

2015

# Major challenges and solutions for utilizing big data in the maritime industry

Sadaharu Koga  
*World Maritime University*

Follow this and additional works at: [http://commons.wmu.se/all\\_dissertations](http://commons.wmu.se/all_dissertations)

 Part of the [Databases and Information Systems Commons](#), and the [Transportation Commons](#)

---

## Recommended Citation

Koga, Sadaharu, "Major challenges and solutions for utilizing big data in the maritime industry" (2015). *World Maritime University Dissertations*. 490.  
[http://commons.wmu.se/all\\_dissertations/490](http://commons.wmu.se/all_dissertations/490)

This Dissertation is brought to you courtesy of Maritime Commons. Open Access items may be downloaded for non-commercial, fair use academic purposes. No items may be hosted on another server or web site without express written permission from the World Maritime University. For more information, please contact [library@wmu.se](mailto:library@wmu.se).

WORLD MARITIME UNIVERSITY

Malmö, Sweden

Major challenges and solutions for utilizing  
big data in the maritime industry

By

**SADAHARU KOGA**

Japan

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

MASTER OF SCIENCE

In

MARITIME AFFAIRS

(MSEA)

2015

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

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

(Signature): 古賀定治  
(Date): September 21, 2015

Supervised by: M. Baldauf

World Maritime University

Assessor: D. Dalaklis

Institution/organisation: World Maritime University

Co-assessor: C. Quass

Institution/organisation: INTERSCHALT maritime systems AG

## Acknowledgement

I am sincerely thankful to all the associates who helped me complete this dissertation. Although they are too many to mention, I would like to express some of my humble appreciations here.

I am deeply appreciative to faculties and staffs of the World Maritime University. Especially I wish to offer my profound thanks to Associate Professor Michael Baldauf. He always gave me insightful comments and kindly shared his knowledge and experience with me. Every single discussion with him was fruitful for me and strongly propelled my work forward. I am also very grateful to Associate Professor Clive Cole for giving me kind and appropriate linguistic supports. Besides, I would like to show my thankfulness to librarians, too. Chris Hoebeke and Anna Volkova provided me with the quiet and comfortable place to concentrate on this dissertation.

I am extremely indebted to Dr. Yohei Sasakawa, who funded my studies as the Chairman of the Nippon Foundation. What he and his foundation gave me was an opportunity not only to study in the World Maritime University, but also to make valued lifelong friends through the precious life in Malmö. I would also like to express my special thanks to the staffs of the Sasakawa Peace Foundation; Mr. Eisuke Kudo and Mr. Shinichi Ichikawa. They offered great supports, directly and indirectly, for my studies in Sweden in various aspects.

I am dearly grateful to Mr. Koichi Kato, Deputy Director-General for Engineering Affairs of Maritime Bureau, Ministry of Land, Infrastructure, Transport and Tourism. When I consulted him about my wish for studying in the World Maritime University, he immediately and graciously agreed and provided me with powerful supports toward the enrollment.

Last and foremost, I would like to thank my beloved parents and brother in Japan. Every time I talk with them, they give me calmness of mind. My achievements in Malmo would not come true without their love and support.

Title of Dissertation: Major challenges and solutions for utilizing big data in the maritime industry

Degree: MSc

The dissertation is a study of big data for the use in the maritime industry. Today's society is information-intensive. The term "big data" is becoming more common. In fact, some maritime companies and institutions have already been trying to utilize big data for enhancing maritime safety and environmental protection. In order to promote this trend, the dissertation tries to identify common and important challenges for the whole maritime industry in terms of the utilization of big data and propose corresponding solutions.

First, by reviewing the definitions of big data, three major features are identified. Big data takes electronic form, is derived through various sensors, and has difficulties in treatment. In terms of difficulty, there are four aspects, which are volume, velocity, variety and veracity. Noting these features, the scope of the dissertation is set as electronic voyage-related data derived on board ships. It is clarified that the major features of big data is applicable to such data.

Second, examples of cutting edge institution/framework related to maritime big data are studied. As such institutions, DNV-GL, Lloyd's Register Foundation and e-navigation framework in the IMO are chosen. Consequently, four major categories of difficulty in further use of maritime big data are identified. They are sound competitive conditions, human resources, technology and security.

Third, possible solutions for identified four categories of problem are discussed and proposed. Regarding sound competitive conditions, instructions on rights and responsibilities of treating big data should be provided. As regards human resources, governments are expected to show its strategy on nurturing experts in need and promote cooperation with the academia and associated industry. Development of technology can be underpinned by research and development aid scheme as well as unification of the style of relevant data sets. Security issue requires well-established legislations and secure and resilient system against cyber-attacks.

**KEYWORDS:** big data, e-navigation, human resource development, electronic/information technology, data security

Declaration	ii
Acknowledgement	iii
Abstract	iv
Table of contents	v
List of Tables	vi
List of Figures	vii
List of Abbreviations	viii
1 Introduction	1
2 Definition and main features of Big Data	6
2-1 Big data in general sense	6
2-2 Maritime big data	12
3 Cutting edge institutions of maritime big data	21
3-1 DNV-GL	21
3-2 Lloyd's Register Foundation (LRF)	28
3-3 E-navigation	34
4 Analysis of challenges and solutions	45
4-1 Sound competitive conditions	46
4-2 Human resources	56
4-3 Technology	64
4-4 Security	74
5 Conclusion	80

## LIST OF TABLES

Table 1	List of information to be recorded in VDR (mandatory & recommendatory)	16
---------	--	----

## LIST OF FIGURES

Figure 1	Essence of “4Vs” as main difficulties of treating big data	11
Figure 2	Overview of the important features of big data	11
Figure 3	Image of the scope of the dissertation	13
Figure 4	Summary of DNV-GL’s stance regarding big data	27
Figure 5	Summary of LRF’s stance regarding big data	34
Figure 6	Overarching e-navigation architecture	38
Figure 7	Summary of the consideration of e-navigation	44
Figure 8	Identified challenges for development of maritime big data	45
Figure 9	Summary of challenges and solutions regarding sound competitive conditions	56
Figure 10	Structure of framework of industry-government-academia cooperation	63
Figure 11	Summary of challenges and approach to them in terms of technology	67
Figure 12	Integrated simulation system adopted in MUNIN	70
Figure 13	Structure of data security problem	77



## LIST OF ABBREVIATIONS

ABS	American Bureau of Shipping
ABS	Anti-lock braking system
AFBF	American Farm Bureau Federation
AIS	Automatic Identification System
ATP	Agriculture-Technology Providers
CMDS	Common Maritime Data Structure
COMSAR	Sub-committee on Radiocommunication and Search and Rescue
CYSPA	European Cyber Security Protection Alliance
DNV	Det Norske Veritas
ECDIS	Electronic Chart Display & Information System
GL	Germanischer Lloyd
GPS	Global Positioning System
ICS	International Chamber of Shipping
IHO	International Hydrographic Organization
IMO	International Maritime Organization
LRF	Lloyd's Register Foundation
MSC	Maritime Safety Committee
MUNIN	Maritime Unmanned Navigation through Intelligence in Networks
NAV	Sub-committee on Safety of Navigation
SIP	E-navigation Strategy Implementation Plan
SOLAS	International Convention for the Safety of Life at Sea
SQA	Software Quality Assurance
STEM	Science, Technology, Engineering and Mathematics
U.S	United States
UMUC	University of Maryland University College
USCG	United States Coast Guard
VDR	Voyage Data Recorder
VHF	Very High Frequency
VTS	Vessel Traffic Service

# **1 Introduction**

## Data explosion

Today's society is an information-intensive society. Throughout our lives, we are creating, receiving and utilizing all sorts of information and, in many cases, such information takes the form of electronic data. Computers and cell-phones may be the most familiar and common examples of devices which deal with information as electronic data. Because of such devices, data is explosively increasing. For example, through twitter, which is one of the world's famous social networking applications, 500,000,000 posts are being made every day and 3.5 terabytes of data is created per year (LRF, 2014). IBM states that, "every day, we create 2.5 quintillion bytes of data; so much that 90% of the data in the world today has been created in the last two years alone" (IBM, n.d.).

## Big data is a trend

Under this circumstance, the term "big data" is becoming more and more common. An example of the definition of big data is "a massive collection of digital data that is so large and complex to make difficult its processing by using traditional data management tools and techniques" (Catlett et al., 2013). From this definition, it can be guessed that increasing data makes it difficult to treat such data with existing tools

or systems and how to utilize them is an emerging issue. Despite such difficulty, many industries consider that big data is a key for innovation and a new opportunity for their business. For example, when we use the internet in our rooms, we sometimes see a web page, such as Amazon or eBay, automatically introducing and recommending some products to which we may have an interest. This brings us a surprise, but it is a commercial strategy of mail-order companies underpinned by technologies related to big data. They process and analyze the data of “what a user has bought in the past, which items they have in their virtual shopping cart, items they’ve rated and liked, and what other customers have viewed and purchased” (Mangalindan, 2012). Another example is found in the transportation sector. Since 2013, TOYOTA has been providing a cloud data service called “Big Data Traffic Information Service” in road traffic sector on the basis of huge data derived through its own telematics system. It says that “the cloud-based service allows the use of TOYOTA's proprietary T-Probe traffic information, route history, traffic volume map, deployment locations of ABS or other on board emergency safety systems, and other map information data via PC, smartphone, or tablet” (TOYOTA, 2013). This system is expected to help people, in the case of disaster, find information about evacuation sites, shelters, and other facilities. It seems safe to say that big data is one of the burning issues of business in general.

#### Maritime big data

Big data is intensively interested in land-based industries. Then, how about maritime industries? Is there such huge data in maritime industries? Yes. For example, the accumulation of data on board ship is found in the Voyage Data Recorder (VDR),

which is required by SOLAS (1974) and has been fitted on many vessels since 2002. In VDR, the data such as a ship's location, speed and bridge audio are recorded throughout voyages. It is true that, on vessels, more and more information is derived and stored as data. However, it is not yet common to process or analyze such data for innovative purposes. Actually, the main official purpose of VDR is data analysis in the case of accidents. Therefore, on most ships, data is just recorded during the voyage and, if no accident has occurred, the data will be abandoned and deleted.

If these data are effectively utilized, great innovation may be achieved in the maritime industry as with Amazon and TOYOTA. In fact, there are some players who are trying to nurture innovation with big data in the maritime field. For example, Class NK (2015) is trying to use big data for machinery condition monitoring to enhance the safety of the voyage. This project is relying on various sensors put inside engines and data derived by them. In addition, DNV-GL (2015) has been running a project, titled ReVolt, which aspires unmanned, zero emission short sea ships. DNV-GL considers that utilization of big data is supposed to be a significant help in this project.

#### Further use of big data in maritime industry

As illustrated by examples of Class NK and DNV-GL, it is considered that utilization of big data enables the maritime industry to drastically enhance safety of ships and environmental protection. Thus, the trend of big data is true for the maritime industries, too. However, this trend is not collective in maritime industries so far; players are just trying to invent their own systems or services individually, without

any unified goal or direction. Thus, it is going to be helpful for the whole maritime industry if critical and common difficulties or bottlenecks for further innovation with big data are identified.

Taking into account such background, this dissertation starts by reviewing the definition of big data and figuring out its main features in general and tries to apply them for maritime affairs. It turns out that general features of big data can be applied for data derived on board vessels. Secondly, strategies and stances of leading institutions in terms of maritime big data, namely DNV-GL and Lloyd's Register Foundation, are studied and significant challenges of the further use of big data which are common for the whole maritime society are identified. In addition, e-navigation, which is one of the IMO's priority issues, is also taken into consideration. Consequently, four aspects, namely sound competitive conditions, human resources, technology and security, are identified as common and important challenges. Finally, for those challenges, possible solutions are discussed and recommended. In terms of sound competitive conditions, clear instructions on the rights and responsibilities of treating big data need to be provided to industry. As regards human resources, the government is expected to show clear strategy and cooperate with the industry and academia to nurture experts on maritime big data, noting that the maritime industry is supposed to face shortages of such specialists in near future. The development of technologies related to maritime big data needs to be supported through governmental research and development aid and unification of format and structure of relevant data items. Security of the system and data has to be ensured by establishing corresponding legislation and developing highly secure and

resilient technologies. If the solutions are fully applied in these four aspects, the maritime industry will see significant advancement in big data innovation.

## **2 Definition and main features of Big Data**

### **2-1 Big data in general sense**

#### What is big data?

The term “big data” has become more and more popular recently. As reviewed in the introduction, big data is underpinned by various digital devices which surround many aspects of our life. Through computers or cell-phones, we watch news, type business documents, blog, buy something and send messages on a daily basis. All of these activities are processes of observing, creating, receiving and utilizing a variety of data. Electronic data always exists beside us and, basically, these are the source of big data. Therefore, it does not seem so difficult to guess a very brief definition of big data. Roughly speaking, it can be like “huge amount of data being derived through various human activities”. But, what are the strict definitions and features of big data?

#### General definition of big data

The term “big data” was created in the field of space-engineering 18 years ago. According to Press (2014) and Friedman (2012), the first use of the term “big data” appeared in 1997 in the work of Michael Cox and David Ellsworth who were scientists at NASA. They defined big data in order to describe certain challenge

which they were facing in the 1990s. This “challenge” was massive information that was generated by supercomputers and could not be processed and visualized by the technology of that time. Since then, thanks to the rapid development of computing and information technology, the notion of “big data” has spread to various industries.

Apart from this initial definition, in recent years, there have been many perceptions of big data especially in the information industry. One of the simplest and most familiar expressions of big data may be one which was introduced by IBM, which is one of the leading companies in the information industry. It states “Data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data” (IBM, n.d.). This seems to imply two very fundamental features of big data. First, big data is regarded as a kind of huge electronic data which is explosively increasing. Although it may be a matter of course that big data is electronic, it is important to recognize and put an emphasis on this feature in order to precisely define big data. Second, big data is being derived through sensors which have been developed rapidly. According to IBM's definition, the term “sensor” means not only ones such as speed meters and gyroscopes but also such as smart phones and digital cameras, which capture information and interpret it into electronic data. Thanks to such sensors, big data can be generated or detected.

In addition to these two fundamental features of big data, another important aspect is pointed out by several institutions. The Oxford English Dictionary (OED), which is one of the most authoritative English dictionaries in the world, has provided more



academic definition, which is “computing (also with capital initials) data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges” (2008). This definition also shows that big data is data of “very large size”, as similarly observed by IBM. However, in addition to this feature, the OED refers to another important feature of big data; it is the “extent” of the amount or size of such data. By describing “its manipulation and management present significant logistical challenges”, OED clearly states that difficulty/challenge of manipulation and management is one of the features of big data. This point is supported by McKinsey (2011) who expresses big data as “datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze”. This simple definition seems to be sharply hitting the point by referring to difficulties of “capture, store, manage, and analyze”. Furthermore, it should be recalled that the “initial definition” of big data presented by NASA also implied this feature of big data, pointing out the impossibility of processing and visualizing by existing tools.

Taking into account the definitions introduced so far, it seems possible to say that there are three fundamental features of big data.

- Electronic form
- Derived through various sensors (which have been rapidly developed)
- Difficulty of capturing, storing, managing and analyzing

#### Difficulty of dealing with big data

With respect to difficulty in treating big data, a question is remaining. What are the

reasons for such a difficulty? In order to grasp what is big data, a breakdown in this respect is inevitable. In 2001, Doug Laney, who is an industry analyst, introduced the notion of “3Vs” in the field of e-commerce in order to describe the difficulty of dealing with relevant information. He stated that the “3Vs” represent “volume”, “velocity” and “variety” (Laney, 2001). The notion of “3Vs” is agreed and accepted by many institutions and companies today and widely regarded as the “mainstream definition of big data” (SAS, n.d.).

Out of these “3Vs”, the first “V” means volume. This simply means the number or amount, depth and breadth of available data. As already illustrated by the example of “2.5 quintillion bytes of data per day”, the volume of existing data is totally different from what it used to be in the past. Because of the volume, it may be difficult not only to detect or select a dataset which is critical for a certain purpose out of an enormous quantity of data, but also to process and analyze it in clusters. The second “V” stands for velocity, which means the pace of generating and using data. Laney (2001) stated that “point-of-interaction speed and, consequently, the pace data used to support interaction and generated by interaction” has been increased. This aspect seems to matter especially for the transportation industries. As introduced in the example of TOYOTA, it is necessary in this sector to immediately receive and process the data and extract proper outputs in real-time to support the transportation which is going on. Thus, the aspect of velocity is certainly important. The third “V” is variety. There are “incompatible data formats, non-aligned data structures, and inconsistent data semantics” (Laney, 2001) and if the data is given in various formats, it is going to be difficult to observe and treat such data in a unified manner.

In addition to these three “V”s, a fourth “V” has been recognized recently; it is “veracity”. According to DNV-GL (2014), the element of “veracity” was introduced by IBM as an issue of trustworthiness of data. This sounds rational because the more amount of data increases, the more attention people have to pay on whether such data is reliable. These four “V”s are finely summarized by IBM (n.d.) with plain interpretations. The essence of IBM's summary is shown in Figure 1.

