/docview/2465724860.

Lin, Bo and Fan Xiao (2020). *Application Engineering of Big Data Technology for Global Marine Environment Forecasting.* DOI: 10.1088/1755-1315/502/1/012025.

Lohr, S. (2017). "The Age of Big Data". English. In: *New York Times.* URL: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-theworld.html.

Luaces, Migual R. and Farid Karimpour (2018). *Miguel R. Luaces Farid Karimipour (Eds.) Web and Wireless Geographical Information Systems.*

Lytra, Ioanna et al. (Jun 2017). "A big data architecture for managing oceans of data and maritime applications". English. In: IEEE, pp. 1216–1226. DOI: 10.1109/ICE.2017.8280019. URL: https://ieeexplore.ieee.org/document/8280019.

Man, Yemao et al. (2020). "From Ethnographic Research to Big Data Analytics—A Case of Maritime Energy-Efficiency Optimization". English. In: *Applied sciences* 10.6, p. 2134. DOI: 10.3390/app10062134. URL: https://search.proquest.com/docview/2383275414.

Mirović, Maris, Mario Miličević, and Ines Obradović (2018). "Big Data in the Maritime Industry". English. In: *Naše more znanstveni časopis za more i pomorstvo* 65.1, p. 56. DOI: 10.17818/NM/2018/1.8. URL: https://hrcak.srce.hr/195330.

Moreira, Lúcia, Roberto Vettor, and C. Guedes Soares (2021). "Neural Network Approach for Predicting Ship Speed and Fuel Consumption". English. In: *Journal of marine science and engineering* 9.2, p. 119. DOI: 10.3390/jmse9020119. URL: https://search.proquest.com/docview/2483971347.

morethanshipping.com (2021). In: URL: https://www.morethanshipping.com/fuel-costs-ocean-shipping/.

Mosavi, Amir et al. (2019). "State of the art of machine learning models in energy systems, a systematic review". English. In: *Energies (Basel)* 12.7, p. 1301. DOI: 10.3390/en12071301. URL: https://search.proquest.com/docview/2316760975.

Munim, Ziaul Haque et al. (2020). "Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions". English. In: *Maritime policy and management* 47.5, pp. 577–597. DOI: 10.1080/03088839.2020.1788731. URL: http://www.tandfonline.com/doi/abs/10.1080/03088839.2020.1788731.

Peng, Peng et al. (2018). "Modelling the competitiveness of the ports along the Maritime Silk Road with big data". English. In: *Transportation research. Part A, Policy and practice* 118, pp. 852–867. DOI: 10.1016/j.tra.2018.10.041. URL: http://dx.doi.org/10.1016/j.tra.2018.10.041.

Perera, Lokukaluge P. and B. Mo (2018). "Ship speed power performance under relative wind profiles in relation to sensor fault detection". English. In: *Journal of ocean engineering and science* 3.4, pp. 355–366. DOI: 10.1016/j.joes.2018.11.001. URL: http://dx.doi.org/10.1016/j.joes.2018.11.001.

Perera, Lokukaluge P. and Brage Mo (2016). "Marine Engine Operating Regions under Principal Component Analysis to evaluate Ship Performance and Navigation Behavior". English. In: *IFAC PapersOnLine* 49.23, pp. 512–517. DOI: 10.1016/j.ifacol.2016.10.487. URL: http://dx.doi.org/10.1016/j.ifacol.2016.10.487.

– (2020). "Ship performance and navigation information under high-dimensional digital models". English. In: *Journal of marine science and technology* 25.1, pp. 81–92. DOI: 10.1007/s00773-019-00632-5. URL: http://hdl.handle.net/10037/17921.

Pezzani, Lorenzo and Charles Heller (2019). "AIS Politics: The Contested Use of Vessel Tracking at the EU's Maritime Frontier". English. In: *Science, technology, & human values* 44.5, pp. 881–899. DOI: 10.1177/0162243919852672. URL: https://journals.sagepub.com/doi/full/10.1177/0162243919852672.

