Master's Degree in
Data Science and Management

Data Privacy and Security

Towards a Secure Data Space for Maritime Data
The Case of the Italian Coast Guard

SUPERVISOR
Paolo Spagnoletti

CO-SUPERVISOR
Blerina Sinaimeri

CANDIDATE
Martina Manno

Academic Year 2022/2023

Abstract

Shipping is the world's oldest sharing economy, and it is still the most efficient way of transporting goods around the globe, handling about 90% of the world's trade. As technology continues to advance, this sector faces new challenges and opportunities. In this context, the maritime informatics discipline meets the need of increasing the efficiency, safety, resilience, and ecological sustainability of the global maritime industry by promoting standardized digital data sharing.

Digitalization has transformed various industries, and the maritime sector is no exception. It involves the integration of advanced technologies and the utilization of modern communications and geolocation systemsand systems to streamline processes, improve decision-making, and optimize resource utilization. In the maritime domain, digitalization encompasses t to enable efficient and real-time data exchange between maritime entities.

On the other hand, digitization has led to such an increase in the amount of data available in the maritime sector that it has also highlighted the need for better data sharing and collaboration. Currently, maritime data are distributed among a wide range of sources, including ships, surveillance systems, port authorities and other organizations. This fragmentation of data makes it difficult to obtain a complete view of the maritime domain and make informed decisions. Under these circumstances, it can be difficult to verify the availability, integrity, and reliability of data when they are dispersed among multiple sources. There are several ways to address the problem of data unreliability in the maritime domain. One important step is to improve data sharing and collaboration among maritime organizations by creating a data ecosystem whose purpose is to support the collection, storage, processing, and sharing of large amounts of maritime data. By improving data sharing and collaboration, and by creating a high-reliability data ecosystem, we can improve the reliability of maritime data. This will create new applications and related services, ensure efficient and real-time data exchange between maritime entities, streamline processes, improve decision making, and optimize resource utilization.

The research goal is to outline a high-reliability data ecosystem in which the rules, standards, and technologies used for data storage, security, and governance are defined to improve data quality, accessibility, and usability. The theoretical contribution is approached with a case study from a user-centered perspective, for which a framework is designed to provide guidelines for achieving a security-by-design approach in the maritime domain, ensuring the safety and efficiency of global maritime operations and ensuring that the data infrastructure is consistent, accurate, and reliable across different systems and applications.

Keywords: maritime informatics, digital transformation, public data spaces, intoroperability, security-by-design

Contents

Introduction

As technology continues to advance, the maritime domain faces new challenges and opportunities in ensuring efficiency, safety, and security. Whithin this context, the emergence of maritime informatics has become a crucial field of study, focusing on leveraging digitalization to enhance operations, communication, and data management within the maritime domain. The research area is data privacy and security in relation to digital transformation processes, big data management and the foundation of the data ecosystem. The research aims to explore the maritime domain as one of the most challenging and diverse. In particular, its complexity can be attributed to its territorial vastness and the number of entities it involves, from objects such as ships to people.

In the early 2000s, the first application of DDS (Digital Data Streams) in the maritime domain was the Automatic Identification System (AIS), introduced by the International Maritime Organisation (IMO) in response to maritime security concerns (1). The system was implemented to improve the safety of vessel traffic through the automatic exchange of real-time information and vessel tracking. As the accumulation speed and scale of AIS data have increased, machine learning algorithms have progressed for anomalies detection practices (2). The mandatory implementation of AIS has led to the development of new open-source data visualisation applications using the data derived from AIS (3). Nowadays, AIS has reached a level of maturity that enables it to serve as a platform for further innovation, such as the integration of new technologies and data sources as satellite imagery (SAT-AIS) (4). Since the introduction of the AIS cyber-physical system, the information value of AIS has increased and with it the demand for data integrity to ensure reliable application design and analysis results. It is evident that the value of the AIS and complementary systems is the data transmitted, and it will be fundamental to build a strategical alignment around them as "a community of hierarchically independent, yet interdependent heterogeneous participants who collectively generate an ecosystem output", described in the management conceptual framework by Llewellyn Thomas (5). The existing literature offers a wide range of perspectives and approaches to address the concept of a data ecosystem. Whithin this context, noteworthy is Daniel Beverungen's (6) work on the emergence of public data spaces. Previous research has mainly focused on digital transformation within organisations, while nowadays it is increasingly important to understand the new trend of interconnection between different stakeholders to co-create services based on the provision and use of data. In particular, the conceptualization and implications of public data spaces and related ecosystems offer promising research opportunities. In this context my research question arises: how can digital transformation enhance collaboration between public and private actors in the maritime domain to align their interests while ensuring data privacy and security?

In the empirical context, there are many reasons that clearly indicate the relevance of the topic, but the research will focus on two aspects in particular: the first one looks at the maritime sector as a the trade sector, the second one looks at the maritime sector as a service provider for citizens. Regarding the first perspective, the maritime transport industry faces strong competitive pressure from other modes of transport, many of which benefit from entire new infrastructures (e.g. China's Belt and Road, autonomous transport systems and new airports such as those in Berlin, Beijing, Mexico City and Sydney). To remain an attractive and contemporary transport option in an increasingly digital economy, the maritime ecosystem needs to create a new digital infrastructure that facilitates the provision of reliable and predictable data on the real-time transit (7). The second position emphasises the role of public administrations in the maritime sector, focusing in particular on the case study of the Coast Guard, whose goal is to remain competitive also through the digital services it provides to citizens.

The research goal is to give input to define the requirements of a public data space for the maritime domain focusing on two key concepts that were highlighted in the theory development section: inter-organisational collaboration and data security. The proposal stems from a study of current European projects, such as Gaia-x [1], in which, however, the maritime sector is not included yet. This is because although integrating the maritime sector would bring benefits in economic and social terms, the maritime sector is very challenging due to its scale and specific requirements, as mentioned above. The promotion of public data spaces aims to ensure data sovereignty for organizations. This is achieved by empowering them to exercise control and enhance data access, thereby establishing a secure, dependable, and transparent environment for data interchange among various entities.The development of a public data space for the maritime sector could improve monitoring of maritime traffic and ensure safe navigation, optimization of maritime operations, protection of the marine environment, monitoring of ship conditions, and identification of potential safety problems. As the maritime domain is a diverse sector in terms of activities and therefore data, the theoretical development chapter highlights what technologies are needed to be integrated into a public data space depending on the activities and results to be achieved, in detail I have chosen to focus on the two activities mentioned above, real time monitoring of ships and the provision of administrative services.

The research design follows a deductive approach. Hence, in order to demonstrate the proposed solution's credibility I identified the Coast Guard case study.

[1] https://gaia-x.eu

The case was chosen mainly as a result of a Luiss Business School project that involved interactions with consultants in contact with the Coast Guards that I followed as a thesis student and an interview I was able to obtain during a seminar held by the Italian Coast Guard [2]at the university. In summary, the findings show regarding the real-time navigation data that the Italian Coast Guard currently uses its own platform that permits the data integration with other players such as Italian Space Agency (ASI) [3] and European Space Agency (ESA) [4]. Instead, regarding the provision of administrative services, the digital transformation could have a significant impact in terms of data security and privacy.

Looking back to the literature review context, some advantages of the proposed solution have been shown, rather a future discussion could be focused on the aspect of European borders, as obviously the maritime domain has a more extensive context. Indeed, although Gaia-x is an open and collaborative project towards the world, currently the use cases presented are limited to the European context because they are based on a series of principles, including security, interoperability, data sovereignty and responsibility, which would be guaranteed at European level and on which other countries should align.

[2] https://www.guardiacostiera.gov.it

[3] https://www.asi.it

[4] https://www.esa.int

Chapter 1

Research Background

This chapter defines the research field of the paper, discussing the background of the research topic and the current problems of interest. In detail, the chapter traces the main topics that led me to develop an interest in the maritime domain and increased my desire to explore further its level of digital transformation in the next chapters.

1.1 The Automatic Identification System: vulnerabilities and attack scenarios

Digital data streams (DDS) have been one of the main innovations in the maritime industry, as they involve the continuous digital encoding and transmission of data describing the state of an entity (8). The idea of using digital data to track vessel movements dates back to the 1970s, when air traffic control engineers began using the first radar and automatic identification systems to manage air traffic. In the early 2000s, the first application of DDS for the maritime sector was the Automatic Identification System (AIS), introduced by the International Maritime Organization (IMO) in response to maritime safety concerns (Appendix 1). The system was implemented to improve the safety of vessel traffic through automatic real-time information exchange and vessel tracking.