#### Important features of big data

In this sub-chapter, as a general consideration of big data, three important features are identified. Big data takes the form of electronic information, it is derived through various sensors, and there are difficulties of capturing, storing, managing and analyzing it. With respect to difficulties, it is commonly considered that there are four main aspects, which are volume, velocity, variety and veracity. These findings can be summarized as in Figure 2.

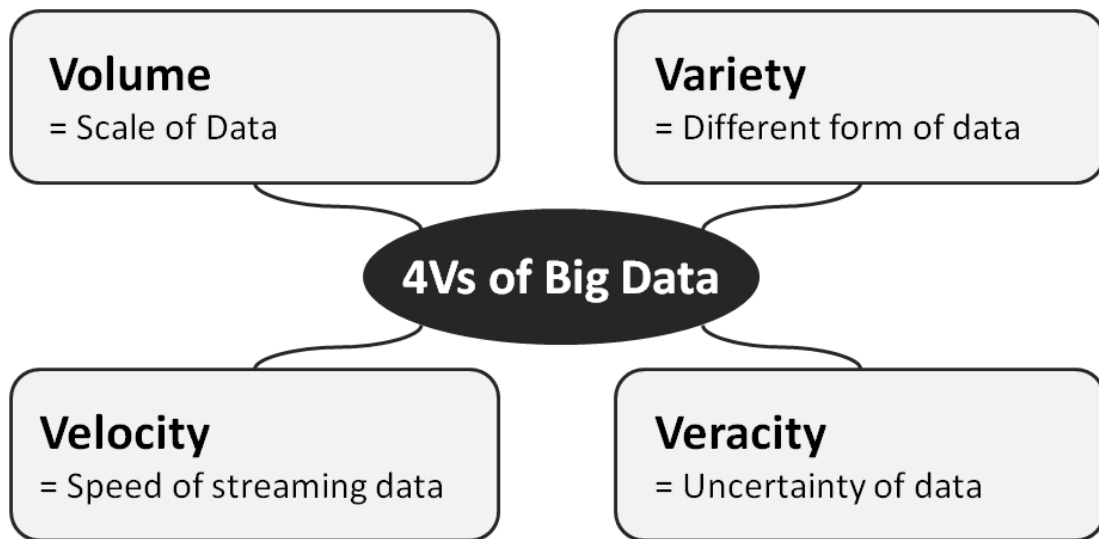


Figure 1: Essence of “4Vs” as main difficulties of treating big data. Reprinted from *The Four V's of Big Data*, by IBM, retrieved from <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

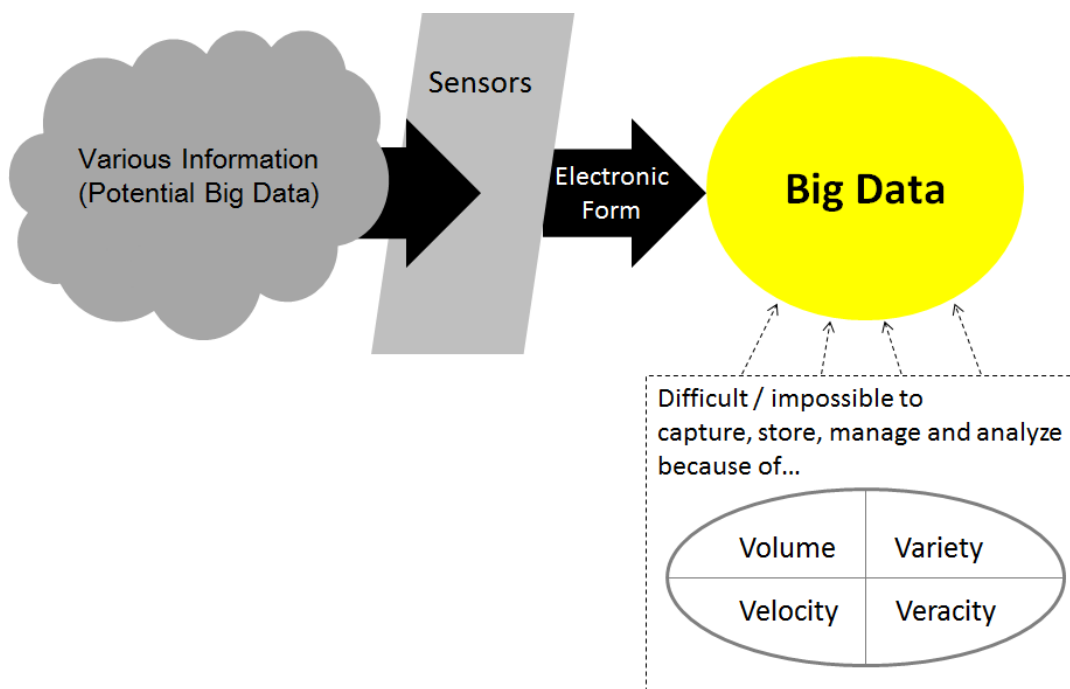


Figure 2: Overview of the important features of big data.

## **2-2 Maritime big data**

### Absolute definition of maritime big data?

In the previous sub-chapter, some notable definitions of big data were reviewed and the main features were shown. Now, it seems rational to consider that big data in the maritime industry is something to which such features are applicable. However, as a matter of fact at this time, there is no internationally established or shared idea about what should be regarded as big data in maritime affairs. Without a common understanding on what is regarded as “maritime big data”, it is going to be quite difficult or impossible to detect common challenges as well as solutions which effectively promote the utilization of such data in the whole of the maritime industry. Thus, what should be regarded as big data in maritime affairs has to be considered.

However, it does not seem feasible to give an absolute definition of maritime big data; because, there are many groups of stakeholders in maritime society and each of them deals with various kinds of data on a daily basis and has a different range of interests. For example, shipowners are generally interested not only in ships’ performance but also freight rates etc. Ship builders should be interested in ships’ performance, too. However they should also be interested in such items as the price of iron, which directly affects the price of ships, and labor costs and performance. Besides, classification societies may be interested in the strength of ship structures and insurance companies may wish to precisely estimate the probability of marine accidents. Although these are just some examples, they clearly show how it is difficult to absolutely define maritime big data.

### Common interests of the maritime industry

As the range of interests to marine-related data depends on stakeholders, it is not possible to perfectly determine what is maritime big data. On the other hand, one of the main purposes of this research is to find the common challenges of maritime industries regarding big data. Taking into account this purpose, the dissertation limits its scope to the most common interests which are shared by most maritime stakeholders. Finding problems or bottlenecks as well as solutions for them within such a scope are expected to benefit broader stakeholders. In this context, the focus of this paper is put on “data which is derived on board vessels”. Information and data derived from such vessels through shipping are crucial for most marine-related industries because they are associated with ships and should be interested in them in light of the optimization of their products or services. As an image, the scope of the dissertation can be shown as in Figure 3.

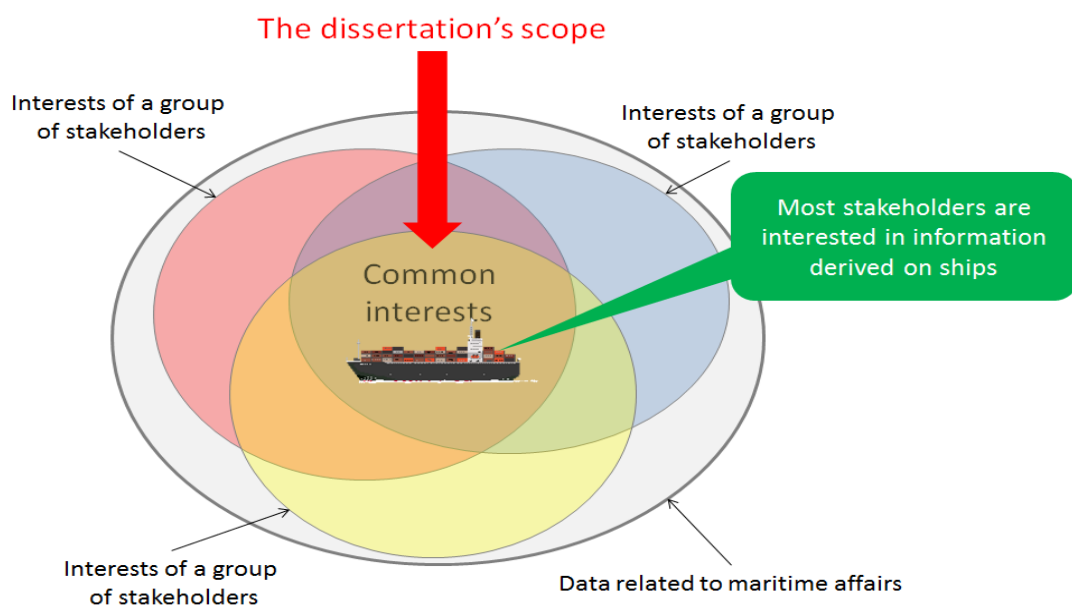


Figure 3: The scope of the dissertation.

### Available data on board vessels

The range of scope has been limited to data available on board vessels. Then, what kinds of electronic data are actually available on board? The equipment which should be firstly recalled is VDR. It is stipulated that, in Regulation 20, Chapter V of SOLAS (1974), vessels have to be fitted with VDR in accordance with the following criteria:

- passenger ships constructed on or after 1 July 2002;
- ro-ro passenger ships constructed before 1 July 2002 not later than the first survey on or after 1 July 2002;
- passenger ships other than ro-ro passenger ships constructed before 1 July 2002 not later than 1 January 2004; and
- ships, other than passenger ships, of 3,000 gross tonnage and upwards constructed on or after 1 July 2002.

In Resolution MSC.333(90) adopted in IMO (2012), it is stated that “the purpose of a voyage data recorder is to maintain a store, in a secure and retrievable form, of information concerning the position, movement, physical status, command and control of a ship”. Thus, the basic purpose of VDR is to record and store the information of each voyage as electronic data.

VDR itself is not equipment which directly observes or derives voyage-related information as a sensor but designed to be connected to tens of electronic devices and to store data items which are derived by those devices. According to the above-mentioned resolution, there are 20 data items to be recorded in VDR, which includes both compulsory and voluntary ones, as shown in Table 1. Those listed in

Table 1 are the actual resources of voyage-related data on board. They include data items regarding ship motion, such as speed, heading, roll motion and accelerations. Not only “motion”, but also “position” of the ships is stored in VDR through AIS and ECDIS. In addition, the order and response of the engine and rudder are also recorded. Furthermore, apart from the physical information related to ship and its equipment, data items of the bridge audio and communication audio are maintained in VDR, too. Equipment which is to be connected to VDR is not limited to a device in Table 1 as it is not an exhaustive list but just a list of compulsory devices and some examples to be equipped on a voluntary basis. Even so, Table 1 is at least a good example to know the variety of data derived on board vessels because most major voyage-related data is designed to be accumulated in VDR.



Table 1: List of information to be recorded in VDR (mandatory & recommendatory)

Date and time
Ship's position
Speed
Heading
Bridge audio
Communications audio
Radar
ECDIS
Echo sounder
Main alarms
Rudder order and response
Engine and thruster order and response
Hull openings status
Watertight and fire door status
Accelerations and hull stresses
Wind speed and direction
AIS
Rolling motion
Configuration data
Electronic logbook

Notes: Reprinted from *Adoption of revised performance standards for shipborne Voyage Data Recorders (VDRs)*, by IMO

### Applicability of the main features of general big data

Major data items to be found on board ships have been reviewed. However, do they suffice for being the main features of big data as discussed in previous sub-chapter? Here, it should be recalled that there are three important features of big data as already summarized in Figure 2.

The first aspect is that big data takes electronic form. It is self-apparent that most voyage-related information takes the form of electronic data because it is recorded in a VDR which keeps the information as electronic data.

The second aspect is that big data is derived through various sensors which have been rapidly developed. This is also applicable to voyage-related information. As a matter of fact, a variety of data items, which are represented by those in Table 1, are coming from different types of sensors on board ships. For example, AIS receives locational information from global navigation satellite systems such as GPS and exchanges information with other vessels or land-based facilities through the VHF radio wave (USCG, n.d.). On the other hand, the heading angle of the ship is measured by gyrocompass (Tokyo Keiki, n.d.). Furthermore, engine performance can be monitored through dedicated sensors. This system is produced by companies such as Kongsberg Maritime (n.g.). As illustrated by these examples, it is obvious that there are various sensors on board vessels which convert various voyage-related information into data.

The third aspect is the difficulty of capturing, storing, managing and analyzing. There are four dimensions of difficulties for treating big data, which are referred to

as the “4Vs”, as already summarized in Table 1. It seems that the “4Vs” are true for the data derived on board ships. The first aspect to be considered is the volume of data. In short, the volume of data on board ship matters. Resolution MSC.333(90) of IMO states that

Recording should be continuous unless terminated in accordance with 5.4.2. The time for which all stored data items are retained should be at least 30 days/720 hours on the long-term recording medium and at least 48 hours on the fixed and float-free recording media. Data items which are older than this may be overwritten with new data (IMO, 2012).

Assuming that a VDR has this minimum capacity of data storage, there are two problems. First, as some international voyages take more than 30 days (Economove Japan, n.d.), old data has to be deleted during such long voyages. Second, even if a voyage does not last more than 30 days, voyage-related data is usually not extracted to other storage devices and just deleted during the next voyage. It is inevitable because VDR is basically equipped for the purpose of post-accident investigation. IMO says that “like the black boxes carried on aircraft, VDRs enable accident investigators to review procedures and instructions in the moments before an incident and help to identify the cause of any accident” (IMO, n.d.). Therefore, ships are not necessarily ready for storing huge amounts of information derived through a number of voyages and processing and analyzing it for the purpose of improvement of such as safety, efficiency, or environmental protection unless the vessel is voluntary fitted with special device with enough capacity. Even for one single vessel, the volume of information is quite heavy. Therefore, it is going to be far more challenging to deal with data from many vessels together.

The second dimension is velocity, which is the pace of generating and using data. Voyage-related information is being produced throughout a voyage by each vessel. According to the International Chamber of Shipping (n.d.), more than 50,000 merchant ships are engaged in international trading. It is not difficult to guess how quickly the information is being created. On the other hand, when a single vessel is focused, the vessel may want to use real-time information to ensure navigational safety or efficiency of fuel consumption. Data is generated in a rapid pace on ships and, at the same time, ships are in need of immediate processing and analyzing of such data. In this context, there are certain challenges in terms of the velocity of generating and using data.

The third dimension is variety. Here again, Laney says the issues of “variety” are “incompatible data formats, non-aligned data structures, and inconsistent data semantics” (Laney, 2001). As already discussed, VDR records various kinds of data. In addition, even for the same kind of data item, the format might not be the same depending on the manufacturer. When these different data items are processed or analyzed, it is inevitable to struggle for unifying or combining them. Therefore, the problem of variety certainly exists in terms of data derived on board ships.

The last element of “4Vs” is veracity. As shown in Figure 1, veracity represents the uncertainty of data. This is a concern mainly for the information industry. On the internet, some people may make up fake data with evil intentions. However, data derived on the vessel is not unrelated to this issue because there is another aspect of uncertainty. This is the reliability of data measured by sensors even without any evil

intentions of humans. The condition of measurement or accuracy of sensor directly affects the credibility of the derived data. As there are a lot of data items on board ships which are physically observed or measured through sensors, this aspect has to be considered very carefully. As discussed here, the four typical difficulties of dealing with big data can be applied to data derived on board ships. In other words, voyage-related data on board vessels can be regarded as significant pieces of “maritime big data”.

#### Scope of the research

Here, the scope of the dissertation can be summarized. Under the purpose of identifying common challenges as well as solutions which effectively promote the utilization of such data in the whole maritime industry, the scope of this paper is limited to data derived on board ships, which is of common interest to most relevant stakeholders. Such data is represented by those stored in VDR. They can be referred to as maritime big data because all of the main features of big data are applicable.

### **3 Cutting edge institutions of maritime big data**

In the previous chapter, a general understanding of maritime big data was developed and the scope of this research was set. On such basis, this chapter reviews activity or strategy of the existing forerunners and identifies common challenges from them for further developing utilization of maritime big data. This chapter picks up three objectives of study, which are namely DNV-GL, Lloyd's Register Foundation (LRF) and IMO's e-navigation. Both of DNV-GL and Lloyd's Register are the world's major and historical classification societies and, therefore, have been accumulating a deep knowledge and experience about ships. On this basis, these institutions have already published their strategy on big data. On the other hand, e-navigation is one of the most prioritized issues treated in the IMO and this concept is sustained by electronic devices and data on board vessels, which may constitute maritime big data. Therefore, the relationship of e-navigation and maritime big data has to be identified in order to avoid a contradiction or duplication between them in the future. In this context, these three examples need to be analyzed and are worth doing so.