Qiao et al. (2019). "M3C: Multimodel-and-Multicue-Based Tracking by Detection of Surrounding Vessels in Maritime Environment for USV". English. In: *Electronics (Basel)* 8.7, p. 723. DOI: 10.3390/electronics8070723. URL: https://explore.openaire.eu/search/publication?articleId=dedup_wf_001::af252a969edfeb94b642088567cbe979.

Ryazanov, Igor et al. (2021). "Deep Learning for Deep Waters: An Expert-in-the-Loop Machine Learning Framework for Marine Sciences". English. In: *Journal of marine science and engineering* 9.2, p. 169. DOI: 10.3390/jmse9020169. URL: https://search.proquest.com/docview/2488760057.

Sanchez-Gonzalez, Pedro-Luis et al. (2019). "Toward Digitalization of Maritime Transport?" English. In: *Sensors (Basel, Switzerland)* 19.4, p. 926. DOI: 10.3390/s19040926. URL: https://www.ncbi.nlm.nih.gov/pubmed/30813277.

Sarabia-Jacome, David et al. (2020). "Seaport Data Space for Improving Logistic Maritime Operations". English. In: *IEEE access* 8, pp. 4372–4382. DOI: 10.1109/ACCESS.2019.2963283. URL: https://ieeexplore.ieee.org/document/8946609.

Schoening, Timm (2019). "SHiPCC—A Sea-going High-Performance Compute Cluster for Image Analysis". English. In: *Frontiers in Marine Science* 6. DOI: 10.3389/fmars.2019.00736. URL: https://search.proquest.com/docview/2319073034.

Shaw, Heiu-Jou and Fu-Ming Tzu (2019). "The Strategy of Energy Saving for Smart Shipping". English. In: *Advances in technology innovation* 4.3, pp. 165–176. URL: https://search.proquest.com/docview/2255467723.

Sheng, Pan and Jingbo Yin (2018). "Extracting Shipping Route Patterns by Trajectory Clustering Model Based on Automatic Identification System Data". English. In: *Sustainability (Basel, Switzerland)* 10.7, p. 2327. DOI: 10.3390/su10072327. URL: https://explore.openaire.eu/search/publication?articleId=dedup_wf_001::d8929133e5544249e4982903f99ed86c.

Snijders, C., U. Matzat, and U.-D. Reips (2012). "'Big Data': Big Gaps of Knowledge in the Field of Internet Science"". English. In: *International Journal of Internet Science*, pp. 1–5.

Snyder, Hannah (2019). "Literature review as a research methodology: An overview and guidelines". In: *Journal of Business Research* 104, pp. 333–339. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2019.07.039. URL: https://www.sciencedirect.com/science/article/pii/S0148296319304564.

Statista. URL: https://www.statista.com/statistics/267605/capacity-of-oil-tankers-in-the-world-maritime-trade-since-1980/.

Thombre, Sarang et al. (2016). "Operational Scenarios for Maritime Safety in the Baltic Sea". English. In: *Navigation (Washington)* 63.4, pp. 521–531. DOI: 10.1002/navi.161. URL: https://api.istex.fr/ark:/67375/WNG-W904F69F-3/fulltext.pdf.

Tranfield, David, David Denyer, and Palminder Smart (2003). "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review". In: *British Journal of Management* 14.3, pp. 207–222. DOI: https://doi.org/10.1111/1467-8551.00375. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8551.00375. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8551.00375.

Tsou, Ming-Cheng (2019a). "Big data analysis of port state control ship detention database". English. In: *Journal of marine engineering and technology* 18.3, pp. 113–121. DOI: 10.1080/20464177.2018.1505029. URL: http://www.tandfonline.com/doi/abs/10.1080/20464177.2018.1505029.

– (2019b). "Big data analytics of safety assessment for a port of entry: A case study in Keelung Harbor". English. In: *Proceedings of the Institution of Mechanical Engineers. Part M, Journal of engineering for the maritime environment* 233.4, pp. 1260–1275. DOI: 10.1177/1475090218805245. URL: https://journals.sagepub.com/doi/full/10.1177/1475090218805245.

Ullo, Silvia Liberata and G. R. Sinha (2020). "Advances in Smart Environment Monitoring Systems Using IoT and Sensors". English. In: *Sensors (Basel, Switzerland)* 20.11, p. 3113. DOI: 10.3390/s20113113. URL: https://search.proquest.com/docview/2409475888.