Since 2002, AIS has been a mandatory installation for international ships with a gross tonnage of at least 300 tons and all passenger ships, regardless of size. It has proven to be useful for the maritime industry, so even pleasure boats and fishing vessels are now often equipped with AIS. According to the International Maritime Organization (IMO), as of December 2021, there are more than 1.2 million AIS installations worldwide. Of these, over 900,000 are shipboard installations and over 300,000 are shore-based installations. Asia is the region with the largest number of AIS installations, followed by Europe, Latin America, and North America. With these estimated numbers, AIS is currently an important technology and solution for:

• maritime security solutions encompass a range of offerings, including aid in security operation support, the monitoring and management of vessel traffic

•law enforcement services cover a broad spectrum, from addressing piracy issues and combating illegal fishing activities to ensuring compliance with international and national regulations, all of which are critical to assisting law enforcement operations

• Search and Rescue (SAR) services

• maritime surveillance services include the monitoring of vessels in sensitive regions, curbing drug smuggling, and bolstering border control

• environmental services encompass the monitoring of hazardous cargo, preventive measures to avert ship-induced pollution, and swift responses to pollution incidents

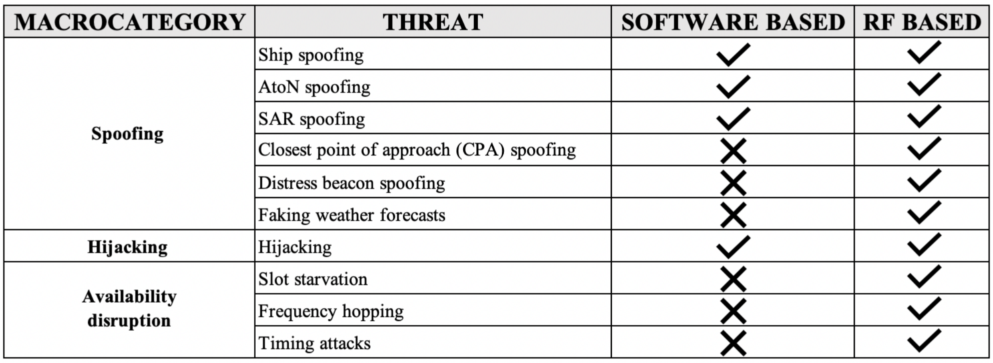
• fleet management services cater to commercial users such as shipping companies and ship owners, offering comprehensive support for their activities

Ship operators and maritime regulatory bodies heavily depend on AIS as a complementary tool to traditional radar systems, aiding in the prevention of collisions and the determination of vessel positions. This is coupled with other technologies such as visual monitoring, audio communication, and long-range identification and detection (LRIT). Due to its ability to furnish precise GPS coordinates, course data, ground speed, and other relevant information, AIS has demonstrated its superior utility in accident inquiries compared to the currently prevalent yet less precise radar technology. Also because ship accidents also occur daily in European waters, as shown by the latest annual overviews (2022) of marine casualties and incidents published by the European Maritime Safety Agency (EMSA)(9), it is important to have the right tool to detect them.

Incidents at sea sometimes have causal and accidental components, but other factors come into play when ships try to conceal their identity, location, or destination. In these situations, vulnerabilities in communication systems can be exploited for economic, geopolitical, and criminal or terrorist purposes. In fact, AIS data has become the industry benchmark, despite its inherent limitations of voluntariness, which is why its oversight is always under the control of European and global authorities, such as the Coast Guards. In this context, maritime safety authorities are facing new challenges with preventive and simulation-based approaches, trying to keep up with digitization to provide efficient services.

Given its primary importance and prevalence in maritime traffic safety, a comprehensive security assessment of AIS is essential. The system can be evaluated from both software and hardware (e.g., radio frequency [RF]) perspectives to identify threats that have affected AIS implementation and protocol specifications. These include disabling AIS communications (i.e., denial of service - DoS); tampering with existing AIS data (i.e., altering the information transmitted by ships); triggering SAR alerts to navigate ships in hostile sea space controlled by attackers; or simulating collisions to take a ship off course. It is interesting to note some real cases where AIS has been contaminated with forged information e.g., the case of Iranian ships falsely identified as belonging to Zanzibar when the U.S. and Europe tightened sanctions related to nuclear programs according to a Bloomberg article (10).

The following paragraphs aim to explain software-based and radio frequency-based threats to AIS in order to understand the consequences of exploiting vulnerabilities in the maritime domain. Three major threat categories can be identified: spoofing, hijacking, and disruption of availability, which can be software-based, radio frequency-based, or a combination of both, as shown in the summary table provided in Figure 1.1:



MACROCATEGORY	THREAT	SOFTWARE BASED	RF BASED
Spoofing	Ship spoofing	✓	✓
	AtoN spoofing	✓	✓
	SAR spoofing	✓	✓
	Closest point of approach (CPA) spoofing	✗	✓
	Distress beacon spoofing	✗	✓
	Faking weather forecasts	✗	✓
Hijacking	Hijacking	✓	✓
Availability disruption	Slot starvation	✗	✓
	Frequency hopping	✗	✓
	Timing attacks	✗	✓

Figure 1.1:
Summary of Identified AIS-Related Threats

In the following, some details about the main AIS threats based on radio frequency are explained:

CPA SPOOFING

The Closest Point of Approach (CPA) works by calculating the minimum distance between two ships, with at least one of them in motion. The CPA algorithm allows ship captains to calculate the time and distance remaining before a potential collision with another ship, assuming the ships are traveling at fixed speed and direction. The CPA alarm is triggered if either parameter (Figure 1.2) falls below the thresholds configured by the transponder. Using this algorithm, AIS allows for the accurate location and identification of vessels within RF coverage, enabling safer navigation when used in conjunction with traditional avoidance mechanisms such as visual observation, voice communications, and radar. CPA spoofing thus consists of simulating a potential collision with a target vessel, triggering a CPA alarm that could cause the target vessel to deviate from its course to hit a reef or run aground during low tide.

$$\begin{cases} T_{CPA} = \dfrac{-w(t_i)(S_r - S_s)}{|S_r - S_s|^2} \\ D_{CPA} = |w(t_i) + T_{CPA}(S_r - S_s)| \end{cases}$$

Figure 1.2:
CPA Algorithm: Tcpa = time remaining before reaching the CPA point. Dcpa = distance between the vessels before reaching the CPA point. Where w(ti) denotes the distance between the vessels at a specific time (ti), and (Sr) and (Ss) are the vessel vectors.

AIS-SART SPOOFING

SARTs represent autonomous devices that play a crucial role in identifying and pinpointing boats and individuals during critical situations. AIS SARTs possess the capability to trigger automatically upon contact with water, transmitting distress signals alongside GPS coordinates. These coordinates are instrumental in aiding rescue teams in locating survivors. The act of manipulating AIS-SARTs, known as spoofing, entails the creation of counterfeit distress signals and the selection of specific coordinates by malicious actors. AIS transponders are mandated to activate alerts upon receipt of distress communications. Perpetrators, including pirates, may exploit SART alerts to lure victims into navigating towards hostile maritime regions under their control. It is essential to bear in mind that, as mandated by legislation, vessels are obligated to engage in Search and Rescue (SAR) operations upon receiving SAR messages.

AVAILABILITY DISRUPTION

• slot starvation entails the fraudulent emulation of maritime regulatory entities to monopolize the complete AIS ”address spectrum,” with the objective of obstructing intercommunication among all the stations within the coverage area. These stations encompass vessels, Aids to Navigation (AtoNs), and AIS access points employed for traffic surveillance. Hence, attackers can disable AIS functionality on a significant scale

• frequency hopping consists in impersonating maritime authorities to issue directives to alter the operational frequencies of one or more AIS transponders. Given that receiving stations are bound to adhere to instructions from maritime authorities, these frequency manipulation tactics endure. Attackers possess the capability to instruct designated vessels to modify their frequencies upon entering specific regions they select, effectively rendering AIS useless

• timing attacks is when malicious users possess the ability to influence the timing of AIS transponder transmissions by altering command inputs, effectively obstructing any subsequent data exchange regarding vessel locations. Hence, vessels may seemingly vanish from the radar screens equipped with AIS technology.

The following are the main software and RF based threats:

SHIP SPOOFING
Ship spoofing refers to the process of creating a valid but nonexistent ship by assigning fictitious information. They can make it appear that ships are within the jurisdiction of an opposing nation or that they are carrying nuclear cargo while sailing in the waters of a denuclearized nation. Ship spoofing could create problems for automated systems that identify data and make inferences based on the collected AIS information. Attackers could falsify information to shift the blame to another ship, for example.

AIS HIJACKING
AIS hijacking involves manipulating any information on existing AIS stations. Attackers can maliciously modify the information provided by AtoNs (Aids to Navigation) installed in ports by authorities to assist and monitor vessels. In the software variant of hijacking, attackers can conduct man-in-the-middle (MitM) attacks to receive ongoing communications and arbitrarily replace AIS information. In the radio frequency version, attackers can replace the original AIS messages with stronger false signals. In both cases, recipients receive modified versions of the victims' original AIS messages from the attacker.

In summary, AIS data has become the industry benchmark, but it has led to increasing manipulation of data at sea. Research by Windward (11) sheds light on the manipulation of AIS data in the real world, the extent of the problem, and its implications, particularly for stakeholders in the financial and intelligence sectors. The introduction of AIS data has tremendous potential for traders, as it enables analysis at the micro level-revealing what a specific ship is carrying and where it is going-and allows them to assess trends at the macro level, such as oil supply projections for a particular country or projected imports and growth in specific regions. Decisions are only as reliable as the data on which they are based, and this axiom is especially true for model-based trading and quantitative analysis, which require a high level of data reliability. The manipulation of AIS has three main implications for the world of finance:

- Distorted view of commodity flows

Understanding cargo flows is directly linked to knowledge of actual ship movements, and incorrect AIS data can lead to inaccurate and misleading analysis of key parameters, such as the amount of a given cargo transported by sea

- Misunderstanding of supply and demand

Freight rates are determined by supply and demand in specific ports and areas. Knowing how many ships are available at a particular port or what cargoes have left the port is extremely valuable information. AIS data that can "hide" ships or cargoes and obscure destinations can have a significant economic impact by influencing the perception of supply and demand

- Impact on trading models

Trading models that rely on the data are designed to account for expected deviations from the data. However, because AIS data are manipulated and there is no validation mechanism to control this phenomenon, there is no way to ascertain the extent of false data and adjust models accordingly

AIS manipulation is a rapidly growing global trend that challenges decision makers who rely, often unknowingly and unwittingly, on inaccurate and increasingly manipulated data. As regulations have increased, the amount of AIS data has also increased, but with diminishing quality as ships realize they are being "watched" through their AIS transmissions. To evade detection, the manipulation practices described above are already being used, and others will surely come to light. Starting with the AIS system, the next section outlines AIS's progress toward the new technologies available.