#### **3-1 DNV-GL**

##### Overview of DNV-GL

One of the most well-known leaders in the field of maritime big data is DNV-GL. In

maritime society, this institution is usually known as a worldwide classification society established through a merger between Det Norske Veritas (DNV) and Germanischer Lloyd (GL) in 2013, both of which have had a long history since the 1860's (DNV-GL, n.d.). DNV-GL mainly covers five sectors, which are the maritime industry, oil and gas industry, energy industry, business assurance and software (DNV-GL, n.d.). Out of these five coverages, maritime industry occupies the largest portion. It is reported that, in 2014, the revenue of DNV-GL was 21,300 million NOK and the revenue from the maritime domain account for more than 40% of the total (DNV-GL, n.d.). At the same time, DNV-GL is one of the most major classification societies for international shipping as it classed “21% of the world’s classed fleet of ships and floating offshore structures” in Gross Tonnage in 2014 (DNV-GL, 2015). In the maritime domain, the kinds of service provided by DNV-GL are ship and offshore classification services, maritime advisory services, certification of materials and components and Flag and Coastal State services and maritime software services (DNV-GL, n.d.).

#### How DNV-GL observes big data

DNV-GL's definition of big data basically follows the “4Vs” which were introduced by Laney and IBM as reviewed in chapter 2. However, DNV-GL also says “these kinds of definitions are relative, and what is currently considered as big data should be expected to vary and change with time” (DNV-GL, 2014). DNV-GL intentionally avoids drawing a clear shape of big data as of now. In DNV-GL's position paper about big data published in 2014, there are six main areas in which the utilization of big data is expected. They are technical operation and maintenance, energy efficiency,

safety performance, management and monitoring of accidents and environmental risks from shipping traffic, commercial operation and automation of ship operation. According to the position paper, a typical expectation for the future in each area is as below (DNV-GL, 2014).

- “Technical operation and maintenance”

Ship's component and system vendors monitor conditions of their products remotely and advise ship management on operation of the system and predictive maintenance.

- “Energy efficiency”

Such as shipping companies, shipbuilding companies, classification societies and component-manufacturers collect, analyze and present data from vessels to drive continuous energy efficiency improvement.

- “Safety performance”

Data from variety of sources onboard vessels become available for continuous safety assessment activities. By requiring disclosure of more data directly from ships, regulatory bodies may be able to increase transparency of safety of shipping.

- “Management and monitoring of accident and environmental risks from shipping traffic”

Shore-based traffic control centers combine real-time information such as course and speed of the ships with facts regarding the safety condition of the ship, cargo etc. that influences the likelihood and consequences of accidents. Then they can determine the most risky shore areas.



- “Commercial operation”

Combining and analyzing data about the availability of cargoes, available space to take cargoes, port slots, weather, ship performance, fuel prices, etc. may be a significant business.

- “Automation of ship operations”

Ships become more and more automated and, maybe one day, autonomous ships controlled from certain onshore operation centers might come true.

Here, as represented by these main areas of interest, DNV-GL seems to be considering data and information derived on board vessels as the epicenter of maritime big data. On the other hand, it is also assumed that big data from ships may be owned or controlled not only by shipowners but also suppliers of components and the classification society etc.

#### Challenges observed by DNV-GL

In its position paper, DNV-GL introduced the following challenges for the further development of utilization of big data in the maritime industry (DNV-GL, 2014).

- Privacy issues

Potential privacy issues, data protection, and privacy-related risks should be addressed early in any big data project.

- Near monopoly providers of data

If the industry has to strongly depend on data from a “near monopoly provider”, it is obviously not a sound market condition. Thus, considerable efforts directed towards reaching mutually acceptable agreements are

expected.

- Ability to handle and analyze data

For effective utilization of big data, skilled workers who have the associated knowledge, skills of visualization and communications, the ability to materialize the idea into models are necessary. Recognizing this aspect is not easy for organizations as it is usually not visible when the final products are completed.

- Availability of powerful tools for analytics, data exploration, data visualization, and model building

In addition to the traditional toolset for modelling and handling data, the new and available one should be used in order to support streams of sensor data and heterogeneous data.

- Flexibility

The possibility of experimenting and exploring tools, algorithms, and data in safe environment is necessary. In other words, the environment on which ideas can be taken from prototype, through pilot, and into production should be established or provided.

- Data governance

When many datasets are stored on the same infrastructure, controlling ownership and access is one of the critical issues. For the sound utilization of big data, data governance has to be taken care of from the early stage of development.

One interesting finding from these six challenges is that it seems possible to

categorize them into three groups.

The first group is “sound competitive conditions”. Out of the six challenges shown above, issues of privacy, near monopoly providers and data governance should be included in this group. As the notion of big data is relatively new in the maritime industry, the competitive condition is not perfectly established or organized at this time. In other words, what is allowed and not allowed have to be determined and shared by players as rules of the game. How the privacy of individuals should be protected, how monopolistic deals should be avoided and how big data should be traded and controlled should be clearly instructed or regulated. These rules or systems are expected to constitute sound competitive conditions. At the same time, the competitive condition has to be designed to contribute further development of big data utilization and must not be unnecessarily burdensome or restrictive for the industry.

The second group is “technology”. The issue of availability of powerful or helpful tools and methods should be in this category as it is obvious that, in order to promote big data utilization, the more affordable, sophisticated or high-performance tools are necessary. In addition, the challenge with regard to flexibility raised above is also related to the aspect of technology because a good test bed on which new technologies can be examined enables relevant technologies to develop faster.

The third group is “human resources”. This is also an essential element. The development of technology and ideas for business are of course created by human resources and the prerequisite for such people is a rich knowledge of the

corresponding domain. Out of the six challenges raised above, “ability to handle and analyze data” is included in this category.

Findings from DNV-GL’s stance and approach

A summary of DNV-GL's stance and categorization of challenges, as well as findings from them, are shown in Figure 4. This sub-chapter mainly focuses on the general direction of DNV-GL and its observation about challenges for the future. In terms of general direction, the major interests are optimization of the relevant businesses such as machinery maintenance and energy efficiency. DNV-GL expects that the main source of big data is going to be data derived on board vessels and big data is owned, analyzed or traded by various stakeholders including shipowners, ship builders, ship-component manufacturers and classification societies. With respect to challenges, DNV-GL’s observation can be categorized into three fundamental issues, which are sound competitive conditions, technology and human resources.

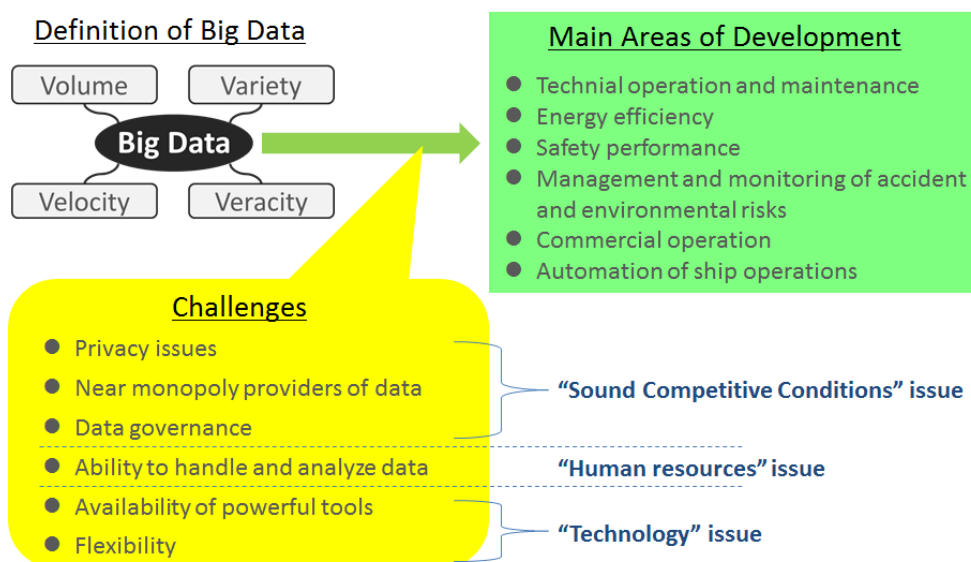


Figure 4: Summary of DNV-GL’s stance regarding big data.

### **3-2 Lloyd's Register Foundation (LRF)**

#### Overview of LRF

LRF is also an institution which is energetically working in the field of big data in the maritime domain. This foundation is young, but contains a long history of one of the most famous classification societies, Lloyd's Register. In July 2012, LRF was established by converting the status of Lloyd's Register from an industrial and provident society to a company limited by shares. LRF is a registered charity and it owns the shares in Lloyd's Register Group Limited as its parent (LRF, n.d.).

LRF's (n.d.) mission is, "to secure for the benefit of the community high technical standards of design, manufacture, construction, maintenance, operation and performance for the purpose of enhancing the safety of life and property at sea and on land and in the air" as well as "the advancement of public education including within the transportation industries and any other engineering and technological disciplines". The main activity of this foundation is awarding grants for research and education and the foundation sets four strategic themes. They are "supporting excellent scientific research; accelerating the application of research; promoting safety and public understanding of risk and; promoting advancement of skills and education" (LRF, n.d.).

#### How LRF observes big data

The issue of big data is covered by "supporting excellent scientific research" and "accelerating the application of research", which are parts of LRF's strategies. These two strategic themes share a funding priority for "emergent technology" which

includes big data, according to “Strategy 2014-2020” (LRF, 2014). LRF published *Foresight review of big data* in 2014 which tells its overall interest with regard to big data (LRF, 2014).

As with DNV-GL, LRF defines big data using the “4Vs”, as already shown in Figure 1 (LRF, 2014). The foundation puts emphasis on, as a background of rapid development of big data utilization, technological developments in terms of generating, storing, processing, understanding and visualizing data, as well as advanced sensor technologies. With this definition and background, LRF is expecting that big data improves “performance, safety reliability and efficiency of assets, infrastructures and complex machines”. Under such an expectation, LRF raises three typical areas of development, as follows (LRF, 2014).

- “Condition-based maintenance”

LRF predicts that “there will be a move away from fixed maintenance intervals, towards tailored predictive maintenance, reducing operator risk and providing better cost-efficiency”. However, it should be kept in mind that considering how to build resilience into a decision support system is important as the network of big data may be heavily affected by some errors or unexpected maintenance stops.

- “Smart factories and autonomous machines”

There are two possible views of the future in this respect shown by LRF. First, “smart materials and products will carry memories of their manufacturing history and plans for how they are meant to be processed

next”. Second, “Advances in autonomous vehicles will impact machine-intense sectors”.

- “Data-enabled prosumers and the quantified worker”

LRF considers that, “personal data can be used on behalf of workers to uncover workload and stress issues”, assuming sensors of vital signs are used. Benefits are not only for the employee side. LRF also states that “employers can use this type of information to measure workforce productivity and identify underperformers”.

Although these ideas are not only for marine-related industries but for a whole engineering industry, they seem deeply associated with the maritime sector. In particular, “smart factories and autonomous machines” and “data-enabled prosumers and the quantified worker” are interesting sights from a different point of view than that of DNV-GL referred to in the previous sub-chapter. The former can be applied to the shipbuilding sector as it treats a tremendous number of materials, machines and vehicles. The latter may be a key for improving safety of shipping drastically because this may be a significant way to reduce the risk of human errors. Taking into account the fact that more than 80% of marine accidents are supposed to be caused by human error (ABS, 2004), developments in this respect are directly associated with improvement of safety of ships.

### Opportunities for LRF

Assuming these possibilities exist in future, LRF states that, in the domain of big data, there are five aspects when the foundation can make contributions. They are

presented as follows (LRF, 2014).

- “Heterogeneous and multi-modal data”

LRF intends to play a role to establish and run an infrastructure by which data from various resources and analytical methods can be shared, exchanged and used. The foundation is trying to become a hub of heterogeneous and multi-modal data.

- “Data analytics”

LRF considers that new methods, tools and technical capabilities for big data analytics need to be developed especially in the field of shipping and energy sectors. The foundation will contribute in this respect cooperating with market players.

- “Interconnected ecosystems”

Thanks to big data analysis, connecting different units of businesses may become possible. For example, marine transport must be affected by such as market of each product to be transported, crude oil market and available port capacity. These “units” of data can be combined to deeply figure out the behavior of the shipping market. LRF targets the promotion of developing such connections which enhance resilience of the supply chain.

- “Data certification”

“Data certification is concerned with metadata that describes what is in the data, who created it, for what purpose, what is the quality of data, and what value arises, among other possible objectives”. The more data are provided to the world, the more attention should be paid for the reliability of such



data. LRF is considering if it can play “a pivotal role in the adoption of data certification and the certification process itself”.

- “Data code and standards”

LRF considers that someone has to take a leadership position over the social capital aspects of big data. In other words, how big data business models are formulated has to be carefully designed in order to avoid allowing certain players to monopolize opportunities of business with big data. The foundation intends to make contributions to this rule-making aspect.

In essence, LRF is trying to take not only advisory roles for establishing associated system or rules of the market but also technical roles for analyzing complex data sets and certifying the quality of such data. The issue of “data code and standards” can be regarded as a part of category of “sound competitive conditions”, which was already derived in a previous sub-chapter as one of the three major kinds of challenges in terms of maritime big data, because it is an issue of controlling the rights and responsibilities of players in the market. On the other hand, all the rest should be categorized as “technology” issues. As regards “Heterogeneous and multi-modal data”, LRF intends to provide expertise on combining, organizing and re-distributing such data sets. Regarding “data analytics” and “interconnected ecosystems”, in short, the foundation is seeking to have a better technology or method of processing big data. “Data certification” underpins technology of generating or processing data, by guaranteeing the quality of such technologies as well as their products. As observed here, LRF’s interest is relatively focused on the aspects of “sound competitive conditions” and “technology”.

### Challenges observed by LRF

LRF presents, in addition to its opportunities, challenges in *Foresight review of big data*, too (LRF, 2014). The challenges basically consist of five dimensions as below.

- Data integration technologies

A sufficient level of technologies has to be achieved; otherwise, incompatible data standards and reference data may cause correlation errors.

- Correlation/causation difficulties

When data analytics are carried out, proper uncertainty quantification is essential in order to avoid deriving significant false-positive results. This aspect requires achievement at a certain level of technology.

- Collection biases

It should not be assumed that derived data is predictive of a larger and unbiased cohort if it is not obvious that obtained data are representative.

- Autonomous machines

There may be vulnerability against cyber-attacks leading to unlawful control of electronic devices or machineries as well as terrorism. Necessary security measures have to be taken into consideration.

- Data quality

Data might be inserted, updated or deleted by unauthorized individuals. An associated security system or framework is necessary.

### Findings from LRF's stance and approach

Taking into account the findings of the previous sub-chapter's discussion, especially three basic categories of challenges, it seems that LRF is focusing more on the

technological aspect than others. “Data integration technologies”, “correlation/causation difficulties” and “collection biases” are considered to be in the category of “technology”. On the other hand, the challenges shown in “autonomous machines” and “data quality” gives a new viewpoint. This is “security”. It seems rational to consider that the benefit of big data can be obtained only on the basis of resistance and resilience against evil actions from outside by assailants. Therefore, adding this new element, four basic challenging areas have been found so far. Consequently, the discussion of this sub-chapter is summarized as in Figure 5.

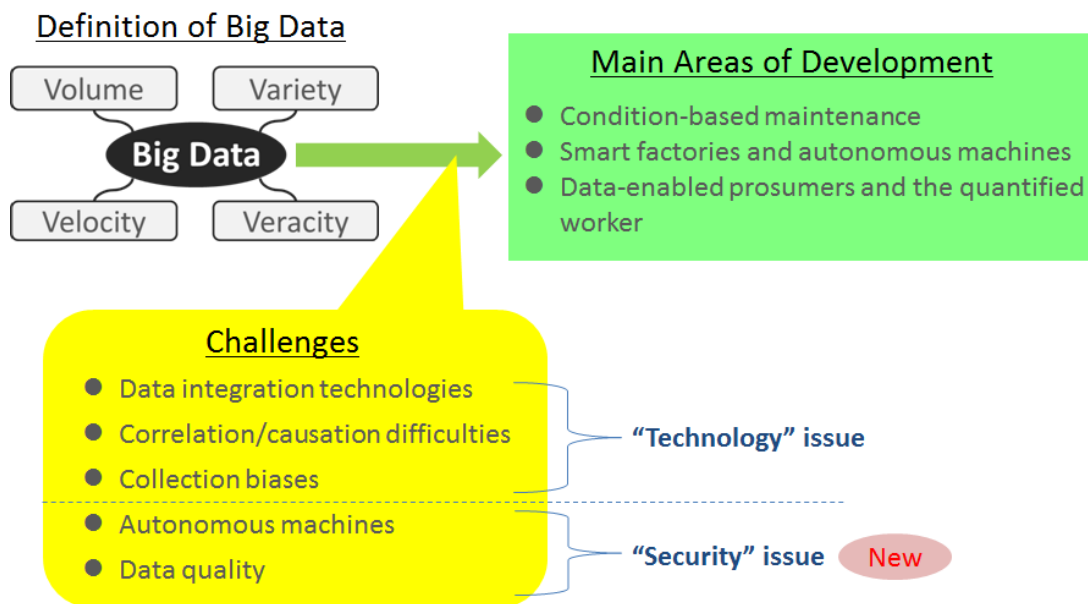


Figure 5: Summary of LRF’s stance regarding big data.