Venskus, Julius et al. (2019). "Real-Time Maritime Traffic Anomaly Detection Based on Sensors and History Data Embedding". English. In: *Sensors (Basel, Switzerland)* 19.17, p. 3782. DOI: 10.3390/s19173782. URL: https://www.ncbi.nlm.nih.gov/pubmed/31480449.

Waluyo, Agustinus et al. *EDITOR-IN-CHIEF ASSOCIATE EDITORS EDITORIAL REVIEW BOARD An official publication of the Information Resources Management Association International Journal of Mobile Computing and Multimedia Communications*.

Wang, Jia et al. (2020). "A Survey of Technologies for Unmanned Merchant Ships". English. In: *IEEE access* 8, pp. 224461–224486. DOI: 10.1109/ACCESS.2020.3044040. URL: https://ieeexplore.ieee.org/document/9291468.

Wang, Jianping, Lijuan Ma, and Wei Chen (2017). "Design of underwater acoustic sensor communication systems based on software-defined networks in big data". English. In: *International* *journal of distributed sensor networks* 13.7, p. 155014771771967. DOI: 10.1177/1550147717719672. URL: https://journals.sagepub.com/doi/full/10.1177/1550147717719672.

Wang, Yuhong et al. (2021). "Incorporation of deficiency data into the analysis of the dependency and interdependency among the risk factors influencing port state control inspection". English. In: *Reliability engineering & system safety* 206, p. 107277. DOI: 10.1016/j.ress.2020.107277. URL: http://dx.doi.org/10.1016/j.ress.2020.107277.

Wang, Zhihuan, Christophe Claramunt, and Yinhai Wang (2019). *Extracting Global Shipping Networks from Massive Historical Automatic Identification System Sensor Data: A Bottom-Up Approach.* DOI: 10.3390/s19153363.

Wen, Jiabao et al. (2020). "Big Data Driven Marine Environment Information Forecasting: A Time Series Prediction Network". English. In: *IEEE transactions on fuzzy systems* 29.1, p. 1. DOI: 10.1109/TFUZZ.2020.3012393. URL: https://ieeexplore.ieee.org/document/9151406.

wLei, Bao (2019). "The MapReduce based statistical model on maritime location big data". English. In: *IOP conference series. Earth and environmental science* 310.2, p. 22044. DOI: 10.1088/1755-1315/310/2/022044. URL: https://iopscience.iop.org/article/10.1088/1755-1315/310/2/022044.

Wong, Geoff et al. (2013). "RAMESES publication standards: meta-narrative reviews". In: *Journal of Advanced Nursing* 69.5, pp. 987–1004. DOI: https://doi.org/10.1111/jan.12092. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jan.12092. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jan.12092.

Wu, Pei-Ju, Mu-Chen Chen, and Chih-Kai Tsau (2017). "The data-driven analytics for investigating cargo loss in logistics systems". English. In: *International journal of physical distribution & logistics management* 47.1, pp. 68–83. DOI: 10.1108/IJPDLM-02-2016-0061. URL: https://www.emerald.com/insight/content/doi/10.1108/IJPDLM-02-2016-0061/full/html.

Xiao, Yi, Zhuo Chen, and Levi Mcneil (2021). *Digital empowerment for shipping development: a framework for establishing a smart shipping index system.* DOI: 10.1080/03088839.2021.1894364.

Xiao, Yijia et al. (2020). "Oil Flow Analysis in the Maritime Silk Road Region Using AIS Data". English. In: *ISPRS international journal of geo-information* 9.4, p. 265. DOI: 10.3390/ijgi9040265. URL: https://search.proquest.com/docview/2394488519.

Yan, Xinping et al. (2018). "Energy-efficient shipping: An application of big data analysis for optimizing engine speed of inland ships considering multiple environmental factors". English. In: *Ocean engineering* 169, pp. 457–468. DOI: 10.1016/j.oceaneng.2018.08.050. URL: http://dx.doi.org/10.1016/j.oceaneng.2018.08.050.