1.2 Digital transformation in maritime domain

Existing maritime communication systems include land, air, and space networks. Most technologies for ship-to-ship and ship-to-shore communication use medium frequency (MF), high frequency (HF), very high frequency (VHF) and very high frequency (UHF) bands (12). Although these bands can provide relatively long propagation distances, they support only basic applications such as half-duplex voice calls, text messages, and automatic identification system (AIS).

Since the introduction of the AIS cyber-physical system, many steps have been taken not only in terms of regulations. By its very nature, AIS is a completely open and standardized communication protocol, and from the very beginning, this feature has enabled interoperability among the AIS systems of different ships, coastal infrastructures, and relevant authorities for safety and maritime traffic monitoring purposes. With the development of information and sensor technology, the era of AIS-originated big data began, making AIS data available and accessible to the public. The author E. Lee et al. (13) summarized several AIS data applications in the Figure 1.3.

Due to the flexibility of this data, it can be collected, analyzed, distributed, stored, and visualized to improve ship monitoring and port management to ensure the safety of people, the environment, operations, and data. The impact of digital transformation on AIS-generated data will be briefly discussed highlighting visualization tools, machine learning algorithms, advanced geolocation techniques and cybersecurity techniques.

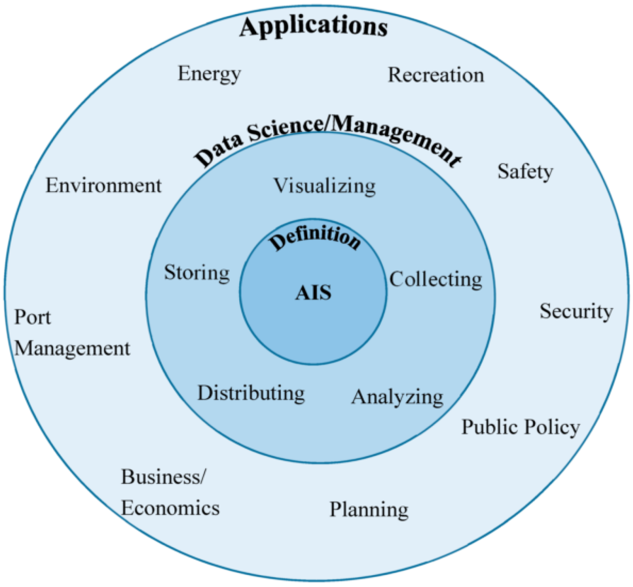


Figure 1.3:
Automatic identification systems (AIS) and its applications by E. Lee et al. (13)

1.2.1 Data visualization techniques and Machine Learning algorithms as a new dimension for value creation

Advanced visualization tools are used to present AIS data more clearly and intuitively, enabling navigation operators to quickly understand important information. Visualization tools enable the creation of maps and graphs of AIS data that can be used to identify problems and improve understanding of the data. For example, they can be used to identify maritime traffic congestion or to monitor the activities of ships in certain areas. Visualization of AIS data is an essential task to facilitate data exploration and decision making. K. Wang et al. (14) provide an overview of maritime traffic data visualization techniques and tools, such as graphs, interactive maps, three-dimensional visualizations, heatmaps, flowcharts, and Geographic Information Systems (GIS) tools, emphasizing that the combined use of these techniques and tools enables effective visualization of maritime traffic data, providing an in-depth understanding of traffic patterns and maritime industry dynamics.

Artificial intelligence algorithms are used to analyze AIS data, making it possible to identify patterns and anomalies in ship movements and predict future behavior. This type of analysis is useful for maritime accident prevention and maritime traffic management. As for artificial intelligence, there are techniques such as cluster classification analysis that can analyze large amounts of AIS data and identify patterns and correlations. For example, these techniques can be used to detect suspicious ship behavior, analyze ship trajectories, analyze weather data, and use heat maps to identify areas at highest risk of accidents or terrorist attacks and alert maritime operators in real time. In addition, artificial intelligence can be used to improve the quality of AIS data, such as by eliminating false positives or correcting positioning errors. The possibilities of analyzing AIS data using big data techniques have already been explored, K. Wolsing et al. (2) proposed a review of recent approaches to anomaly detection. The various techniques evaluated can be identified in 3 main categories: one group is based on machine learning (with neural networks or clustering), another group on stochastic methods such as Gaussian models, and the other group uses geometric properties. The table in Appendix 2 summarizes the results of the 44 anomaly detection approaches for maritime AIS tracks. Although the methodologies differ widely, there are many commonalities in the type of anomaly, such as course deviation being the most important. In addition, most detectors focus on a specific region and thus require new training to be applied to other regions.

1.2.2 From maritime to space: new cybsersecurity challenges

Advanced geolocation techniques, such as satellite triangulation, are being used to improve the accuracy of AIS data and provide greater visibility into the location of ships for maritime traffic management and accident prevention. Maritime authorities are already working with the European Space Agency (ESA) to integrate AIS data with satellite imagery. In particular, Italy has proven to play a leading role, with significant investment and strong involvement in the agency's key programs. According to recent estimates, Italy boasts positions at the top of international rankings, ranking fifth in terms of innovation in the sector and seventh in terms of public investment in space activities as a proportion of national GDP, as well as second at the European level in terms of the number of assets in orbit. (15). Indeed, ESA has developed several projects aimed at integrating AIS data and satellite imagery for maritime surveillance, including the SAT-AIS (Satellite Automatic Identification System) project, the MAR-SAT (Maritime Security Services) project, and the GMES (Global Monitoring for Environment and Security) project. A practical example is the E-SAIL microsatellite (4) which aims to increase the coverage of the Automatic Identification System, a short-range coastal tracking system. Its main limitation is the earth's curvature, which limits its horizontal range to about 74 kilometers (16). With the use of satellites this problem can be solved, making it possible to identify and track maritime vessels around the world. With satellites circling the Earth, as shown in Figure 1.4, it is possible to scan the entire planet and have a complete image of the Earth that makes the application possible.

When a ship's AIS signal is transmitted to the satellite, it records and decodes the ship's identity before sending the signal to ground stations for further processing and distribution, meaning that some of the data processing is done aboard the

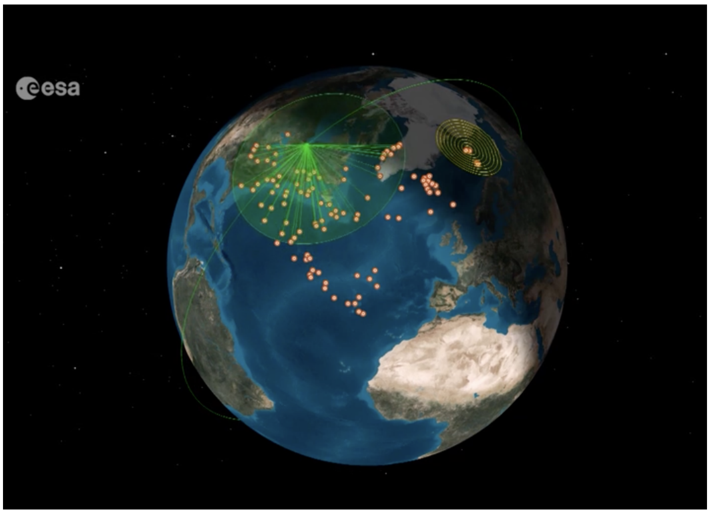


Figure 1.4:
ESAIL maritime microsatellite by ESA (4)

E-SAIL satellite. It is also capable of tracking more vessels than other satellites, an impressive feat considering the size of this satellite.

In 2022, the EU identified space as a strategic domain in the Strategic Compass and called for an EU space strategy for security and defense. It may seem that space services are a rather distant frontier from cyber attacks compared to others whose media resonance is certainly greater, but instead many of the activities carried out in space are critical to the functioning of society and the economy because they keep essential services running for government, private companies, and citizens. There are two cases to better understand the impact of cybersecurity attacks on satellite technologies. The first dates back to April 2023 and concerns China, which has declared its intention to build "cyber weapons" to intercept and monitor enemy satellites in order to knock out communications in possible future scenarios of war on Earth (17). Somewhat like what has already happened with the ViaSat KA-SAT satellite, which provides Internet connectivity to Europe, which was the target of a cyber attack at the beginning of the Russia-Ukraine war on Feb. 24, 2022, and in May 2023 is still under attack (18). Using wiper-type malware to format the modems of those who connected to that network segment, called "Acid Rain," with the intent of making service unavailable by preventing access to data or systems. In addition to the command center of the Ukrainian Armed Forces, however, a German energy company was also hit, losing control of more than 5,000 wind turbines. Thus, it is possible to define an interrelationship between cybersecurity and space: cybersecurity for space (examples are security of space applications, satellite technologies, protection of space assets and all satellite operating infrastructure) and cybersecurity from space for the use of satellite space technologies to strengthen cybersecurity on earth. In summary, technology development in this field can benefit the country, but at the same time exposes it to the risk of increasing the cyber attack surface. As the graph in Figure 1.5 shows, the scenario is constantly changing: as the number of operational satellites increases, so do satellite incidents.