### 3-3 E-navigation

#### Overview of e-navigation

E-navigation is currently one of the most prioritized frameworks in the IMO, the

development of which was proposed in 2006. The definition of e-navigation is “the harmonized collection, integration, exchange, presentation and analysis of marine information on board and ashore by electronic means to enhance berth to berth navigation and related services for safety and security at sea and protection of the marine environment” (IMO, n.d.) As clearly shown in this definition, this framework is heavily depending on “marine information by electronic means”, which should be regarded as a significant piece of maritime big data. In addition, it is a priority of the IMO. Therefore, it is rational and necessary to review the scope, direction, and future prospects of this framework and consider possible collaboration or application to develop big data.

The origin of this framework was found in 2006. In the eighty-first session of the Maritime Safety Committee (MSC) meeting, Japan, Marshall Islands, Netherlands, Norway, Singapore, United Kingdom and United States proposed “to develop a broad strategic vision for incorporating the use of new technologies in a structured way and ensuring that their use is compliant with the various navigational communication technologies and services that are already available, with the aim of developing an overarching accurate, secure and cost-effective system with the potential to provide global coverage for ships of all sizes” (IMO, 2006). The first significant milestone of the framework of e-navigation was presented in the eighty-fifth MSC meeting in 2008, as the committee adopted *Strategy for the development and implementation of e-navigation* (IMO, 2008). It set a general direction for this issue, by presenting visions, core objectives, benefits, basic requirements, potential users and key strategies. Since then, the IMO secretariat and

Member States have worked for this topic through discussions in two sub-committees; NAV and COMSAR. The sub-committees carried out such items as identification of user needs and gap analysis. Finally, in 2014, the MSC adopted *E-navigation Strategy Implementation Plan (SIP)* in its ninety-fourth session (IMO, 2014). In this plan, discussions on e-navigation which had taken place in the MSC for a long time are reflected. Thus, this latest paper is the most reliable document for grasping the current situation of e-navigation. As discussed later in this sub-chapter, this plan set the five prioritized solutions and identified tasks which are necessary to achieve such solutions. Recently, the issue of e-navigation still remains in the High-level Action Plan, which is a plan of implementing high-priority issues of the IMO, according to the decision made at the ninety-fifth session of the MSC (IMO, 2015).

With regard to e-navigation, the Danish Maritime Authority has been developing the project titled “Maritime Cloud” (Christensen, 2014). In IMO, this is defined as “a communication framework enabling efficient, secure, reliable and seamless electronic information exchange between all authorized maritime stakeholders across available communication systems” and regarded as a component of e-navigation (IMO, 2014). However, as found in this definition, the framework is dedicated to the improvement of communication. This means that, in terms of range of interest, big data and the maritime cloud share only a small area. Therefore, it was necessary to exclude the review or analysis of the maritime cloud from the discussion in this dissertation.

### Scope of e-navigation

The scope and general direction of the framework of e-navigation can be found in SIP as well as *the Strategy for the development and implementation of e-navigation*. The latter states “The overall goal is to improve safety of navigation and to reduce errors. However, if current technological advances continue without proper co-ordination there is a risk that the future development of marine navigation systems will be hampered through a lack of standardization on board and ashore, incompatibility between vessels and an increased and unnecessary level of complexity” (IMO, 2008). The whole architecture which the e-navigation framework is trying to achieve is presented in SIP, as Figure 6 (IMO, 2014).

There are both similarities and differences between the scope of e-navigation and that of big data. As regards similarities, both are highly depending on the electronic devices on board vessels. Thus, the developments associated with e-navigation may naturally contribute to a better utilization of big data from ships. On the other hand, regarding difficulties, e-navigation basically does not pay much attention to the accumulation of data and analyzing it after voyages for such purposes as research and development. E-navigation mainly focuses on safety and the efficiency of navigation during voyages and is interested in real-time, streaming data. In addition, one of the most important goals of e-navigation is to enhance and optimize the human-machine interface, as shown in Figure 6, in order to improve the bridge environment for mariners' communication. This is a different point of view from big data. Taking into account such a difference of scope between e-navigation and big data, this sub-chapter is aimed at detecting the similarities and possible contributions

of e-navigation to develop big data utilization.

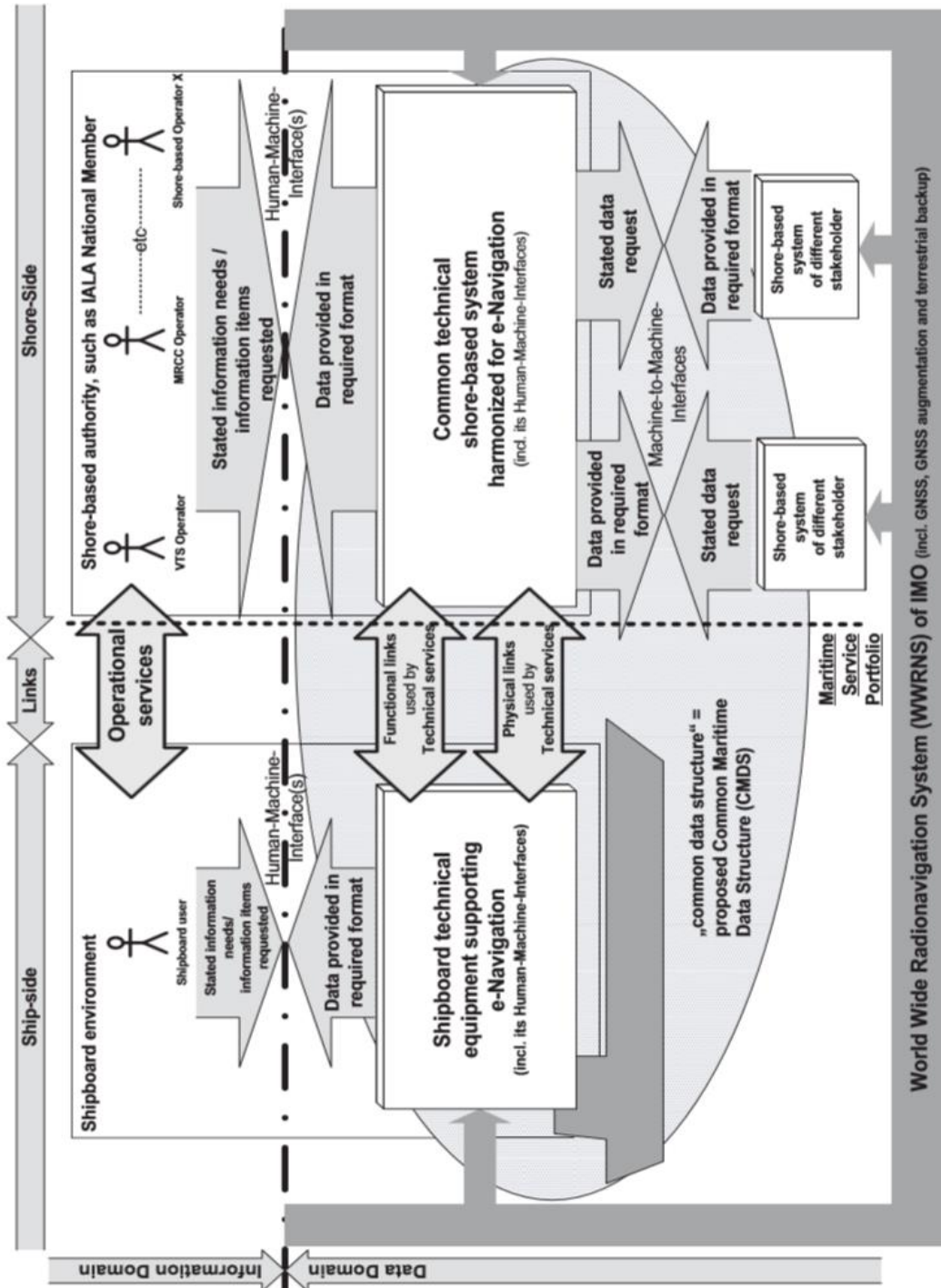


Figure 6: Overarching e-navigation architecture. Reprinted from *Draft e-navigation strategy implementation plan*, by IMO (NCSR 1/28)

### General direction of e-navigation

As already mentioned, SIP identified five prioritized e-navigation solutions. They are small goals to be achieved for reaching the absolute goal of e-navigation (IMO, 2014). They are as below.

- S1: Improved, harmonized and user-friendly bridge design

This solution is basically aimed at increasing the user-friendliness of bridge equipment, and dedicated to better working conditions of seafarers and takes care of items such as how relevant information should be displayed.

- S2: Means for standardized and automated reporting

The standardized and automated reporting system reduces unnecessary burden on seafarers. For this purpose, such as the establishment of automated voyage data collection and standardized digital reporting formats are attempted.

- S3: Improved reliability, resilience and integrity of bridge equipment and navigation information

This solution consists of enhancing the quality of relevant equipment and software which are used for navigational operations and increasing the reliability of voyage-related information referred to throughout navigation.

- S4: Integration and presentation of available information in graphical displays received via communication equipment

The key of this solution is to find an optimal way of displaying information which is necessary for safety of navigation on the bridge. For this purpose, such as the “Common Maritime Data Structure (CMDS)”, standardized



interfaces for data exchange and mapping of specific services on monitor are attempted. The quality assurance process to ensure reliability of data is also considered under this solution.

- S9: Improved communication of Vessel Traffic Service (VTS) Service Portfolio  
This solution seeks for better methods of communication. As the central interest of this solution is human communication, technology of treating data, such as recording and processing, is not covered in this issue.

According to the NAV Sub-Committee (IMO, 2013), prioritization was done taking into account two criteria, which are “1: seamless transfer of data between various items of equipment on board” and “2: seamless transfer of electronic information/data between ship and shore and vice versa and from ship to ship and shore to shore”. As both of these concern electronic data on board ships, selected solutions should have something to do with maritime big data. Therefore, these prioritized solutions are worth analyzing.

In SIP, there are so-called “sub-solutions” given under each prioritized solution and, according to those sub-solutions, S1 and S9 seem to have nothing to do with big data. The interest of S1 is not storage and usage of data but designing a user-friendly bridge. S9 is an issue of how seafarers and personnel in shore-based facilities can communicate effectively.

However, the rest of the solutions partially include certain associations with big data. For example, S2 contains a sub-solution titled “automated collection of internal ship data for reporting” (IMO, 2014). Literally, the purpose of this sub-solution is

efficient reporting. However, the technology of “automated collection of data” can be a help for easier and more effective accumulation and exchange of data. S3 includes an issue of “software quality assurance” (IMO, 2014). This is directly related to the reliability of the data management system on board. In addition, this solution also attempts to improve the reliability and resilience of position, navigation and timing information. These approaches are expected to contribute to enhance the quality of data derived on board. Under S4, “Common Maritime Data Structure” is being considered as mentioned above. This is a kind of uniform format of data set and is supposed to be a basis of effective data exchange and communication. This effort should be welcomed from the point of view of utilizing big data, too. As exemplified here, prioritized solutions of e-navigation have some directions of development which can be shared with those who are interested in maritime big data. Thus, it is useful to review the challenges treated in the discussion of e-navigation, focusing on S2, S3 and S4.

#### Challenges relevant to e-navigation

In terms of challenges of the e-navigation framework, MSC has identified eighteen tasks to be cleared (IMO, 2014). As the discussion of e-navigation has been taking place for a long time since 2006, the challenges have been broken down into “task” levels. Taking into account the discussion on S2, S3 and S4, four tasks which seem to be related to big data utilization are detected from SIP (IMO 2014). They are as below.

- “Investigating the best way to automate the collection of internal ship data for reporting including static and dynamic information”

As already discussed, automated data collection can be a significant help not only for ship reporting but also for better utilization of big data on board vessels. In MSC, it is planned to draw up a technical report on the automated collection of internal ship data for reporting by 2016. As this is just a “technical report”, actual development and endorsement of the new system is supposed to take longer time. However, the time frame does not matter here. The fact to which attention has to be paid is that e-navigation is moving toward the development of an automated data collection system.

- “Developing draft guidelines for Software Quality Assurance (SQA) in e-navigation”

This is an ongoing discussion on how the quality of software which deals with data associated with navigation should be ensured and proved. According to the Korean delegation to IMO, which is the main proposer of SQA, SQA is regarded as “a set of activities designed to evaluate the process by which the products are developed or manufactured” (IMO, 2013). In other words, the criteria for evaluating the quality of relevant software are needed. MSC is planning to develop the corresponding guidelines.

- “Developing guidelines on how to improve reliability and resilience of position, navigation and timing systems on board by integration with external systems”

This is an issue not about software which deals with data but about data in itself. As a position, navigation and timing is critical information for the safety of shipping, the MSC is trying to enhance the reliability and resilience of these data. The committee intends to develop relevant guidelines by 2016.

- “Developing a CMDS and further the standardized interfaces for data exchange used on board”

In the strategy paper of e-navigation (IMO, 2008), it is stated that “information should be provided in an internationally agreed common data structure”. As such a structure, it was agreed that the “S-100” data model should be used. This model has been developed by the International Hydrographic Organization (IHO). As with the data model, interfaces for data exchange used on board are also being standardized.

The four tasks raised above out of the eighteen presented in SIP should be considered from the point of view of maritime big data as they are potentially beneficial in this respect. First, the task of “automate the collection of internal ship data” gives a good insight for the further development of maritime big data. Automated data collection tools make it efficient to obtain and store voyage-related data as maritime big data. Second, the task of “SQA” shows that it is important to ensure a certain level of reliability in terms of software which treats voyage-related information on board vessels. This idea can be applied to big data, too. Third, the task of “reliability and resilience of position, navigation and timing systems” is directly related to big data. Out of the four general difficulties of treating big data, “veracity” embraces this challenge, as already discussed. Fourth, the task of “CMDS” illustrates the importance of having a uniform data structure in order for easier and more efficient data exchange. When these four tasks are observed in the light of maritime big data, they should be categorized as issues of “technology” out of the four aspects of challenges obtained through studying DNV-GL and LRF.

## Findings from the review of e-navigation

As already mentioned, the scope of e-navigation is not the same as that of big data. However, it may be possible to share some interests and create synergy between both domains at least in terms of the four aspects discussed here. The most important point is the fact that there is potential for collaboration. An example of possible collaboration is discussed later in sub-chapter 4-3. Consequently, the findings regarding e-navigation can be summarized as Figure 7.

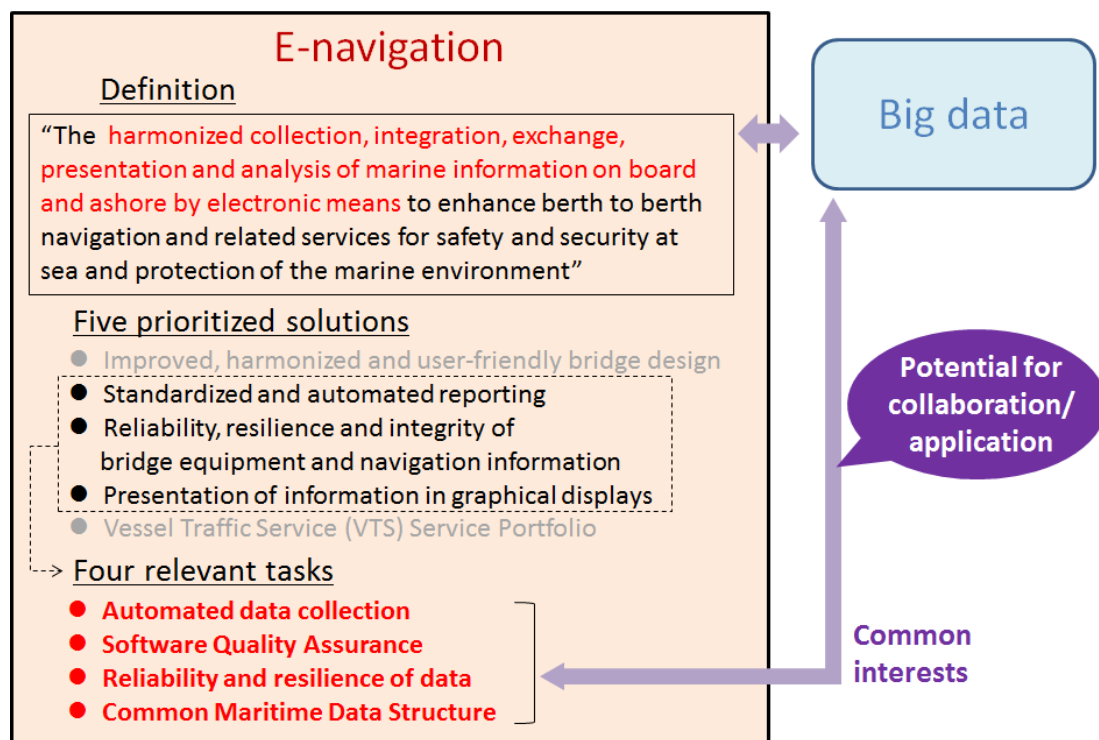


Figure 7: Summary of the consideration of e-navigation.