Yang, Dong et al. (2019a). "How big data enriches maritime research - a critical review of Automatic Identification System (AIS) data applications". English. In: *Transport reviews* 39.6, pp. 755–773. DOI: 10.1080/01441647.2019.1649315. URL: http://www.tandfonline.com/doi/abs/10.1080/01441647.2019.1649315.

Yang, Jiachen et al. (2019b). "Prediction of Marine Pycnocline Based on Kernel Support Vector Machine and Convex Optimization Technology". English. In: *Sensors (Basel, Switzerland)* 19.7, p. 1562. DOI: 10.3390/s19071562. URL: https://www.ncbi.nlm.nih.gov/pubmed/30935145.

Yang, Tingting et al. (2018). "Resource allocation in cooperative cognitive radio networks towards secure communications for maritime big data systems". English. In: *Peer-to-peer networking and applications* 11.2, pp. 265–276. DOI: 10.1007/s12083-016-0482-z. URL: https://search.proquest.com/docview/2002184239.

Yoo, Sang-Lok (2018). "Near-miss density map for safe navigation of ships". English. In: *Ocean engineering* 163, pp. 15–21. DOI: 10.1016/j.oceaneng.2018.05.065. URL: http://dx.doi.org/10.1016/j.oceaneng.2018.05.065.

Yuen, Kum Fai, Gangyan Xu, and Jasmine Siu Lee Lam (2020). "Special issue on 'Artificial Intelligence & big data in shipping'". English. In: *Maritime policy and management* 47.5, pp. 575–576. DOI: 10.1080/03088839.2020.1790052. URL: http://www.tandfonline.com/doi/abs/10.1080/03088839.2020.1790052.

Zaman, Ibna et al. (2017). "Challenges and Opportunities of Big Data Analytics for Upcoming Regulations and Future Transformation of the Shipping Industry". English. In: *Procedia engineering* 194, pp. 537–544. DOI: 10.1016/j.proeng.2017.08.182. URL: http://dx.doi.org/10.1016/j.proeng.2017.08.182.

Zerbino, Pierluigi et al. (2019). "Towards Analytics-Enabled Efficiency Improvements in Maritime Transportation: A Case Study in a Mediterranean Port". English. In: *Sustainability (Basel, Switzerland)* 11.16, p. 4473. DOI: 10.3390/su11164473. URL: https://explore.openaire.eu.

Zhang, Nuo and Kun Zheng (2020). "Research and Design of the Architecture of the Marine Logistics Information Platform Based on Big Data". English. In: *Journal of coastal research* 106.sp1, p. 628. DOI: 10.2112/SI106-142.1.

Zhang, Shu kai et al. (2018). "Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity". English. In: *Ocean engineering* 155, pp. 240–250. DOI: 10.1016/j.oceaneng.2018.02.060. URL: http://dx.doi.org/10.1016/j.oceaneng.2018.02.060.

Zhang, Xiunian and Jasmine Siu Lee Lam (2019). "A fuzzy Delphi-AHP-TOPSIS framework to identify barriers in big data analytics adoption: case of maritime organizations". English. In: *Maritime policy and management* 46.7, pp. 781–801. DOI: 10.1080/03088839.2019.1628318. URL: http://www.tandfonline.com/doi/abs/10.1080/03088839.2019.1628318.

Zhang, Yihong et al. (2020). "Parallel Three-Branch Correlation Filters for Complex Marine Environmental Object Tracking Based on a Confidence Mechanism". English. In: *Sensors (Basel, Switzerland)* 20.18, p. 5210. DOI: 10.3390/s20185210. URL: https://search.proquest.com/docview/2443281914.

Zhang, Zhihua, Donald Huisingh, and Malin Song (2019). "Exploitation of trans-Arctic maritime transportation". English. In: *Journal of cleaner production* 212, pp. 960–973. DOI: 10.1016/j.jclepro.2018.12.070. URL: http://dx.doi.org/10.1016/j.jclepro.2018.12.070.

Zhou, Xiangyu et al. (2020). *Using Deep Learning to Forecast Maritime Vessel Flows*. DOI: 10.3390/s20061761.