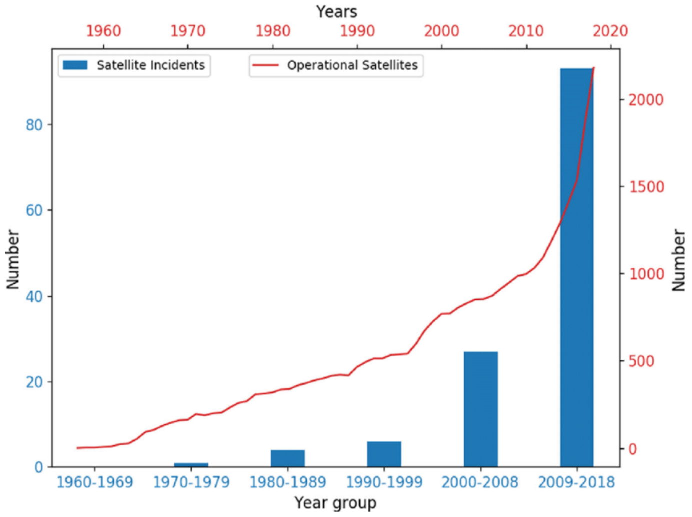


Figure 1.5:
Satellite incidents VS operational satellites (19)

“Cyber-attacks on the maritime industry’s operational technology (OT) systems have increased by 900% over the last three years with the number of reported incidents set to reach record volumes by year-end, latest data by cyber security firm Naval Dome reveals.” (20).

F. Akpan et al. (21), highlight the cybersecurity challenges faced by the maritime industry. One of the challenges mentioned is the threat of ransomware attacks targeting shipping companies. For example, in 2017, Danish shipping giant Maersk fell victim to the NotPetya ransomware attack, resulting in significant financial losses. This incident underscores the need for robust cybersecurity measures to prevent and mitigate such attacks. The same example is also reported by L. Jensen (22) as a cybersecurity risk in the supply chain. This incident underscores the need to strengthen cybersecurity collaboration and information sharing among all actors in the maritime supply chain. In fact, this paper focuses on the challenges to achieve cyber resilience in the maritime industry. The paper highlights the growing interconnectivity between ships, ports and other stakeholders, which opens potential entry points for cyber threats. For example, the integration of electronic data interchange (EDI) systems between ports and ships enables efficient data exchange, but also introduces potential risks if not adequately protected. Another challenge discussed is the vulnerability of operational technology (OT) systems on ships. OT systems, such as engine control systems and navigation systems, are increasingly connected to networks, making them potential targets for cyber attacks. For example, the U.S. Coast Guard reported an incident in which malware infected critical ship systems, causing disruptions in ship operations. This highlights the importance of protecting OT systems through network segmentation, access control, and regular patching.

In the paper, Y. Jo et al. (23) delve into cyberattack models targeting ship equipment using the MITRE ATTCK Framework, a comprehensive tool for categorizing cyber-attack techniques. The study presents a hypothetical scenario illustrating how adversaries could exploit vulnerabilities in vital ship systems, like navigation and engine controls, to compromise a vessel's functionality. An example discussed is the potential manipulation of the Automatic Identification System (AIS), allowing attackers to tamper with vessel identities or disrupt communications. This could lead to collisions or disruptions in maritime traffic by impersonating other vessels or altering their positions. The article underscores the importance of comprehending and countering various attack methods, including network reconnaissance, credential theft, and command and control manipulation. It emphasizes the necessity of implementing cybersecurity measures like network segmentation, intrusion detection systems, and robust access controls to thwart these attack vectors. These articles shed light on cybersecurity challenges confronting the maritime industry, encompassing issues such as ransomware attacks, vulnerabilities in operational technology, and the imperative for cyber-resilience and supply chain security. Real-world incidents and potential attack scenarios were discussed, underscoring the significance of robust cybersecurity strategies to safeguard maritime assets and operations. The utilization of Open-Source Intelligence (OSINT) within maritime intelligence was highlighted as crucial for gathering, monitoring, and analyzing publicly available data across online platforms, social media, and hacking forums. This serves to enhance situational awareness, risk assessment, and decision-making regarding potential threats or signs of compromise.

It was pointed out that as the rate of accumulation and scale of AIS data has increased, algorithms have progressed and become faster across various sources. This has led to an increase in the information value of AIS and the demand for data integrity to ensure reliable application design and analysis results. The mandatory implementation of AIS has led to the development of new applications using AIS-derived data. Today, AIS has reached a level of maturity that allows it to serve as a platform for further innovation, such as the integration of new technologies and data sources. For example, the European space agency ESA has developed a program called Integrated Applications Promotion (IAP) (24), which aims to promote the development of innovative commercial applications based on satellite data, including those using AIS data. These collaborations are important because they allow AIS data to be integrated with other types of data, such as satellite imagery, to improve understanding of maritime traffic and the situation on the high seas. The vast potential of satellite systems is currently underutilized. The added value that these systems can offer is considerable and is unknown to users and service providers. However, there are regulatory, organizational, and security challenges that need to be addressed in order to fully realize the potential of AIS as a platform for innovation.

Starting with the introduction of AIS as the first major innovation and explaining its evolution and new challenges, the status quo in terms of digitalization in the maritime industry was outlined. From this emerged the strategic role of data in the maritime sector, as a key concept. Hence, the research goal is to outline a high-reliability data ecosystem in which the rules, standards, and technologies used for data storage, security, and governance are defined to improve data quality, accessibility, and usability. In order to achieve this objective, it is fundamental to understantd digitalization in the maritime domain and the main benefits and challenges associated to the introduction of data spaces in this domain.

Chapter 2

Literature Review

This chapter presents the literature relevant to this research. The goal is to highlight how the literature address the need of a high-reliability data ecosystem creation.

2.1 Data Value

From all that has been said so far, it is evident that the value of the AIS and complementary systems is the data transmitted, and it will be fundamental to rethink an ecosystem around them as "the strategical alignment structure of the multilateral set of partners that need to interact in order for a focal value proposition to materialize" (25) or as described in the management conceptual framework by L. Thomas as "a community of hierarchically independent, yet interdependent heterogeneous participants who collectively generate an ecosystem output" (5). The existing literature offers a wide range of perspectives and approaches to address the concept of a data ecosystem. On this matter, the "pie model" by G. L. Romme et al. (26) offers a wide overview of the main feature of an ecosystem. A data ecosystem based on AIS data could involve heterogeneous actors and sectors to create a coordinated and modular environment (27) for the processing and use of such data. The minimum requirement of such an ecosystem would certainly lie in the quality of the data, so an integrated collection of data from other different sources (such as satellite data) and the capability to process and manage the collected data is required. Therefore, the entire data architecture that deals with the overall design and organization of the organization's data resources, including the rules, policies, standards, and technology used to manage and maintain the data needs to be reviewed. This must embrace data models, data structures, data flow, data storage, data security, and data governance. The final purpose is to have a well-designed data architecture that could improve the quality, accessibility, and usability of data and ensure that data is consistent, accurate, and reliable across different systems and applications.

From the literature, it is possible to extrapolate a key fundamental concept: collaboration among ecosystem participants is a crucial factor in the success and resilience of the ecosystem. This is not possible without adequate data security and protection that enables secure information sharing. As already mentioned, the value of data in the digital society increases day by day becoming a strategic resource, generating economic, social, and cultural value, as highlighted by C. Alaimo et al. (28). However, this value can be compromised whether data is not adequately protected. Therefore, data security becomes a key element in ensuring the continuity and prosperity of organizations and society as a whole. As reported by M. Jovanovic et al. (29), data security, together with platform architecture and governance, contribute to ensuring the trust of users and industry players but also to expanding the value of digital platforms that depend closely on data and are increasingly interconnected to deliver advanced services and functionality. In the paper "When Data Becomes Infrastructure and our Lives Depend on it" (30)the author E. B. Swanson affirms that data has become a critical infrastructure element for the functioning of our modern societies. Our lives are increasingly dependent on data, which is used in a wide range of sectors, from critical infrastructure to communications, from healthcare to the economy. Therefore, data security is essential to protect our society and preserve user trust in new technologies.

These concepts can be leveraged in rethinking data as the fulcrum of an ecosystem in which players act to realize a shared value proposition. The collaboration and sharing of information will have positive consequences on products, prices, services, and businesses on a diverse range of applications. The focal point is the design a data ecosystem with a properly defined and reliable data architecture that could have the potential to influence other fields of society, from monitoring raw materials to the tourism industry or the agriculture sector, from insurance to urban planning services and natural catastrophes estimation, and even monitoring illegal activities.

2.2 The emergence of public data spaces

Public data spaces (Appendix 3), unlike the private digital platforms discussed in the academic literature, present an ecosystem approach to digital commerce models. This approach creates new challenges and opportunities for both providers and users participating in such spaces. While private platforms often focus on direct interactions between the parties involved, public data spaces highlight a broader vision, involving multiple actors in a collaborative environment. Moreover, the public ecosystem consisting of a public data space differs significantly from an ecosystem based on private digital platforms. While these additional dependencies can increase the level of complexity for organizations seeking to use public data spaces, data spaces can provide new perspectives for digital innovation at the ecosystem level. Within platform ecosystems, the locus of digital innovations seems to be increasingly shifting from the organizational level to the ecosystem level. In the context of public data spaces, this implies that innovation is not driven solely by individual organizations, but by the collective efforts and interactions of the various participants within the ecosystem. This change requires a new mindset in which organizations recognize the importance of collaboration and co-innovation with external partners. In addition, public data spaces lack a single platform owner who attempts to manage the platform ecosystem, holding the actor accountable. The absence of a single platform owner responsible for managing the ecosystem introduces a particular challenge. Unlike private platforms with clear ownership structures, public data spaces lack a central authority that can oversee and govern the entire ecosystem. This decentralized nature can make decision-making and coordination more complex. However, it also opens the door to more democratic and community-driven innovation, where different actors collectively contribute to shaping the direction of the ecosystem. This offers the opportunity for distributed ownership and a shared sense of responsibility among ecosystem participants. In more detail, the comparison between private digital platforms and public data spaces is well summarised by the authors D. Beverungen, et al. (6) through a table, shown in the Figure 2.1.