## 4 Analysis of challenges and solutions

Through the review of DNV-GL and LRF, four important aspects of the challenges have been identified; sound competitive conditions, human resources, technology and security. These categories and the concrete issues in each category can be rearranged as in Figure 8. In this chapter, based on these findings, each category of challenges in terms of maritime big data is analyzed in order to detect the core of the problem. Then, possible solutions for solving detected problems are considered.

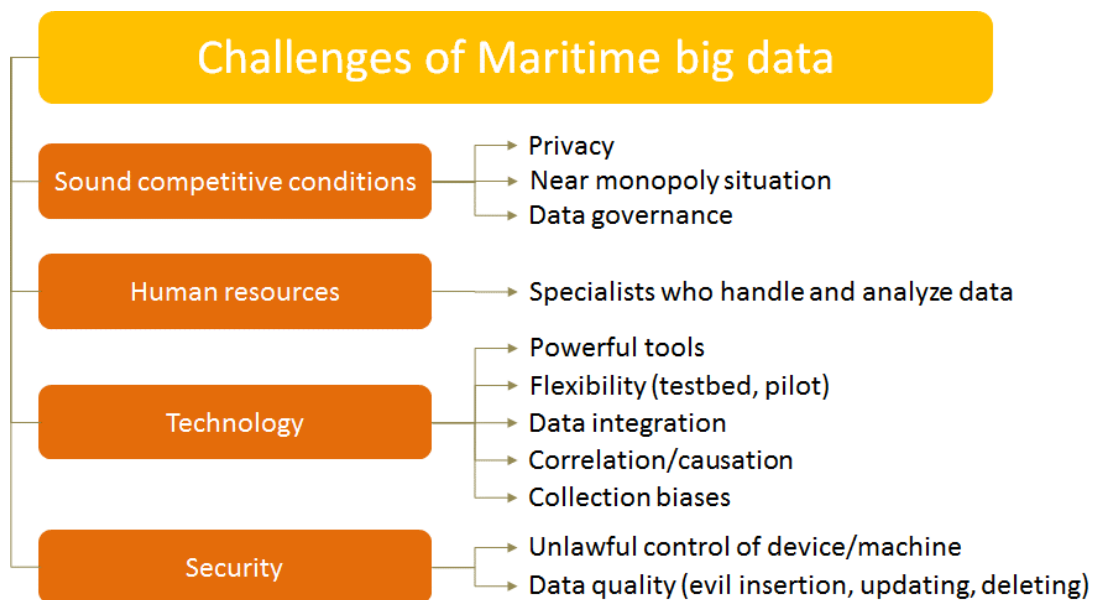


Figure 8: Identified challenges for development of maritime big data.

#### **4-1 Sound competitive conditions**

The first category of the challenge is the sound competitive conditions. The privacy issue, near monopoly of data providers and data governance have been detected in the previous chapter as elements of this category. As regards the privacy issue, for example, personal communication on the bridge may have to be properly protected. With respect to the monopoly problem, if rights of certain kinds of data are monopolized, the number of stakeholders who cannot feel any incentive to exchange data may increase and such tendency prevents the sound development of the use of maritime big data. In terms of data governance, it should be noticed that, if different kinds of data are concerned, the stakeholders who should have property rights may differ. For example, it is expected that the right holders of engine performance and hydrographic data are different. Then, how can these challenges be refined? In other words, what is the core problem of this category?

##### What is the core of the problem?

These pieces of the challenges reviewed above can be integrated into one general problem. This is the absence of “rules of the game” in terms of balance of rights and responsibilities of the holder or acquirer of data and secondary user. If there is no set of rules to control this issue, many relevant entities may be prevented from sharing or exchanging big data with others because of the fear of effluence of their important business-related information and losing their profit and business opportunities.

Actually, this theme is referred to as a very important issue of big data. McKinsey (2011) points out the importance of this issue. It states, “Advances in storage and

analytics will continue as organizations increasingly need to shore ever greater amount of data, access and analyze them, sometimes in near real time. These innovations require an effective intellectual property system, which will both ensure incentives for creating valuable data and enable the effective sharing and integration of different pools of data”. The significance of this problem is also pointed out by marine-related institution, too. For example, DNV-GL considers that well-established governance of data is one of the fundamental issues of big data, stating that good solutions have to be built into the system from a very early stage of development in this field (DNV-GL, 2014). As illustrated in these examples, the problem of absence of obvious rules in terms of the rights and responsibilities on treatment of big data should be referred to as a kind of prerequisite for developing the use of maritime big data.

#### How to address the problem

In order to address this problem, it seems useful to divide the discussion into two patterns as below.

- Pattern A: Data shared and exchanged by some specific stakeholders for certain collaborative activities

This pattern assumes local cooperation. A possible example is a collaboration project among a ship building company, a shipowner and a ship machinery company for improving efficiency of fuel oil consumption of designated series of vessels. It is expected that, in this kind of project, relevant data such as ship speed and order and performance of the engine need to be shared and jointly analyzed. However, as a matter of course, they



do not want to open their private data and consequence of research to other competitors.

- Case B: Data opened and shared by all concerned parties

This pattern is concerned with open data, such as AIS data, which is available for every interested entity. This pattern also includes cases in which an institution willingly opens their private data to the public.

For each of these two patterns, how to control the rights and responsibilities regarding such as ownership of and access to big data is considered and, consequently, a possible solution is proposed for each pattern introducing some model cases to be consulted.

Data shared and exchanged by some specific stakeholders

As this case is just a local cooperation by a limited number of companies or institutions, establishing local rules to be promised by such entities should be a basic approach. There is no single absolutely right answer as to how rights and obligations should be organized because the answer depends on the condition of agreement to be made by members of the collaborative activity. In this context, what they should consider and how they can develop such agreements should be instructed or supported by an even-handed third party. This approach seems rational because most of marine-related stakeholders are not originally specialists of data engineering. They should need guidance for developing agreements or contracts on who have ownership and access to the data.

In this respect, a very interesting model case has been already established in the

agricultural sector in the U.S.A. According to Kassner (2014), “The agriculture community is being dragged into the big-data era by Agriculture-Technology Providers (ATP)”. Farmers are interested in big data for getting over the uncertainties inherent to farming. However, they were initially “hesitant to lose control of their data and want assurances before they sign contracts”. The American Farm Bureau Federation (AFBF) tackled this problem and got an answer. They have developed a document called “Privacy and Security Principles for Farm Data”. The function of this document is to support establishing a set of fundamental terms and conditions when an individual or organization is to share data with other entities. There are thirteen principles presented in this document and Kassner introduces the following three principles as representative ones (Kassner, 2014).

- Ownership

“Farmers own information generated on their farming operations. However, it is the responsibility of the farmer to agree upon data use and sharing with the other stakeholders. The farmer contracting with the ATP ensures that only the data they own or have permission to use is included in the account with the ATP”.

- Collection, Access and Control

“An ATP's collection, access, and use of farm data should be granted only with the affirmative and explicit consent of the farmer. This will be by contract agreements whether signed or digital”.

- Transparency and Consistency

“ATPs shall notify farmers about the purposes for which they collect and use

farm data. They should provide information about how farmers can contact the ATP with any inquiries or complaints, the third parties to which they disclose the data, and the choices the ATP offers for limiting its use and disclosure”.

By signing the document, ATP will be bound by these principles and farmers will obtain reasonable assurance about rights over their farm data.

Now, when the “farmer” is replaced with “initial owner of marine-related data”, such as shipowner, and “ATP” is replaced with “maritime service provider” which is represented by a classification society or consulting agency, it turns out that the structure of this framework is applicable to the maritime industry. Therefore, it seems useful to develop a similar framework for maritime big data. In order to achieve this, as with the example of the U.S. farm industry, the public sector has to contribute at the governmental level. Namely, government should develop a set of recommendatory guidelines. Of course, the possibility of compulsory instruments cannot be denied at all. However, as of now, there is no international regulation which is applicable to all kinds of industry and legally binding ownership of electronic data. Thus, the reasonable approach is to start with a recommendatory framework.

As discussed so far, these guidelines are expected to provide advice on how relevant stakeholders should make an agreement on the rights and responsibilities on the treatment of data in question. Then, how can it be achieved? There are three main steps. First, the government establishes a working group which is going to draft the

guidelines. In this working group, specialists from various areas of expertise need to be involved. For example, lawyers, classification societies and representatives from relevant maritime industries are supposed to attend. Second, in the working group, the study and analysis of existing models, such as AFBF's case, are carried out. The group considers the necessary descriptions to be articulated in documents of agreement on data sharing and exchanging. The output from this group is a set of draft guidelines. Then, the government finally publishes these guidelines as legally non-binding instruments.

However, the form of output is not limited to the guidelines. Actually, the output of the example of AFBF is a kind of “memorandum of understanding (MOU)”. Even if this form is adopted, the basic steps are the same. Instead of developing the guidelines, the working group develops a draft MOU and the government publishes or authorizes the concluded MOU. Here, in case of MOU, there is an additional step. After publication or authorization of the MOU, it needs to be signed by relevant maritime entities for ensuring that they comply with it. Then, the MOU comes into effect among signers.

The most important benefit of this solution is to overcome the bottleneck of vague rights and responsibilities on treating data within certain groups with a limited number of stakeholders. Thanks to these guidelines or MOU, the relevant entities can move forward to the innovation of maritime big data under clear-cut rules. In other words, it is expected that the proposed solution enables more and more companies and institutions to feel free to launch innovative collaborations on maritime big data

without being concerned of the ambiguity of rights and responsibilities on data.

Data opened and shared by all concerned parties in the public domain

In terms of data, which is opened and shared in the public domain, the root of the problem is the same as that of “data shared and exchanged by some specific stakeholders”, namely, the ambiguity of rights and responsibilities of relevant stakeholders. However, there is a significant difference in terms of the feature of the problem from those discussed in the previous part. The key term is the “unspecified number” of entities to be involved. As far as unspecified number of stakeholders are involved, it is not reasonable to adopt guidelines or MOUs for instructing local contracts because it is impossible to make all of the relevant entities to explicitly agree to a single contract or sign a single MOU. Furthermore, nobody knows who are the “all of relevant entities”. Therefore, this category needs standardized criteria or rules which apply to everyone and such a common rule should be provided by the government. In short, the government needs to develop an equal and reasonable standard of rights and responsibilities on the treatment of open and public maritime big data.

In order to address this problem, it is useful to refer to the existing instruments of license concerning open-data. James (2013) expressed the definition of open-data as “data that can be freely used, shared and built-on by anyone, anywhere, for any purpose”. This definition seems useful to grasp what is open-data. According to Research data management service group of Cornell University (n.d.), there are at least two famous institutions which have been developing tools to govern the use of

such open-data. One is the Open Data Commons group and another is Creative Commons. The group of Cornell University (n.d.) says that tools introduced by these institutions enable the description of a set of “community norms”.

First, the Open Data Commons group (n.d.) advocates three kinds of standard licenses. They can be roughly summarized as below.

- Public Domain Dedication and License

The database and its contents are in the public domain and free for everyone to use.

- Attribution License

The database and its contents are free for everyone to use, provided that the user shows attribution to the source of data.

- Open Database License

Any secondary use of the database must show attribution and the new product created from such data must be accessible. In addition, any new products made by this material must be distributed using the same terms.

This is the most restrictive option.

On the other hand, Creative Commons (n.d.) introduces two kinds of standardized licenses.

- CC Zero

When an owner of the database willingly waives copyright, this owner can use the CC Zero mark. It explicitly places the database or data into the public domain. Public Domain Dedication and License which is introduced by the Open Data Commons is functionally equal to CC Zero.

- Public Domain mark

If the data or database for which there are no known copyright or restrictions, this mark can be used. Creative Commons considers that, for example, factual data can be flagged with Public Domain mark so as to make it clear it is free to use.

Cornell University (n.d.) points out that, out of those introduced above, there are two most reasonable choices in general, stating that “anything other than an Public Domain Dedication and License or CC Zero license may cause serious problems for subsequent scientists and other users”. This is because a possible mess of source acknowledgement occurs. For example, it is quite difficult to follow the restriction given by each source of data if a product is created from hundreds of data sources. This is going to be a disincentive for future developers of maritime big data tools.

Taking into account existing standardized licenses as well as their features, the government should develop the country's common rules of rights and responsibilities on the treatment of open maritime big data in the public domain. The actual steps are basically the same as those in the previous part of the discussion. First, the government establishes a working group which consists of relevant stakeholders, such as shipowners, ship building companies, machinery manufacturers and experts of law, and the group drafts the new rules for open maritime big data taking into account the existing models as introduced above. Then, the government finalizes the concluded draft of the working group and publishes to make it come into force. In terms of the status of this new rule, the government has two options; mandatory

instruments or recommendatory ones. As this rule is equally applied to every entity, both options are expected to work in a fair manner. In either case, everyone is under the control of the same rule. Thus, the government has to consider in this respect taking into account such relationships with other relevant national rules which already exist.

#### Summary: Sound competitive conditions

In this sub-chapter, it turned out that the challenge of sound competitive conditions is an issue of rights and responsibilities on the treatment of maritime big data. It is necessary to have a common norm in this respect. This problem can be divided into two patterns. One is the local and closed partnership among the limited number of entities. Another is sharing data with an unspecified number of entities in the public domain. In both, the government is expected to take a significant role. For the first pattern, the government needs to develop a set of guidelines or corresponding MOU to provide the maritime industry with instruction on how to make an agreement on associated rights and responsibilities. For the second pattern, a common rule or guideline of treatment of open big data, which is applied to every entity involved in the use of such data, is needed. By establishing these frameworks, it is expected that the fair and smooth development of the use of maritime big data is promoted. The summary of this sub-chapter is shown in Figure 9.



## Sound competitive conditions

Problem = Absence of obvious rules in terms of rights and responsibilities on the treatment of big data

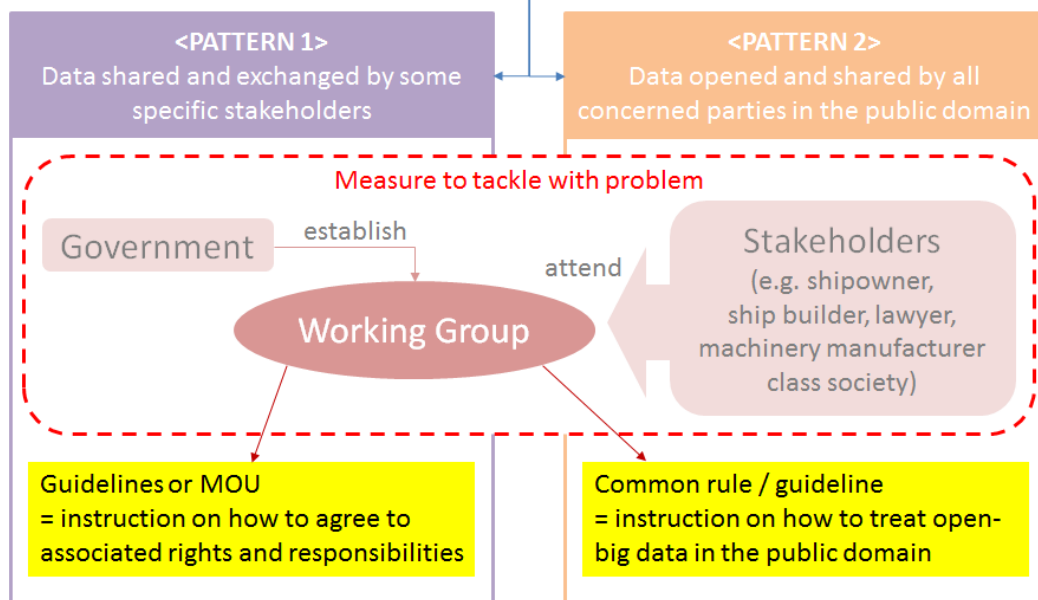


Figure 9: Summary of challenges and solutions regarding sound competitive conditions.

### 4-2 Human resources

Human resources with certain expertise for dealing with big data, which may be called “data scientist”, is obviously important because no data-related software and hardware can be developed or properly operated without these specialists. The functions to be fulfilled by data scientists were explained by Inter (2013) who stated “Data scientists are responsible for modelling complex business problems, discovering business insights, and identifying opportunities”. It also says that major benefits to be brought by data scientists are “skills for integrating and preparing large,

varied data sets”, “advanced analytics and modelling skills to reveal and understand hidden relationships”, “business knowledge to apply context” and “communication skills to present results”.