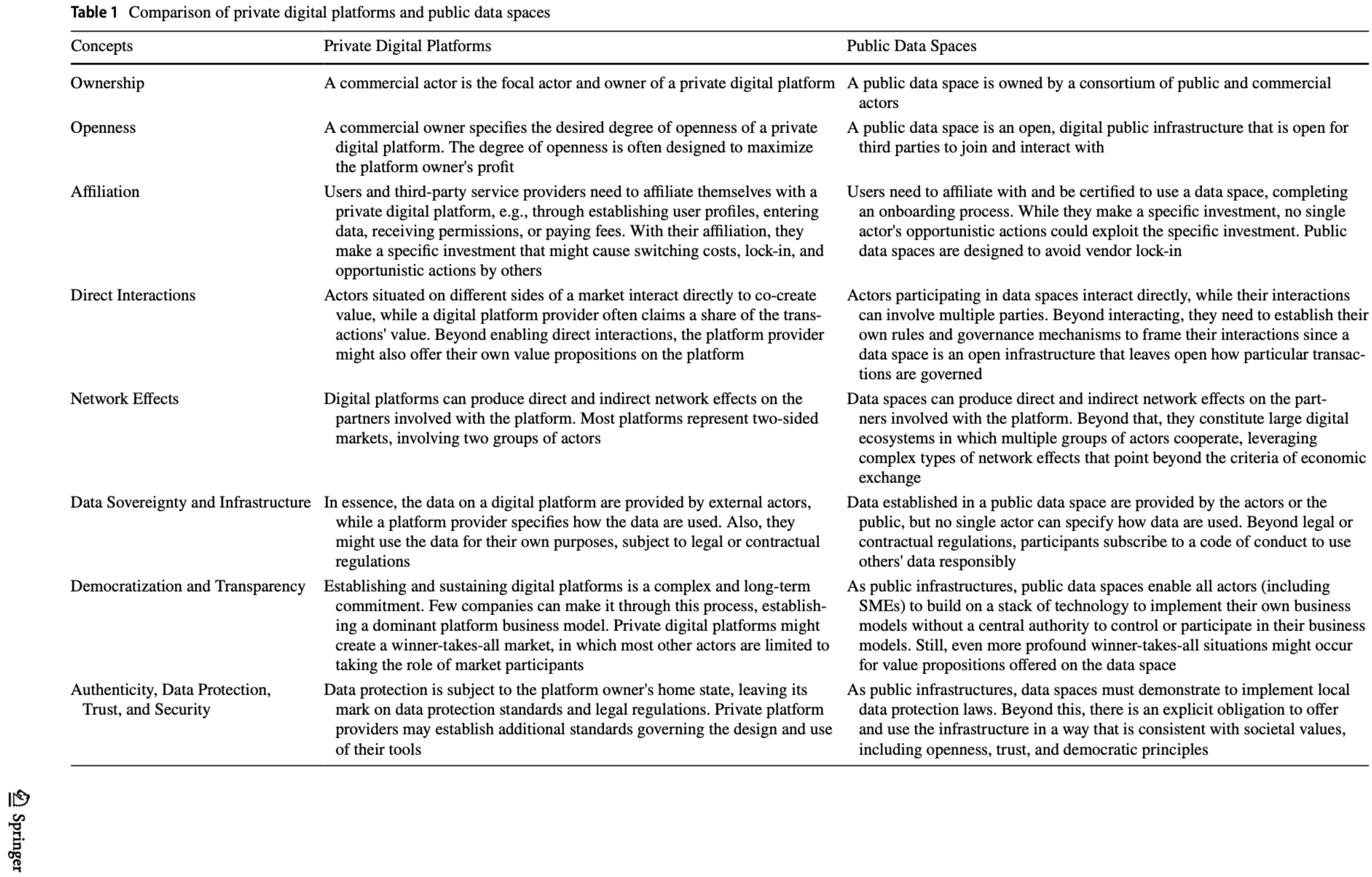


Table 1 Comparison of private digital platforms and public data spaces

Concepts	Private Digital Platforms	Public Data Spaces
Ownership	A commercial actor is the focal actor and owner of a private digital platform	A public data space is owned by a consortium of public and commercial actors
Openness	A commercial owner specifies the desired degree of openness of a private digital platform. The degree of openness is often designed to maximize the platform owner's profit	A public data space is an open, digital public infrastructure that is open for third parties to join and interact with
Affiliation	Users and third-party service providers need to affiliate themselves with a private digital platform, e.g., through establishing user profiles, entering data, receiving permissions, or paying fees. With their affiliation, they make a specific investment that might cause switching costs, lock-in, and opportunistic actions by others	Users need to affiliate with and be certified to use a data space, completing an onboarding process. While they make a specific investment, no single actor's opportunistic actions could exploit the specific investment. Public data spaces are designed to avoid vendor lock-in
Direct Interactions	Actors situated on different sides of a market interact directly to co-create value, while a digital platform provider often claims a share of the transactions' value. Beyond enabling direct interactions, the platform provider might also offer their own value propositions on the platform	Actors participating in data spaces interact directly, while their interactions can involve multiple parties. Beyond interacting, they need to establish their own rules and governance mechanisms to frame their interactions since a data space is an open infrastructure that leaves open how particular transactions are governed
Network Effects	Digital platforms can produce direct and indirect network effects on the partners involved with the platform. Most platforms represent two-sided markets, involving two groups of actors	Data spaces can produce direct and indirect network effects on the partners involved with the platform. Beyond that, they constitute large digital ecosystems in which multiple groups of actors cooperate, leveraging complex types of network effects that point beyond the criteria of economic exchange
Data Sovereignty and Infrastructure	In essence, the data on a digital platform are provided by external actors, while a platform provider specifies how the data are used. Also, they might use the data for their own purposes, subject to legal or contractual regulations	Data established in a public data space are provided by the actors or the public, but no single actor can specify how data are used. Beyond legal or contractual regulations, participants subscribe to a code of conduct to use others' data responsibly
Democratization and Transparency	Establishing and sustaining digital platforms is a complex and long-term commitment. Few companies can make it through this process, establishing a dominant platform business model. Private digital platforms might create a winner-takes-all market, in which most other actors are limited to taking the role of market participants	As public infrastructures, public data spaces enable all actors (including SMEs) to build on a stack of technology to implement their own business models without a central authority to control or participate in their business models. Still, even more profound winner-takes-all situations might occur for value propositions offered on the data space
Authenticity, Data Protection, Trust, and Security	Data protection is subject to the platform owner's home state, leaving its mark on data protection standards and legal regulations. Private platform providers may establish additional standards governing the design and use of their tools	As public infrastructures, data spaces must demonstrate to implement local data protection laws. Beyond this, there is an explicit obligation to offer and use the infrastructure in a way that is consistent with societal values, including openness, trust, and democratic principles

Figure 2.1: *Comparison of private digital platforms and public data spaces*

Use of the term has increased significantly in recent years, driven by growing awareness of the importance of data sharing, interoperability and information networking. In recent years, the concept of public data space is beginning to be embraced by a growing number of governments and businesses. In Europe, for example, the European Commission launched the Open Data 500 project in 2018, with the goal of making 500 datasets publicly available by the end of 2020.

The reference paper for this part of the literature is "From Private Digital Platforms to Public Data Spaces: Implications for the Digital Transformation" by D. Beverungen et al. (6), in which the author discusses the potential of public data spaces to foster innovative progress and digital transformation. The author argues that public data spaces can provide a more open and collaborative environment for data sharing and collaboration, which can lead to new insights and innovations. In particular, the paper gives a precise definition of public data spaces as "digital infrastructures that enable the open and coordinated use of data by a wide range of actors, including governments, businesses, and citizens."

Furthermore, the author highlights the potential benefits of public data spaces that could play a significant role in digital transformation, with the acknowledgment that there are challenges that need to be addressed before public data spaces can be fully realized. For my research, it could be helpful to draw attention to some of the main strengths and challenges of public data spaces.