The importance of data scientists is not only in the character of their jobs but also in the quantity of the work. According to DNV-GL (2014), “preparing data for analysis is very labor-intensive, typically accounting for as much as 80% of the time spent in projects. But without it, results simply cannot be trusted”. Thus, it is safe to say that finding and hiring skilled personnel is certainly one of the most important aspects associated with the development of big data utilization. In other words, ensuring enough quantity and quality of human resources is inevitable for developing the use of maritime big data.

#### What is the core of the problem?

The importance and necessity of such human resources are understood, but, what is the actual problem to be tackled in this respect? In essence, the problem is that finding or hiring such well-qualified personnel is getting more and more difficult. This problem seems more severe for the maritime industry. Statistics, which is one of most typical areas of expertise related to big data, can be raised as an example. With regard to this domain, Hal Varian, chief economist at Google stated that “the sexy job in the next 10 years will be statisticians” (Lohr, 2009). In addition, McKinsey (2011) raises an example of the shortage of human resources to work as a specialist of big data. It states “The United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze

big data and make decisions based on their findings”. The example of Hal Varian implies that not only the maritime industry but also many other industries which intend to take advantage of big data will desire to catch more and more engineers in the field of big data. Such tendency justifies the shortage of human resources observed by McKinsey. Industries will face the difficulty of finding the specialists of data-related engineering.

The foresight of the shortage of human resources is a general tendency of society and applicable to all relevant industries. Thus, even the information-intensive industry, which includes companies such as IBM and Google, may also suffer from this shortage. This means that the maritime industry, which is not originally so information-intensive and does not have much experience and know-how of nurturing such specialists, suffers more. In short, taking into account the situation of the labor market in the field of big data, maritime stakeholders have to make intensive efforts to get enough numbers of engineers with sufficient expertise on treating big data.

#### How to address the problem

Who has to tackle this problem? Without doubt, the private sector, including shipowners, ship builders and machinery manufacturers, will make the effort as much as they can, based on the economic incentives. These players accept paying out more money for specialists of big data as long as they can generate enough profit. However, efforts made by the private sector are not enough because of the foreseen shortage of human resources which is a structural problem. In other words,

regardless of how much money the private stakeholders pay, the gap in the demand and supply in the big data-related labor market does not disappear; the problem remains. Therefore, certain measures need to be taken by the public sector at the government level. The measure has to be aimed at increasing the number of data scientists which meet the requirement of the industry. To be more precise, governments should improve and expand their national education systems.

Some countries which have noticed such needs have already been making changes. The U.S.A. is one of them. In November 2009, the President launched an initiative for the innovation of education, aiming at moving “American students from the middle to the top of the pack in science and math achievement over the next decade” (The White House, 2013). The government of the U.S.A. is still intensively supporting science, technology, engineering, and mathematics (STEM) education. One of the most significant achievements of the U.S.A. is publishing strategy on STEM education (Committee on STEM education, 2013). There are five capital objectives in the strategy as below.

- To prepare 100,000 well-qualified new STEM teachers by 2020, and support the existing STEM teacher workforce
- To achieve a 50 percent increase in the number of U.S. youth who have an authentic STEM experience each year prior to completing high school
- To graduate 1,000,000 additional students with degrees in STEM fields over a decade
- To increase the number of students from groups that have been underrepresented in STEM fields that graduate with STEM degrees in the coming decade and

improve women's participation in areas of STEM where they are significantly underrepresented

- To provide graduate-trained STEM professionals with basic and applied research expertise, options to acquire specialized skills in areas of national importance, mission-critical workforce needs for the Committee on STEM education agencies, and ancillary skills needed for success in a broad range of careers

Under these clear and quantified objectives, various frameworks are being run by the government of the U.S.A. The first example is “Investing in innovation Fund”. According to the U.S. Department of Education (n.d.), this project is “to provide competitive grants to applicants with a record of improving student achievement and attainment in order to expand the implementation of, and investment in, innovative practices that are demonstrated to have an impact on improving student achievement or student growth, closing achievement gaps, decreasing dropout rates, increasing high school graduation rates, or increasing college enrollment and completion rates”. In 2015, a budget of around 112 million US dollars is allocated to this project. The objective of this fund is not limited to STEM, but, the STEM education strategy states that “STEM initiatives have been a competitive priority” in this fund (Committee on STEM education, 2013). The second example is the “Mathematics and Science Partnership”. This program is aimed at supporting the improvement of mathematics and science education through partnerships, which include high-need local education agencies and the mathematics, science, or engineering department of an institution of higher education. In 2014, the government allocated 150 million U.S. dollars to this project.

It is recommended that, taking into account these examples, governments which intend to enhance the development of maritime big data in their countries promote the cultivation of human resources by reference to these models. However, the introduced examples of the U.S.A. are not maritime specific solutions but just general ones. Thus, further consideration for solutions which are specified to the maritime industry is necessary. This dissertation proposes two possible answers as below.

#### Developing a strategy for human resource development

The first solution simply follows what has been done by the U.S. government; provide a strategy for developing human resources for maritime big data. Corresponding government institutions, such as the maritime administration, should firstly present a future vision. In other words, the goal for the whole of the maritime industry in terms of human resource development needs to be identified as a starting point. As with the example of the U.S strategy, the objectives have to be clear and quantified. The strategy has to show the number of specialists to be nurtured, kinds of expertise to be developed and when such objectives have to be achieved. These objectives should stand on the basis of the present survey and analysis of the associated labor market. Then, taking into account these objectives, the relevant measures to be taken by the public sector should also be introduced in the strategy in order to provide a bridge from the current situation to the future vision. If there is such a strategy shown by the government, stakeholders in the maritime private sector can plan and decide their direction of investment and development taking into account the existing challenges and what kind of support is going to be provided by

the government to get over them.

#### Industry-government-academia collaboration

Another recommendation is a framework of industry-government-academia cooperation for education at university level. The aim of this framework is to enable existing universities to provide the maritime industry with data scientists at the designated level of expertise by establishing a dedicated specialization course in universities.

The structure of the framework is shown in Figure 10. To begin with, the government establishes a joint committee in which the government, designated university and private stakeholders work together. The purpose of this committee is to establish a new specialization course in the university which is dedicated to nurturing human resources with expertise on maritime big data. In order to design the new course, each participant contributes to this committee. The university brings academic expertise such as ship engineering and information technology to the committee, while the private sector provides information relevant to the ship and its equipment, concrete needs for human resources and knowledge of each field of business. The government may also contribute by funding a pilot course in the university, as well as subsidizing relevant studies or research carried out under the committee.

Establishing this framework generates at least two kinds of significant benefits. First, the quality and quantity of the associated human resources will be improved and, as a consequence, structural problems of the labor market, which has a severe shortage of qualified data scientists, will be eased at least in the maritime field. Second, it is

expected that the gap between demand and supply in terms of expertise or ability of human resources is going to be minimized as the educational side can fully understand the needs of industry and establish tailor-made academic courses.

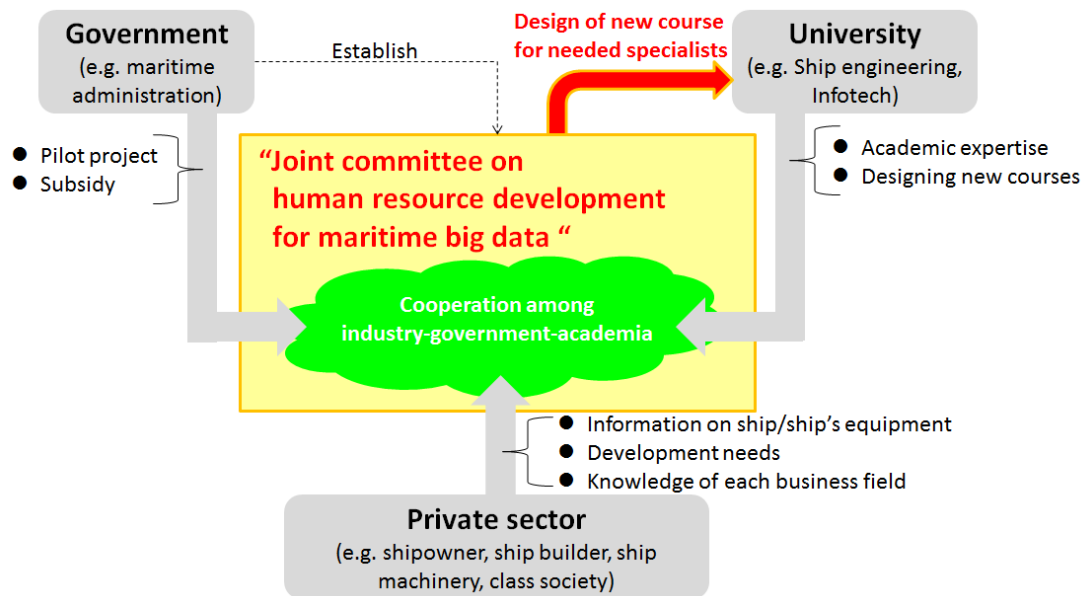


Figure 10: Structure of framework of industry-government-academia cooperation.

Summary: Human resources

With respect to human resources, main problem is the difficulty in getting enough specialists who are sufficiently qualified in terms of maritime big data despite the coming shortage of such workers in the labor market. By referring to existing examples, two possible measures were proposed. One is developing a national strategy for nurturing associated human resources. Another is establishing a cooperative framework among the government, university and private sector. The former is expected to encourage and support the decision of investment to be made



by stakeholders in the private sector. The latter eases the constant shortage of human resources related to big data and contributes a matching demand and supply in the labor market.

### **4-3 Technology**

The third challenge to be considered is technology. With regard to this aspect, “availability of powerful tools” and “flexibility” have been raised by DNV-GL. On the other hand, LRF introduced “data integration”, “correlation and causation” and “collection biases”. As a brief review of the previous chapter, these components of technology-related challenges are summarized as below.

- Availability of powerful tools

- The equipment or system with more capacity and capability needs to be invented and produced.

- Flexibility

- Test-bed for experimenting and exploring new tools in a safe environment is necessary.

- Data integration

- Standardized format or structure of data may be helpful for further developing utilization of big data.

- Correlation/causation difficulties

- When data analytics are carried out, such as proper uncertainty quantification is going to be essential.

- Collection biases

It should not be assumed derived data is predictive of a larger and unbiased cohort if it is not obvious that the data are representative.

#### What is the core of problem?

As an overview of these five elements, it turns out that there is a fundamental and common point. Out of the five elements of challenge, “powerful tools”, “correlation/causation” and “collection biases” can be lumped together into inventing new electronic equipment with higher performance and upgrading technology of information-processing. The rest of the two are considered later on. It is obvious that such an invention and upgrade are essential for the development of big data because the data is stored, processed and analyzed through these tools. In other words, the performance or achievement of utilizing big data directly depends on the capacity and capability of relevant tools. When the better tool is used, the more innovation can be achieved.

#### How to address the problem

The problem has already been refined. Then, how should such better tools be developed? Basically, it is going to be done by each company or research institution. They develop the associated equipment and information technology based on the principles of market competition. In fact, there are such fore-running companies. One of the typical examples is Interschalt maritime systems AG. It produces VDRs and the corresponding software with advanced information technology. The product called “VDR-G4/G4e” is a VDR which makes it possible to access to the recorded data in real time. Besides, “Maritime Data Engine”, which can be installed on board

ships or on shore side, enables to receive data from one or more VDR's "in order to convert them and send them out on multiple ports with different data structure" (Interschalt, 2015). As already discussed in the sub-chapter 2-2, electronic data stored in VDR has a vast potential to be used as maritime big data. This company has had such insight and developed new products. As illustrated by this example, the private sector naturally tries to develop the technological level in terms of big data as much as it can.

As those who develop the relevant tools are private entities, the role and contribution of the public sector should be auxiliary. Here, the structure of technology-related challenges and approaches to cope with this issue can be summarized as in Figure 11. Basically, the activities of private companies are implemented based on their own determination and investment. Therefore, it is not reasonable to try to control their behavior by presuming their behavior or forcing them to do something. However, some complementary support can be provided by the public sector or other third-party entities to encourage stakeholders in the private sector to promote the development of tools related to maritime big data. In this respect, such as providing a test-bed and funding pilot project, which were regarded as issues of "flexibility", are categorized as such complementary support as in Figure 11. In addition, "data integration", which means the standardization of the format or structure of data, can be facilitated by governments or other third-party institutions. Thus, it is also categorized as an issue of "support". Therefore, noting the structure of challenge in Figure 11, solutions to enhance these "supports" is focused on in the discussion below. Two aspects, namely "flexibility" and "data integration", are considered.

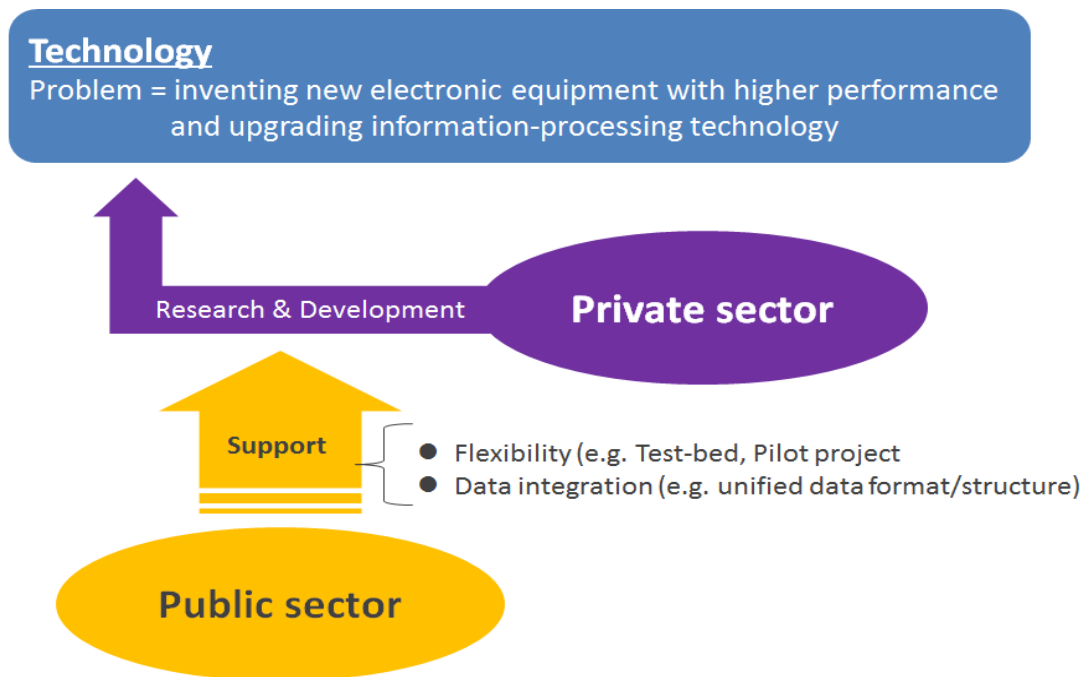


Figure 11: Summary of challenges and approach to them in terms of technology.

### Flexibility

As already mentioned, the problem of “flexibility” is represented by the needs for the provision of a test-bed and funding for the pilot project. The necessity for such support is pointed out by DNV-GL (2014), as discussed in the previous chapter. In addition, McKinsey (2011) states that “policy makers can play a role in helping to overcome technology issues related to the use of big data”. McKinsey also recognizes the “helping role” should be taken up by the public sector.

In order to get an idea on how to support the development of technology regarding big data, EU's scheme for research and development can be seen as a model case. In the EU, there is a huge framework which provides various types of support for the

development of science and technology. This framework was titled the Seventh Framework Program for Research (FP7). According to the EU, there are two major strategic objectives of FP7. They are “strengthening Europe’s scientific and technological base and supporting its international competitiveness and the EU policies, through research cooperation among Member States and with international partners” (European Union, 2015). This framework started in 2007 and final year of adopting projects was 2013. In FP7, projects are selected and funded through the open application scheme and the total budget throughout the implementation period is more than 50 billion Euro (European Commission, n.d.).