STRENGHTS

- increase transparency and accountability: public data platforms have the potential to simplify citizens' ability to oversee the utilization of their data and ensure that governments and corporations are answerable for their conduct. Through facilitating data accessibility in a more inclusive and available manner, individuals can gain insights into the utilization of their information, fostering an atmosphere of transparency in data handling. This enhanced transparency can cultivate increased trust among citizens, governments, and businesses, as their actions become more apparent and open to examination. Moreover, it promotes conscientious data administration and ethical deliberations, adhering to the principles of FRAND (fair, reasonable, and non-discriminatory)

- improving efficiency and innovation: public data spaces hold the promise of boosting government service efficiency and fostering innovation by facilitating businesses' access to and utilization of data. Governments can streamline their processes by harnessing data for well-informed decision-making, optimizing resource allocation, and crafting more precise policies. Furthermore, these data spaces serve as a fertile breeding ground for innovation, allowing businesses to tap into the data resources to create novel products, services, and solutions that tackle societal issues. This harmonious relationship between data accessibility, government efficacy, and entrepreneurial innovation contributes to a more agile and adaptable ecosystem

• enhancing social welfare: public data spaces can have a substantial impact on enhancing the overall welfare of society, especially during critical periods like public health crises or conflicts. These platforms, which offer access to up-to-date information, enable governments and institutions to better monitor and respond to emergencies. This accessibility of real-time data promotes greater collaboration among different entities, resulting in more efficient crisis management and ultimately contributing to the well-being of society

CHALLENGES

• data quality: The effectiveness of public data spaces hinges on the quality of the data they contain. It is of paramount importance to ensure that the data is not only accurate but also complete and regularly updated. Data quality stands as a foundational concern in the context of public data spaces. To be a valuable resource for users and organizations, the information shared within these spaces must be dependable and trustworthy. If the data is inaccurate or outdated, it can lead to misguided decisions and potentially damage the overall credibility of the entire public data ecosystem. To maximize the utility of public data spaces, it is crucial to establish mechanisms for verifying and maintaining data quality. In the Information System Research field, the existing body of literature emphasizes data as a pivotal resource for innovation and value creation. Additionally, the emerging interest in promoting data spaces and related ecosystems for public benefit is beginning to attract research attention. (31) (32). A particular attention goes to the paper (33), in which the authors explore analytics-based services (ABS) for generating value for customers through data analysis. The researchers have categorized four unique typologies of ABS that underscore the intentions of providers when furnishing these services: 1) enhancing data accessibility for customers, 2) presenting insights derived from data, 3) offering data-driven suggestions, and 4) facilitating innovative approaches to conducting business operations. This research underscores the vital role of data as a strategic asset and demonstrates how companies are leveraging data and analytics to create innovative, customer-centric, and value-generating business opportunities. It provides valuable insights into the potential for innovation and value creation through existing data and analytics solutions. Furthermore, the emergence of data spaces is expected to expand the volume and diversity of available data, along with the techniques for analyzing them.This expansion is anticipated to lead to the development of new analytics-based services, opening up additional opportunities for businesses to create value.

• data governance: The establishment of clear and well-defined rules and regulations governing data sharing and usage within public data spaces is of utmost importance. This serves the dual purpose of safeguarding individuals' privacy and enhancing security while fostering a sense of trust among all participants involved. Effective data governance is a fundamental requirement for creating a framework that strikes a balance between openness and responsible utilization. By outlining specific guidelines pertaining to data access, sharing, and management, public data spaces can effectively mitigate potential risks associated with unauthorized access or the inappropriate use of sensitive information. Robust data governance practices ensure that individuals engaging with the public data ecosystem can do so with confidence, knowing that their data rights and privacy are upheld and respected. It's important to note that public data spaces are constructed upon infrastructure provided by public entities. While this configuration offers various advantages, such as enabling data protection and security measures, there may still be incentives for stakeholders to influence and shape the infrastructure to their advantage. Consequently, questions regarding how the governance structure of the ecosystem can influence the digital transformation of organizations and how these organizations are adapting to the evolving platform landscape become crucial considerations.

From what has been said so far, public data spaces can be applied to maritime domain in order to collect and share maritime data from a variety of sources, such as ships, ports, and weather stations. This can help to improve the quality and reliability of maritime data by providing a more comprehensive and up-to-date picture of the maritime environment. Furthermore, public data spaces can be the first step to develop and deploy new maritime data applications. For example, public data spaces could be used to develop new reliable applications that track maritime traffic, predict weather conditions, or identify potential hazards. These applications can help to improve the safety and efficiency of maritime operations. Last but not least, public data spaces can be used to foster collaboration between different stakeholders in the maritime industry, but not only, creating a true data ecosystem.For example, public data spaces could be used to bring together ship owners, port operators, and government agencies to share data and develop solutions to common problems. This can help to improve the overall safety and efficiency of the maritime industry. Of course, the public data spaces challenges persist also in this case, in terms of security and privacy of maritime data, and rules and regulations for how maritime data can be shared and used.

In a nutshell, public data spaces are promoted to guarantee data sovereignty to organizations by enabling control and optimization of data access, thus ensuring a secure, reliable, and transparent space for data access and exchange between different parties. The pooling of data of the same type or complementary nature may enable organizations to co-create innovations that are not possible with the resources of a single organization.

2.3 European data spaces frameworks

In February of 2020, the European Commission announced the European Data Strategy (34). The primary objective of this strategy is to establish a unified data market within the EU, facilitating the efficient and secure exchange of data across various sectors. This initiative aligns with the Commission's overarching goal of advancing the European data economy while upholding fundamental European values, including self-determination, privacy, transparency, security, and fair competition. To realize this vision, it is imperative that the regulations governing data access and utilization are equitable, transparent, and easily implementable. This necessity becomes even more pronounced as the European data economy experiences rapid growth, surging from 301 billion euros (2.4% of GDP) in 2018 to an anticipated 829 billion euros (5.8% of GDP) by 2025.

Initiatives such as International Data Spaces (IDS) and GAIA-X are essential for creating domain-specific data spaces and for their interconnection in a European data and infrastructure ecosystem. They define standards and provide software architectures for fair data sharing and reliable data spaces. The IDS and GAIA-X projects are closely aligned to enable seamless integration of the architectures and supporting processes.

2.3.1 International Data Spaces - IDS

The International Data Spaces (IDS) (35) project was launched in 2015 by the German Federal Ministry for Education and Research with the goal of designing and prototyping a distributed software architecture for data sovereignty. In parallel, the IDS Association (IDSA) was established as a non-profit industrial association and resumed the research work to further develop it within the IDS Reference Architecture Model (IDS RAM). The IDS RAM is a description of the architecture of a technology-independent data space. The central software components described in the IDS RAM (e.g., IDS Connector, IDS Broker, IDS Clearing House) have been incorporated into the DIN SPEC 27070 standard, which provides a blueprint for a secure gateway for reliable data exchange. As of August 2023, IDSA counts 145 members from 28 countries. The main goals of IDS can be summarized by 5 concepts strongly linked to each other:

- Trust. It is an essential element in facilitating data sharing within a data ecosystem. Participants need to have confidence in both the reliability of the systems involved and the assurance that other members of the data ecosystem acquire valuable data. It is crucial that this data is utilized in strict adherence to the usage policies set forth by the data provider. Prior to gaining access to the ecosystem, each participant and IDS software will undergo a certification process.

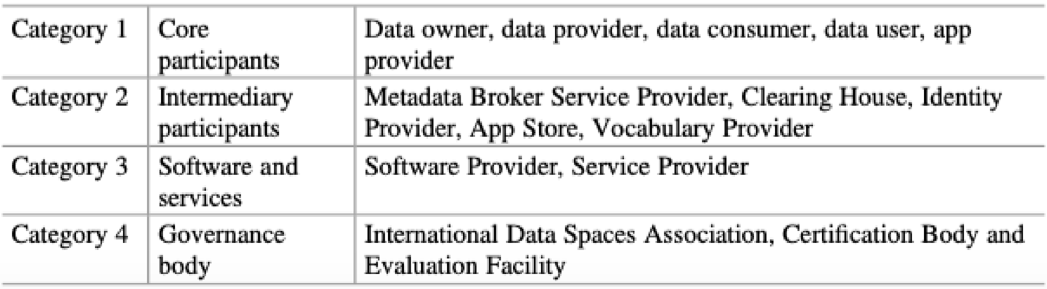
- Security. All IDS systems must meet state-of-the-art security standards to ensure data trust and sovereignty. Therefore, security requirements must be included in the certification criteria.

- Data sovereignty. It can be defined as a natural person's or corporate entity's capability of being entirely self-determined with regard to its data. This means that the data owner can define usage restrictions for their data before sharing it with consumers. Data consumers must accept the usage restrictions.

- Data ecosystem. It enables new business models that individual actors cannot achieve alone due to not having access to the entirety of the data. No single actor possesses all the required data to offer an innovative service. Hence, a data ecosystem requires a data space to enable these novel and innovative services.

- Standardized interoperability. It is crucial for establishing the data environment, given that diverse data ecosystems will engage in the exchange of assorted data formats and communication protocols. The critical aspect of achieving this lies in the standardization of interoperability, which allows each system to seamlessly interact within the framework of the IDS. To achieve this, the IDS architecture is meticulously outlined within a reference architecture model. This model encompasses a comprehensive semantic description of data and endpoints, and certification serves as the mechanism to validate that each system aligns with the prescribed architecture and effectively utilizes the information model as stipulated. Additionally, it's worth noting the DIN 27070 standard, which specifically defines the IDS connector.

As already mentioned, data spaces must be seen in the context of the ecosystem they support and the underlying software infrastructure. If we focus on the ecosystem point of view, we can immediately define different actors that may choose to enter the ecosystem under different roles depending on the level of interaction and organization, detailed in the Figure 2.2.



Category 1	Core participants	Data owner, data provider, data consumer, data user, app provider
Category 2	Intermediary participants	Metadata Broker Service Provider, Clearing House, Identity Provider, App Store, Vocabulary Provider
Category 3	Software and services	Software Provider, Service Provider
Category 4	Governance body	International Data Spaces Association, Certification Body and Evaluation Facility

Figure 2.2:
International Data Space role categories

The core participants are actively engaged and mandatory whenever data interchange occurs within the IDS:

- A Data Owner is described as an individual or a legal entity responsible for creating and/or maintaining control over data. This control is established through the formulation of usage policies and the authorization of data access. Data ownership needs, at least, the technical capability and accountability for defining usage agreements and enabling data access.