Under this general framework of science and technology development, there is a good example of a maritime project in which a data-related test-bed has been provided. Maritime Unmanned Navigation through Intelligence in Networks, which is the so-called MUNIN, is a project which has been funded under FP7. The objective of this project is to develop technology which enables operating unmanned vessels during deep-sea-voyages. It is stated that “the vision of MUNIN is to show the update possibilities of today’s fleet to autonomous vessel” (European Commission, n.d.). According to the European Commission’s database (n.d.), the project has been run by a group of eight institutions from five European countries, headed by Fraunhofer CML. The project started in August 2012 and is planned to conclude in 2015. For the three years, around 2.9 million Euro of budget was provided through FP7 framework. This project is being used as a test-bed in which integrated simulation can be carried out. For example, the remote maneuvering support system, deep-sea navigation system and engine monitoring and control

system, an image of which is shown in Figure 12, have been tested in this project (Fraunhofer CML, 2015). The absolute goal of this project, unmanned vessel, is attractive. However, at the same time, it is quite challenging. If no support had been provided by EU government, the group of eight challengers might not have seen any opportunity to examine their innovative ideas because of the risk of failure. Conversely, providing a test-bed or other corresponding budgetary support gives an incentive for those who have a concrete and interesting idea of innovation to materialize such idea. In this context, so as to support the effort of the private sector to develop technologies associated with maritime big data, it is effective to provide aids for an experimental project which tries to achieve innovation in technology, as represented by FP7. If a country already has established a system for research and development of science and technology, supports for projects related to maritime big data can be provided under such a framework. If there is no such existing system in a country, the government needs to newly establish a specific aid scheme dedicated to maritime big data technologies.

## MUNIN Test-bed

Integrated simulation-based environment

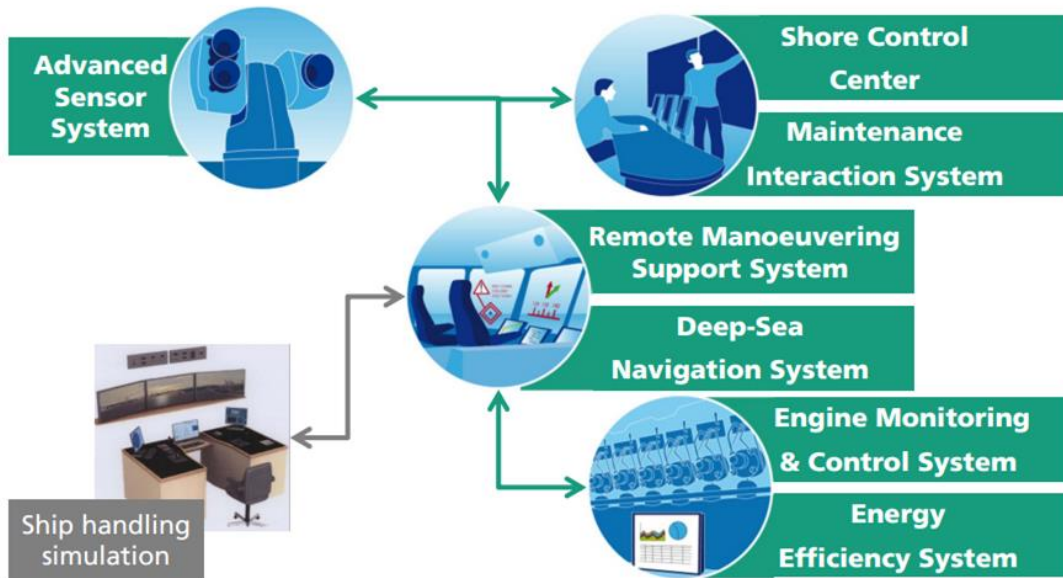


Figure 12: Integrated simulation system adopted in MUNIN. Reprinted from *Autonomous Technologies for Unmanned Cargo Ships*, by Burmeister, H.

### Data integration

Data integration can be interpreted as “standardizing data”. As already discussed in chapter 3, incompatible data standards and reference data may make it difficult to assemble different sorts of data (LRF, 2014). This means that if various kinds of data are compatible, processing, combining and analyzing data will be possible or easier.

There are two possible scenarios in which compatibility of data is achieved. First, it may be achieved among stakeholders in the private sector through voluntary and collaborative work. Different companies may share the same format or structure of

data which is applied to each company's electronic products. However, as mentioned in the part of “flexibility”, the behavior of private entities cannot be presumed or predicted. Nobody can control them in terms of such collaborative action.

Second, it may be possible that a unified data format or structure is created or provided by a reliable third-party entity. This approach is justified by e-navigation, which was reviewed in the previous chapter. As a finding of study on e-navigation, four tasks which possibly have concern with maritime big data were found, as shown in Figure 7. Out of those four tasks, “Common Maritime Data Structure” is the same kind of concept as the compatibility of maritime big data. The concept of CMDS is basically agreed in the MSC. Member States of IMO are trying to share a unified format of data which is going to be used in e-navigation. According to SIP (IMO, 2014), it is expected that data providers adopt “IMO recognized data standards such as IHO's S-100”. Regarding S-100 standard, IHO (2015) explains that “S-100 provides a contemporary hydrographic geospatial data standard that can support a wide variety of hydrographic-related digital data sources, and is fully aligned with mainstream international geospatial standards, in particular the ISO 19100 series of geographic standards, thereby enabling the easier integration of hydrographic data and applications into geospatial solutions”. It also says “these standards specify, for hydrographic and related information, methods and tools for data management, processing, analyzing, accessing, presenting, and transferring such data in digital/electronic form between different users, systems and locations”. In short, S-100 is a standard of hydrographic data which complies with the standards set by the International Organization for Standardization (ISO) and the IMO adopted it as a



unified format for e-navigation. This means that the convenience of a unifying format or structure of certain kinds of data is recognized and agreed in the international arena. Therefore, it is rational to consider that the unified format or structure of maritime big data may be welcomed by maritime industry.

Unification of style of maritime big data can be done through discussion in the IMO, as with CMDS. First of all, one or more voluntary Member States make a proposal on this issue to a MSC meeting. The necessity or convenience of unifying the style of data, as with the example of e-navigation, should be presented in the proposal paper. If the proposal is basically agreed, the scope of discussion should be determined. As mentioned in sub-chapter 2-2, there is a vast potential which can be regarded as maritime big data and it is impossible to cover all. Member States have to clearly choose their objectives of unification. After determining the objectives, the relationship of this issue and CMDS framework has to be subsequently defined. The Member States need to decide if this issue should be a part of the CMDS framework and the scope of CMDS should be expanded. Otherwise, this issue is to be discussed as a new independent topic. Then, under the agreed scope and orientation of discussion, the substance of this topic is discussed in designated committees and sub-committees.

In terms of unifying data style, there is one thing to be carefully considered. This solution should be subject to agreement among the relevant industries and, even if it is agreed, scalability needs to be ensured for further development of format or structure of data in the future. Once the “recommended data format and structure”

are internationally fixed, there is a risk that momentum for further evolving them might be shrunk or lost. As already introduced in this sub-chapter, Interschalt has invented “Maritime Data Engine” which can convert or normalize the data with different structures (Interschalt, 2015). This kind of innovative progress should not be discouraged by inflexible unifications. Therefore, stakeholders have to be careful about the scalability when they standardize the style of maritime big data.

As a consequence of unification of data style, a complementary topic should be considered. This is the certification of data. LRF (2014) states that “certification is a measure of how well a standard has been implemented”. That is to say, if there is a certain format or structure of data to be followed, compliance with such a standard should be checked and guaranteed. Otherwise, no one can believe that a certain data set complies with the designated format and the standard may not effectively work. Actually, LRF predicts that a need in this respect emerges in the near future and regards it as a business opportunity, as already reviewed in chapter 3. It is recalled that LRF stated that it can play “a pivotal role in the adoption of data certification and the certification process itself” (LRF, 2014). As illustrated by this example, certification of data is basically expected to be done by a third-party with corresponding expertise, such as classification society.

#### Summary: Technology

Regarding technologies, the challenge of maritime big data was simple. It is required to invent new electronic equipment with higher performance and upgrade information-processing technology. Although these challenges should be basically

tackled by the private sector, the public sector can support such an effort. Based on the discussion in chapter 2, it turns out that the government and other third party entities may help the private sector in terms of “flexibility” and “data integration”. In terms of “flexibility”, providing a test-bed or corresponding budgetary support through a national subsidy framework for research and development of science technology seems effective. The example of MUNIN shows that innovative ideas tend to need a test-bed in order to be achieved. MUNIN’s test-bed is so large that it is quite risky or impossible for private entities to do it without help given by the public sector. With respect to “data integration”, unification of format or structure of maritime big data can be carried out in IMO, if it is internationally agreed. During such consideration, the existing framework of CMDS in e-navigation should be consulted. If the unification of data style is achieved successfully, classification societies, among others, are expected to play a role in the certification of the data.

#### **4-4 Security**

Security, as an aspect of challenges in terms of maritime big data, was derived through reviewing LRF’s recognition and strategy. LRF pointed out that the unlawful control of device/machine and abusive insertion, update and deletion of data are the actual risks of big data security. These seem to represent the typical elements of risk to be considered because they assume unlawful access to both the system which treats big data and data itself.

The issue of security is getting more and more crucial, as the maritime industry is becoming more and more computerized. ESC Global Security's head of cyber security, Joseph Carson says that "advancement in broadband technologies and the move towards 'Big Data' will leave the maritime industry vulnerable to cyber-crime unless it develops a better awareness of ICT security and adopts security best practices" (Digital Ship, 2015). As more extreme expression, Espen Barth Eide, who is World Economic Forum managing director, said that "every conflict we see in the future will be a cyber-conflict" at Nor-Shipping 2015 in Oslo (Digital Ship 2015). These examples show that many maritime stakeholders are considering that security of big data matters.

#### What is the core of the problem?

Data security can be called in a different way, such as "cyber security". According to the University of Maryland University College, "cyber security, also referred to as information technology security, focuses on protecting computers, networks, programs and data from unintended or unauthorized access, change or destruction". As illustrated by this definition, the key factor of data security is the existence of the assailant.

Taking into account this feature, the problem of data security can be structured. There are three dimensions to be considered. First, as a matter of course, what is legal/illegal has to be distinguished. The second dimension is whether a person intends to perpetrate a cyber-crime or not. The third dimension is, in case a person

commits a crime, if such an attack will be successful or not. The approach to enhance big data security should be considered in each of these dimensions.

With regard to the first and second dimension, legislation seems to be the proper approach. The definition of illegal action has to be given by the national law of each country and a person's determination on whether to perpetrate a crime should be affected by penalties stipulated in the corresponding law. The necessity of the legislative approach is advocated by McKinsey (2011), who states that "businesses and governments will almost certainly want strong laws prohibiting hacking and other security invasions to protect their own operations as much as possible, and they will want effective enforcement of those laws". On the other hand, as regards the third dimension, technology is the key approach. This is an issue that systems which treat maritime big data, as well as data itself, have to be secure enough to make a criminal access unsuccessful. In this respect, how sophisticated and advanced a crime-prevention system is installed, is important. The structure of the problem discussed here can be summarized in Figure 13.

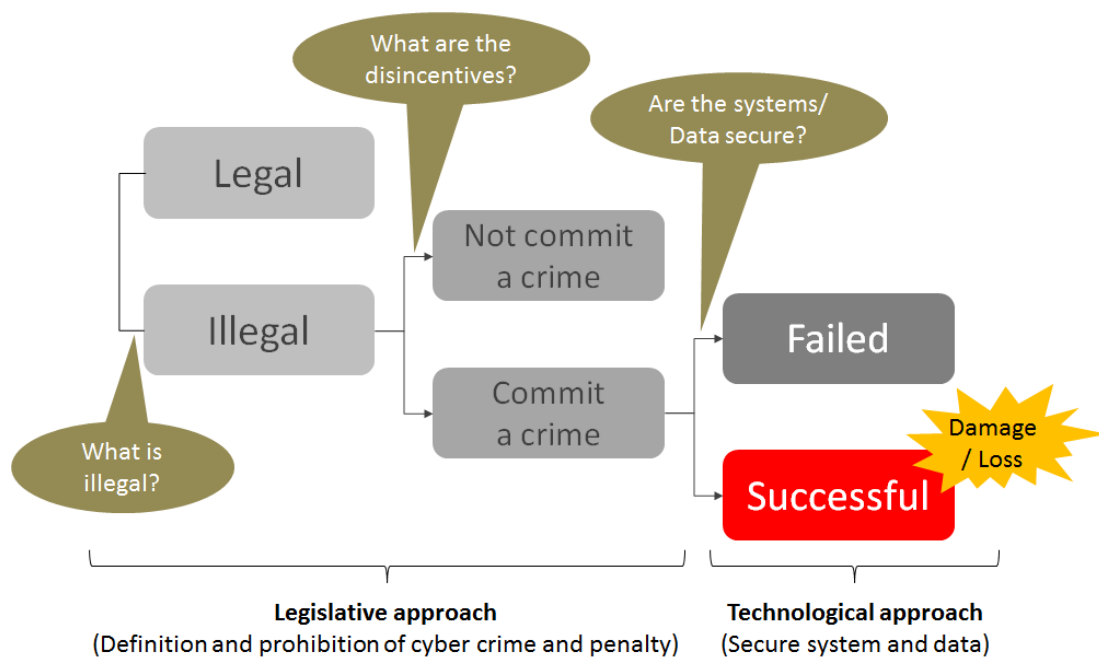


Figure 13: Structure of data security problem.

### Legislative aspect

In this aspect, it is essential to define data-related crimes, prohibit them and set up corresponding penalties in case of violation. The main purpose of this approach is to prevent crimes from happening. Useful examples of such legislation can be found in Japan. There are at least two major national laws associated with cyber-crime in Japan. First is the law of Unfair Competition Prevention Act (1993), which sets out regulations in terms of corporate practice. In this law, the “trade secret information” of companies is defined and abusive access to such information is regarded as a crime and an objective of sanction. The second example is an Act concerning the Prohibition of Unauthorized Computer Access (2000), which directly regulates unauthorized access to computers. The scope of this law is not limited to corporate

activities; attacking an individuals' computer is also prohibited. These laws have been frequently amended. Most recently, the former was amended in July 2015 and the latter in May 2013. This is because the whole society is becoming information-intensive. The more developed information-driven society has a bigger risk of cyber-crimes. In terms of range of coverage, those laws introduced above take care not only of the maritime industry but all relevant industries. As the maritime industry is just one of many industries which are under the risk of such crimes, the corresponding legislation naturally should be developed for all of them.

#### Technological aspect

Even if the legislative condition is perfectly developed, potential assailants are not supposed to completely disappear. Thus, a secure and resilient system and data design is necessary in order to make such illegal attacks unsuccessful. As already mentioned, such a system and data totally depends on the associated technologies. Here, the discussion in sub-chapter 4-3 should be recalled. How to promote development of relevant technologies was discussed in that sub-chapter and the same approach or method can be applied to the security-related technology issue.

As such effort, an EU-backed research and development project can be raised. The European Cyber Security Protection Alliance (CYSPA) is a project which has run from 2012 to 2015. The objective of this project is “to address trustworthy ICT through a European strategy to protect cyberspace, with target audiences ranging from research communities and industry to public authorities and infrastructure operators” (European Commission, 2014). The project creates “an Alliance for EU

stakeholders working together to articulate, embody and deliver the concrete actions needed to reduce cyber disruption” (CYSPA, n.d.). In short, interested stakeholders from various industries work together to enhance their cyber security and the foundation for such cooperative activity is provided by the European Commission.

### Summary: Security

By structuring the problems regarding the security of big data, it turns out that there are two necessary approaches. One is the legislative approach and another is the technological approach. In terms of the legislative aspect, as represented by the example of Japanese laws, it is necessary to define illegal activities, prohibit them and set penalties for violations through national laws. This will be a disincentive for potential assailants to commit crimes. With respect to the technological aspects, it is important to develop or invent a better system and data set which are secure enough to make cyber-attacks unsuccessful, in case such crime is attempted. As a method to tackle this challenge of technology, research and development aid scheme, as discussed in the previous sub-chapter, can be applied.



## **5 Conclusion**

### The significance of big data and the main purpose of this dissertation

Recently, the world has become an information-intensive society. In every aspect, our lives are surrounded by more and more electronic devices and information. Under such circumstances, the term “big data” is popular and attracts many industries. Big data is a significant business opportunity for them. In fact, data-oriented customer services are provided by the companies such as Amazon and TOYOTA, which are two examples. This trend is also applicable to the maritime industries. It is expected that effective use of big data can enhance the safety of ships and environmental protection and, actually, some institutions have already been trying to make innovations of big data in the maritime field. For example, Class NK is seeking to utilize big data for machinery condition monitoring, while DNV-GL is aspiring to develop unmanned, zero emission short sea ships. However, these are just individual actions. As of now, it seems there is no common understanding of what are crucial challenges for the further utilization of big data in the maritime field. This dissertation tries to make a contribution in this respect, identifying the common and important challenges for the use of big data in the maritime industries and presenting possible solutions for such challenges.

### Approaches

For these purposes, three main steps were followed. First, the definition and main features of big data were reviewed in a general sense and the outline of “maritime big data” was considered. Second, some cutting edge marine-related institutions’ stances and strategies of big data were studied in order to identify common and important challenges. Third, identified problems were structured and possible solutions were considered and presented.