- A Data Provider role involves supplying data for exchange between a data owner and a data consumer, utilizing software components aligned with the IDS reference architecture model. In most instances, although not obligatory, the data provider and the data owner may be one and the same entity. Establishing a direct connection between a data provider and a data consumer is feasible. To facilitate a data request, the data provider can transmit relevant metadata to a broker service. Additional actions may include transaction logging at a clearing house and the enhancement or transformation of data through the utilization of data apps.

- A Data Consumer receives data from a data provider. From a standpoint of business process modeling, the data consumer represents the counterpart to the data provider, and as such, the activities carried out by the data consumer closely resemble those performed by the data provider. Before connecting to a data provider, the data consumer has the option to search for existing datasets by submitting an inquiry to a Broker Service Provider.

- A Data User, such a data owner, is the legal entity holding the legal authority to utilize the data owned by a data owner, as stipulated by the usage policy. In many cases, the data user is the same of data consumer. Nevertheless, situations may arise where different participants assume these roles.

- An App Provider develops data applications designed for use within the IDS framework. To be deployed, a data application must adhere to the IDS system architecture. Furthermore, data applications can undergo certification by an authoritative body to enhance trust in these applications. Application providers are obligated to furnish comprehensive descriptions of each data application, including semantics, functionality, interfaces, and other relevant aspects, using metadata conforming to an established metadata model.

Intermediary participants function as trusted entities, encompassing roles such as Metadata Broker Service Provider, Clearing House, App Store, Vocabulary Provider, and Identity Provider. These roles should exclusively be entrusted to reputable organizations. The federated architecture of IDS encompasses the operation of (virtually) centralized components that oversee distinct aspects of service provisioning within the data realm.

Software and service providers encompass IT companies that offer software and/or services (for instance, in a software-as-a-service model) to IDS participants. Roles falling under this category include App Providers, Service Providers, and Software Providers. A software provider furnishes the software to implement the functionalities required by the IDS. Unlike data applications, the software isn't supplied via the App Store, but rather distributed through the usual channels of software providers and employed based on individual agreements between the software provider and the user (such as a data consumer, data provider, or intermediary service provider). If a participant lacks the necessary technical infrastructure to engage with the IDS, they can transfer the data intended to be available within the IDS to a Service Provider that hosts the essential infrastructure for other organizations. This role also encompasses providers offering additional data services (such as data analysis, data integration, data cleansing, or semantic enrichment) to enhance the quality of data exchanged within the IDS.

The governance role is based on the Certification Body and the International Data Spaces Association. The International Data Spaces Association operates as a non-profit organization dedicated to fostering the ongoing advancement of Data Spaces, with its membership comprising predominantly of major industrial corporations, IT firms, small and medium-sized enterprises, research organizations, and industry associations. Oversight of participant certification and the technical core components within the IDS falls under the responsibility of the Certification Body and the Evaluation Facility.

2.3.2 Global Architecture for Interoperable Automated Data Spaces - Gaia-X

The initiative known as Gaia-X (36), which was initiated in 2019 under the collaboration of the German and French Ministers of Economy, aimed to safeguard the autonomy of businesses over their industrial data, even in situations where such data is stored on cloud service platforms located outside the European Union, preventing any loss of control.It was officially authorized on December 21, 2021. The project aims at data sovereignty in a broad context,such as IDSA does, from data sharing and exchange to data storage and data handling on cloud platforms, as stated in the position paper (37). The Federation Services form the core of the GAIA-X architecture. It will offer a range of "federation services" that will enable interactions among all participants without any specific company assuming a dominant role in controlling the flow of information.They comprise a federated catalogue of distributed services, sovereign data exchange, identity and trust management, and compliance services.

Nowadays, Gaia-X counts more than 370 members, representing both users and 21 hubs across the world. The establishment of national Gaia-X hubs, functioning as independent think tanks, ambassadors of the Association, further accelerates the emergence of new data spaces and use cases at the national level before these are subsequently extended to a European and global scale. The membership increases for 40% since 2021 and the country membership increases by 24% since 2021. Gaia-X partners share the notion that data spaces will play a role in the digital business similar to what the web did 40 years ago in propelling the growth of the Internet. In order to show the increasing relevance of data spaces and specifically of the Gaia-X initiatives, I found an histogram of Gaia-X Member distribution by country dates on end of April 2022 and so I decided to investigate the actual members dates on August 2023. In order to do that, I collected the data about the actual members from the website and in analyzing the data I used the Python programming language and some supporting libraries, including Pandas [1] for data manipulation and Matplotlib [2] for visualization in order to create the bar graph with the comparison of the members from April 2022 to August 2023, as shown in Figure 2.3.Furthermore, I decided to used the Python programming language to create a map to have a graphical impact of the distribution of members in the world (Figure 2.4). For manipulating geospatial data, I used the Geopandas [3] library, and for creating the interactive map I leveraged Folium [4].