### Definition and main features of big data

As regards definitions and the main features of big data, three major conditions to define big data were found, through reviewing definitions including those of NASA, OED, IBM and McKinsey. First, big data takes the form of electronic data. Second, it is derived through various sensors. Third, there are difficulties in capturing, storing, managing and analyzing it. Furthermore, in terms of the third point, it turned out that the difficulty is categorized into four aspects, as a kind of common understanding of industries. The four aspects are called “4Vs”, which are volume, velocity, variety and veracity. Volume represents scale of data. Velocity is speed of streaming data. Variety means different form of data and Veracity expresses uncertainty of data. The main features of big data were summarized in Figures 1 and 2.

### Maritime big data

Taking into account these general features, the scope of the dissertation was determined as a foundation of discussing challenges and solutions of maritime big data. There are so many kinds of data in the maritime field that it is quite difficult to

describe an absolute definition of “maritime big data”. Thus, in order to make the discussion compact and useful, the scope was limited to a common interest of relevant industries, namely “data derived on board vessels”. It was discovered that the general features of big data are applicable to such data. Voyage-related data on board ships is not the only component of maritime big data. However, it can be regarded as a “significant piece of maritime big data”.

#### Cutting edge institutions of maritime big data

In order to identify common and important challenges regarding maritime big data, two famous cutting edge institutions’ stance and strategy of big data were reviewed. They are DNV-GL and LRF, both of which adopted the “4Vs” as the definition of big data. In addition, e-navigation was also studied as it is a recent framework of IMO which is closely related to electronic data on board ships.

DNV-GL observes that developing technical operation and maintenance and improving energy efficiency are the major potential areas of maritime big data. At the same time, DNV-GL considers that there are six main challenges for the development of the use of maritime big data. It was considered that these six elements of problems can be sorted into three main categories, which are “sound competitive conditions”, “human resources” and “technology”. These findings were summarized in Figure 4. On the other hand, LRF raises such items as enhancing machinery maintenance and the management of workload and workers’ stress as the typical innovations to be achieved by big data. There were five problems introduced by LRF. Three of these were included in the category of “technology”, while the remaining two should be

categorized as a fourth group, the “security” issue. The most significant output from this consideration was finding four important challenges with some concrete elements for each. These findings were used as ingredients of the later discussion.

As regards e-navigation, there is both a similarity and a difference when compared with the notion of big data. The similarity is that both e-navigation and maritime big data totally depend on electronic devices and the data derived and processed by them. The difference is that e-navigation concentrates on safety and efficiency of navigation during voyages and is interested in real-time, streaming data, while maritime big data covers not only real-time data, but also the accumulation of data for use after voyages. Taking into account this similarity, the direction of development and tasks of e-navigation were reviewed and it was concluded that there are four tasks which may have a linkage with maritime big data. The four tasks are automated data collection, software quality assurance, reliability and resilience of data and Common Maritime Data Structure.

#### Analysis of challenges and solutions

Four main aspects of challenges identified through review of cutting edge institutions were structured and possible solutions for them were discussed and presented, referring to existing model cases in other industries.

With respect to sound competitive conditions, it was recognized that the core of the problem is an absence of obvious rules in terms of the rights and responsibilities on the treatment of big data. In order to tackle this problem, discussion was divided into two patterns. One is a case in which a limited number of stakeholders share and

exchange data for certain collaborative activities. Another case is that data is open and shared in the public domain. It was recommended that such guidelines and MOUs be developed in order to provide industries with clear instructions on their rights and responsibilities. This solution is expected to promote data exchange and collaborations among companies, removing unnecessary concerns or precautions.

As regards the aspect of human resources, the maritime industry is going to face serious structural problems; this concerns the shortage of data scientists with expertise on maritime big data. In order to cope with this difficulty, two possible solutions are presented. First, it is recommended that governments develop their country's strategy on human resource development in terms of maritime big data. Such "general direction" shown by governments can be a help for investment decision of entities in the private sector. Second, the cooperation framework among industry, government and academia should be enhanced. As an example, establishing joint committees dedicated to human resource development is proposed. This framework is established by a government along with representatives from academia and industry who collaboratively design a new course for nurturing experts on maritime electronics. The latter solution can be one of the direct answers which solve the structural problem of human resource shortages.

The issue of technology is a crucial foundation of development of maritime big data. However, this issue should be basically tackled by the private sector depending on the principle of market competitiveness. Therefore, the key is how it should be supported by governments or other third-parties. As such support, two main

approaches were considered; they are “flexibility” and “data integration”. Regarding “flexibility”, the government is supposed to provide test-bed or corresponding budgetary aid for innovative projects through research and development projects which are represented by FP7 in the EU. This kind of support will give incentives for private companies and institutions to venture to materialize their innovative ideas. In terms of “data integration”, unification of format or structure of data may be helpful for the maritime industry to exchange their data smoothly. This can be achieved in IMO by reference to CMDS framework in the discussion of e-navigation. However, the Member States have to be careful to ensure the scalability of format or structure of data. Too rigid standards sometimes prevent further innovations. When unification of the data style is carried out, the classification societies can play a supplementary role providing data certification services.

With regard to Security, the problem was refined into legislative and technological aspects. In the aspect of legislation, associated national laws have to be developed. The law defines illegal actions in terms of security of data and network, prohibits such evil actions and stipulates penalties in case of violations. Such laws may be established to cover not only the maritime industry but also the whole of related industries. A set of examples was found in the Japanese legislative system. On the other hand, the technological aspect becomes critical when an assailant attempts to commit a cyber-crime. The relevant system and data set should be secure enough to make such crime unsuccessful. In order to invent more secure tools, frameworks such as governmental research and development aid can be a significant help.

### Sum-up

To conclude, this dissertation studies the basic features of big data, clarifies that such features can be applied to voyage-related data derived on board ships, reviews cutting edge institutions' stances and strategies on maritime big data, identifies common and important challenges for development, and introduces some possible solutions to solve such problems. There are two particularly significant achievements to be highlighted. First, four fundamental challenges for further developing maritime big data have been identified. Second, possible solutions for each of four problems have been proposed. In each aspect of challenges, it seems impossible for the private sector to conquer all of the challenges by itself; the private sector also has to take certain roles. By fully applying these solutions proposed in this dissertation, the utilization of maritime big data is supposed to be accelerated in the whole of the maritime society. It is expected, one day, all of these challenges will be cleared and the society will see an enormous contribution of maritime big data for the safety of ships and human lives as well as environmental protection.

## References

- ABS. (2004). *Maritime Accidents and Human Performance: The Statistical Trail*. Retrieved August 19, 2015, from <http://www.eagle.org/eagleExternalPortalWEB/ShowProperty/BEA%20Repository/References/Technical%20Papers/2004/MaritimeAccidentsHumanPerformance>
- Act concerning the Prohibition of Unauthorized Computer Access, Japan. (2000). Retrieved September 5, 2015, from <http%3A%2F%2Fwww.e-gov.go.jp%2Fhtmldata%2FH11%2FH11HO128.html>
- Burmeister, H. (2015). *Autonomous Technologies for Unmanned Cargo Ships*. Retrieved September 3, 2015, from <http://www.thedigitalship.com/conferences/presentations/2015hamburg/15-MUNIN-150318-Digital-Ship-to-publish.pdf>
- Catlett, C., Gentzsch, W., Grandinetti, L., Joubert, G. R., & Vasquez-Poletti, J. L. (2013). *Cloud computing and big data*.
- Christensen, T. (2014). *The Maritime Cloud and its applications*. Danish Maritime Authority.
- Class NK. (2015). *Application of Maritime Big Data to Machinery Condition Monitoring*.
- Committee on STEM Education of National Science and Technology Council. (2013). Federal Science Technology, Engineering, and Mathematics (STEM) Education 5-year Strategic Plan. Retrieved August 27, 2015, from [https://www.whitehouse.gov/sites/default/files/microsites/ostp/stem\\_stratplan\\_2013.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/stem_stratplan_2013.pdf)
- Creative Commons. (n.d.). Keep the internet creative, free and open. Retrieved August 31, 2015, from <https://creativecommons.org/publicdomain/>
- CYSPA. (n.d.). CYSPA stand at the CSP Innovation Forum in Brussels. Retrieved September 7, 2015, from <http://www.cyspa.eu/default.aspx?page=home>
- Digital Ship. (2015, July 11). Plenty of 'phish' in the sea, warns security company. Retrieved September 4, 2015, from <http%3A%2F%2Fthedigitalship.com%2Fnews%2Fmaritime-software%2Fitem%2F3889-plenty-of-phish-in-the-sea>
- DNV-GL. (2014). *BIG DATA - the new data reality and industry impact* (Vol. 4, DNV GL Strategic Research & Innovation Position Paper). Retrieved August 17, 2015.
- DNV-GL. (2015). *Annual Report 2014*. Retrieved August 17, 2015, from <https://www.dnvgl.com/about/publications/annual-reports.html>



- DNV-GL. (2015). *ReVolt: The unmanned, zero emission short sea ship of the future*.
- DNV-GL. (n.d.). About DNV-GL. Retrieved August 17, 2015, from <https://www.dnvgl.com/about/in-brief/our-history.html>
- DNV-GL. (n.d.). Key figures for DNV GL. Retrieved August 17, 2015, from <https://www.dnvgl.com/about/in-brief/key-figures.html>
- DNV-GL. (n.d.). Maritime. Retrieved August 18, 2015, from <https://www.dnvgl.com/about/in-brief/maritime.html>
- Economove Japan Co., Ltd. (n.d.). Sailing date to major destinations from a Japanese port. Retrieved August 7, 2015, from [http://www.economovejapan.com/html/date\\_salling.html](http://www.economovejapan.com/html/date_salling.html)
- European Commission. (2014). European CYber Security Protection Alliance. Retrieved September 7, 2015, from [http://cordis.europa.eu/project/rcn/106313\\_en.html](http://cordis.europa.eu/project/rcn/106313_en.html)
- European Commission. (2015). MUNIN. Retrieved September 3, 2015, from [http://cordis.europa.eu/project/rcn/104631\\_en.html](http://cordis.europa.eu/project/rcn/104631_en.html)
- European Commission. (n.d.). What is FP? The basics. Retrieved September 3, 2015, from [http://ec.europa.eu/research/fp7/understanding/fp7inbrief/what-is\\_en.html](http://ec.europa.eu/research/fp7/understanding/fp7inbrief/what-is_en.html)
- European Union. (2011). *New Practical Guide to EU Funding Opportunities for Research and Innovation*. Retrieved September 3, 2015, from [http://ec.europa.eu/research/participants/data/ref/fp7/204008/practical-guide-rev3\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/fp7/204008/practical-guide-rev3_en.pdf)
- Friedman, U. (2012, October 8). Big Data: A Short History. *Foreign Policy*. Retrieved July 8, 2015, from <http://foreignpolicy.com/2012/10/08/big-data-a-short-history/>
- IBM. (n.d.). The Four V's of Big Data. Retrieved June 29, 2015, from <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>
- IBM. (n.d.). *What is Big Data* (Unpublished master's thesis). Retrieved June 28, 2015, from <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- Intel Corporation. (2013). *A vision for big data*.
- International Chamber of Shipping. (n.d.). Shipping and World Trade. Retrieved August 11, 2015, from <http://www.ics-shipping.org/shipping-facts/shipping-and-world-trade>
- International Convention for the Safety of Life at Sea, 1974. IMO. (1974).
- International Maritime Organization (IMO). (2006). *Report of the Maritime Safety Committee on its eighty-first session* (MSC 81/25).
- International Maritime Organization (IMO). (2008). *Report of the Maritime Safety Committee on its eighty-fifth session* (MSC 85/26).

- International Maritime Organization (IMO). (2012). *Adoption of revised performance standards for shipborne Voyage Data Recorders (VDRs)* (Resolution MSC.333(90)).
- International Maritime Organization (IMO). (2013). *Development of draft Software Quality Assurance (SQA) Guidelines for e-navigation* (NAV 59/6/2).
- International Maritime Organization (IMO). (2013). *Report to the Maritime Safety Committee* (NAV 59/20).
- International Maritime Organization (IMO). (2014). *Overview of the Maritime Cloud concept* (NCSR 1/INF.21).
- International Maritime Organization (IMO). (2014). *Report of Maritime Safety Committee on its ninety-fourth session* (MSC 94/21).
- International Maritime Organization (IMO). (2014). *Report to the Maritime Safety Committee* (NCSR 1/28).
- International Maritime Organization (IMO). (2015). *Report of the Maritime Safety Committee on its ninety-fifth session* (MSC 95/22).
- International Maritime Organization (IMO). (n.d.). E-navigation. Retrieved August 24, 2015, from <http://www.imo.org/en/OurWork/Safety/Navigation/Pages/eNavigation.aspx>
- International Maritime Organization (IMO). (n.d.). Voyage Data Recorder (VDR). Retrieved August 21, 2015, from <http://www.imo.org/en/OurWork/Safety/Navigation/Pages/VDR.aspx>
- Interschalt maritime systems AG. (2015). Innovation for shipping.
- Kassner, M. (2014, December). Who owns the data once it becomes part of big data? Retrieved August 29, 2015, from <http://www.techrepublic.com/article/who-owns-the-data-once-it-becomes-part-of-big-data/>
- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. *Application Delivery Strategies, META Group, 949*. Retrieved June 10, 2015, from <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Lipman, B. (2014, July 3). Today's Cyber Security is a Game of Cat-and-Mouse - Wall Street & Technology. Retrieved September 4, 2015, from <http://www.wallstreetandtech.com/security/todays-cyber-security-is-a-game-of-cat-and-mouse/d/d-id/1268787>

- Lloyd's Register Foundation. (2014). *Foresight review of big data* (Vol. 2, Report series 2014).
- Lloyd's Register Foundation. (2014). *Foundation review 2014*. Retrieved August 18, 2015, from <http://www.lrfoundation.org.uk/Images/49109-lloyds-register-foundation.pdf>
- Lloyd's Register Foundation. (n.d.). Funding for the future. Retrieved August 18, 2015, from <http://www.lrfoundation.org.uk/funding/index.aspx>
- Lloyd's Register Foundation. (n.d.). Mission. Retrieved August 18, 2015, from <http://www.lrfoundation.org.uk/about-us/mission/>
- Lloyd's Register Foundation. (n.d.). A proud tradition. Retrieved August 18, 2015, from <http://www.lrfoundation.org.uk/about-us/history/>
- Lohr, S. (2009, August 05). For Today's Graduate, Just One Word: Statistics. Retrieved August 27, 2015, from [http://www.nytimes.com/2009/08/06/technology/06stats.html?\\_r=0](http://www.nytimes.com/2009/08/06/technology/06stats.html?_r=0)
- Mangalindan, J. (2012, July 30). Amazon's recommendation secret. Retrieved August 20, 2015, from <http://fortune.com/2012/07/30/amazons-recommendation-secret/>
- Ongsberg Maritime AS. (n.d.). Vessel Performance System. Retrieved August 7, 2015, from <http://www.km.kongsberg.com/ks/web/nokbg0240.nsf/AllWeb/AC15755EC2442854C125788F0031E4F6?OpenDocument>
- Open Data Commons. (n.d.). Licenses. Retrieved August 30, 2015, from <http://opendatacommons.org/licenses/>
- Oxford University Press. (2008, September). OED (Oxford English Dictionary). Retrieved June 28, 2015, from <http://www.oed.com/view/Entry/18833?redirectedFrom=big%2Bdata#eid301162177>
- Press, G. (2014). 12 Big Data Definitions: What's Yours? *Forbes*. Retrieved June 27, 2015, from <http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/>
- Privacy and security principles for farm data. (2015). Retrieved August 30, 2015, from <http://www.fb.org/tmp/uploads/PrivacyAndSecurityPrinciplesForFarmData.pdf>
- Research data management service group, Cornell university. (n.d.). Introduction to intellectual property rights in data management. Retrieved August 30, 2015, from <http://data.research.cornell.edu/content/intellectual-property>
- SAS Institute Inc. (n.d.). What Is Big Data? Retrieved July 8, 2015, from [http://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](http://www.sas.com/en_us/insights/big-data/what-is-big-data.html)

Tokyo Keiki Inc. (n.d.). ジャイロコンパス(Gyrocompass). Retrieved August 7, 2015, from <http://www.tokyo-keiki.co.jp/marine/products/gyro.html>

TOYOTA. (2013, May). TMC Announces New Big Data Traffic Information Service. Retrieved June 28, 2015, from <http://www2.toyota.co.jp/en/news/13/05/0529.pdf>

Unfair Competition Prevention Act, Japan. (1993). Retrieved September 5, 2015, from <http://law.e-gov.go.jp/htmlldata/H05/H05HO047.html>

United States Coast Guard (USCG). (n.d.). How AIS Works. Retrieved August 9, 2015, from <http://www.navcen.uscg.gov/?pageName=AISworks>

U.S. Department of Education. (n.d.). Science, Technology, Engineering and Math: Education for Global Leadership. Retrieved August 27, 2015, from <http://www.ed.gov/stem>

The White House. (2013). Educate to Innovate. Retrieved August 27, 2015, from <https://www.whitehouse.gov/issues/education/k-12/educate-innovate>