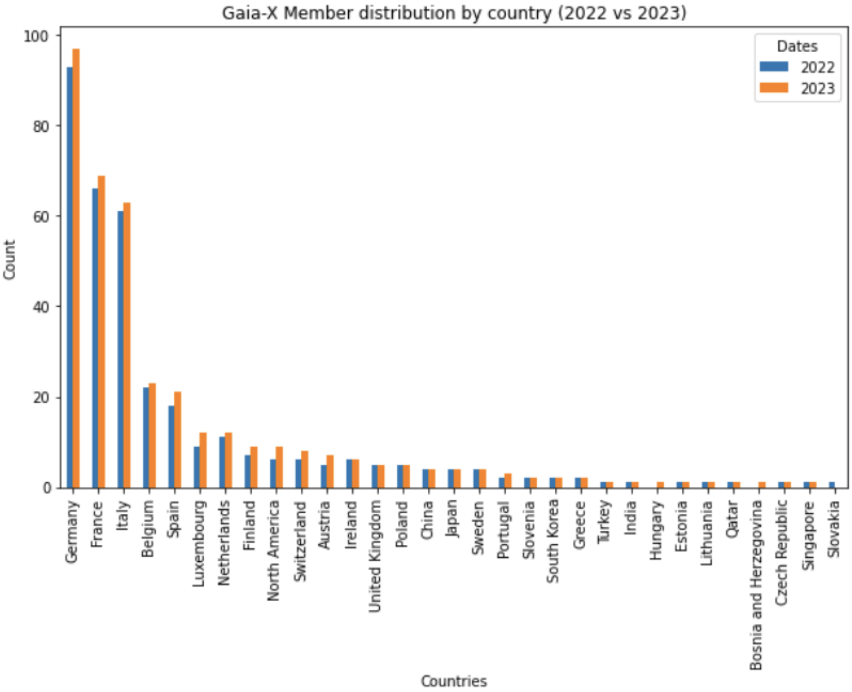


Figure 2.3: *Gaia-X Member distribution by country (2022 vs 2023)*

[1] https://pypi.org/project/pandas/

[2] https://pypi.org/project/matplotlib/

[3] https://pypi.org/project/geopandas/

[4] https://pypi.org/project/folium/

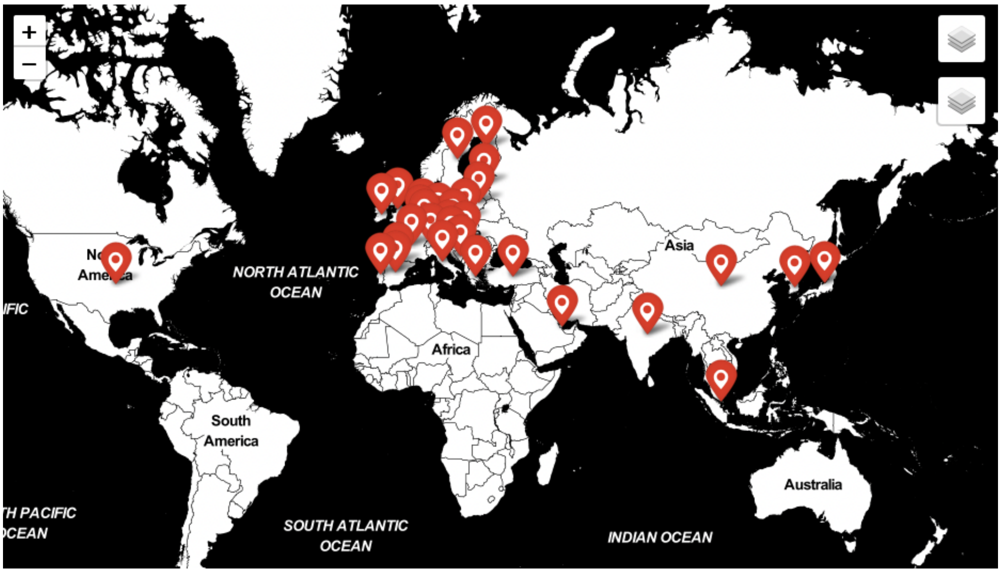
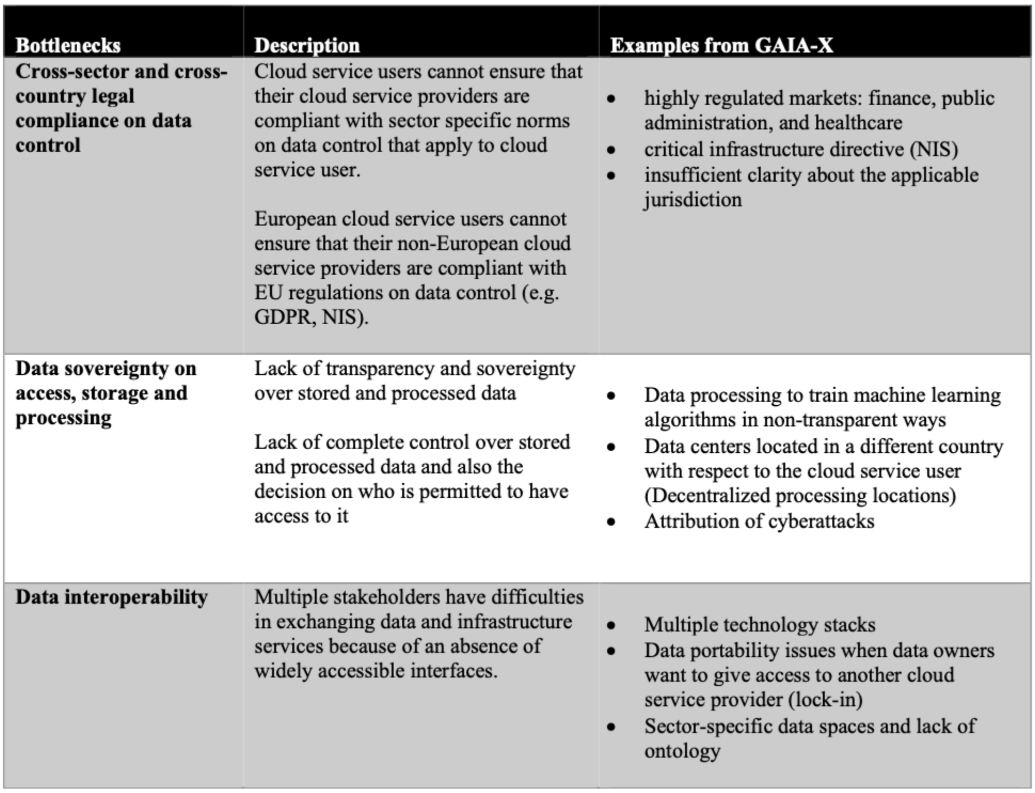


Figure 2.4: *Gaia-X members distribution in the world*

Gaia-X is an initiative aimed at providing a regulatory and technical framework that supports the creation of data ecosystems and infrastructures.The infrastructure serves as the underlying framework for creating data spaces (as described in the European Data Strategy), which will lay the groundwork for "advanced intelligent services." These services will be centered around "collaborative use cases," wherein multiple parties agree to share data with the aim of enhancing supply chain process efficiency (smart and interconnected supply chains result in savings of around 20%) and bringing new value to competitive situations (such as data spaces concerning mobility where each provider shares their routes, real-time traffic data, and costs, facilitating the development of innovative mobility services).

As identified by the authors in the paper "Data control coordination in cloud-based ecosystems: The EU GAIA-X Ecosystem" (38), the main challenge of this ecosystems is about control bottlenecks: uncoordinated data control limits value creation. To address these data control bottlenecks (summarized in the table in the Figure 2.5), coordination among actors within an ecosystem is essential, along with establishing multilateral data control agreements prior to engaging in innovative activities that lead to value co-creation. For instance, Gaia-X is constantly ongoing and employs the IDS RAM to ensure data usage controls and compliance.



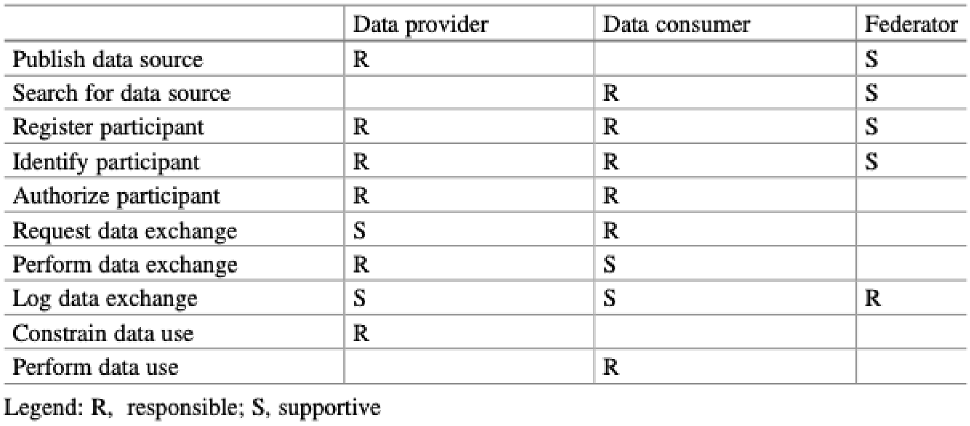
Bottlenecks	Description	Examples from GAIA-X
Cross-sector and cross-country legal compliance on data control	Cloud service users cannot ensure that their cloud service providers are compliant with sector specific norms on data control that apply to cloud service user. European cloud service users cannot ensure that their non-European cloud service providers are compliant with EU regulations on data control (e.g. GDPR, NIS).	• highly regulated markets: finance, public administration, and healthcare • critical infrastructure directive (NIS) • insufficient clarity about the applicable jurisdiction
Data sovereignty on access, storage and processing	Lack of transparency and sovereignty over stored and processed data Lack of complete control over stored and processed data and also the decision on who is permitted to have access to it	• Data processing to train machine learning algorithms in non-transparent ways • Data centers located in a different country with respect to the cloud service user (Decentralized processing locations) • Attribution of cyberattacks
Data interoperability	Multiple stakeholders have difficulties in exchanging data and infrastructure services because of an absence of widely accessible interfaces.	• Multiple technology stacks • Data portability issues when data owners want to give access to another cloud service provider (lock-in) • Sector-specific data spaces and lack of ontology

Figure 2.5: *Data control bottlenecks*

The Framework builds upon the evolution of the X-Model and enables the transition from disjoint data and infrastructure ecosystems, to composable, interoperable and portable cross-sector data sets and services. The Gaia-X framework aims to connect the Data and Infrastructure Ecosystems and relies on 3 conceptual pillars to achieve that:

• Gaia-X Compliance: decentralized services to enable objective and measurable trust. Gaia-X defines a set of rules and principles that must be followed by organizations that want to participate in the Gaia-X ecosystem. These rules and principles are designed to ensure that data is shared in a secure and reliable manner.

• Data Spaces and Federations: interoperable and portable cross-sector datasets and services. Gaia-X defines the concept of a data space, which is a virtual environment where data can be shared and exchanged. Data spaces can be public or private, and they can be used to share a variety of data types, including personal data, business data, and scientific data. Gaia-X defines the concept of a Federator, which is a software component that facilitates the interaction between different data spaces. The Federator ensures that data is exchanged in a secure and reliable manner, and it also ensures that the rules and principles of Gaia-X are followed. The Figure 2.6 shows the role of the data provider, data consumer and Federator:



	Data provider	Data consumer	Federator
Publish data source	R		S
Search for data source		R	S
Register participant	R	R	S
Identify participant	R	R	S
Authorize participant	R	R	
Request data exchange	S	R	
Perform data exchange	R	S	
Log data exchange	S	S	R
Constrain data use	R		
Perform data use		R	

Legend: R, responsible; S, supportive

Figure 2.6:
Data space roles and responsibilities

• Data Exchange: anchored contract rules for access and data usage. Gaia-X defines a trust framework that provides a way for organizations to assess the trustworthiness of other organizations that want to participate in the Gaia-X ecosystem. The trust framework is based on a set of criteria, such as the organization's security practices and its compliance with the Gaia-X rules and principles.

For each of these pillars there are 3 types of deliverables: Functional specifications, Technical Specifications and Software.

The Gaia-X framework is still under development and it has the potential to revolutionize the way data is shared and exchanged in Europe, and it could help to boost innovation and economic growth. The framework aims to provide:

• Security. Gaia-X is designed to ensure that data is shared in a secure and reliable manner. The framework includes a number of security features, such as encryption and access control.

• Interoperability. Gaia-X supports a variety of standards and technologies, which makes it interoperable with other systems. This makes it easy to share data between different organizations.

• Federated governance. Gaia-X is a federated framework, which means that there is no central authority. This makes it more flexible and adaptable to different needs.

• Openness. Gaia-X is an open framework, which means that it is open to all organizations. This makes it more inclusive and accessible.

2.4 Research Question

The authors of the paper ”From Private Digital Platforms to Public Data Spaces: Implications for the Digital Transformation” (6) present a set of relevant research questions in order to inspire researchers aiming to investigate this topic, summarized in the Figure 2.7. Based on the questions already formulated in the literature, the research question to address is:

How to build a public data space for the digital transformation of maritime activites? What technical foundations are required to support secure and trusted data and application sharing in the maritime sector?

The aim of this research is to investigate a new field of application of public data spaces to provide interesting considerations about the possible adoption and development of public data spaces in the maritime domain.Hence, the research contribution is the definition of the requirements for a reliable data space where public and private players of the maritime domain can align their interests to pursue digital transformation in their activities.

From an academic perspective, the research can extend the existing literature by providing interesting theoretical considerations on the benefits and challenges of the technology in the maritime sector. From a managerial perspective, the contribution of this research is about providing an inspiring and innovative example that can serve as a guide on how to develop and implement this technology at the European level in the maritime sector or in sectors with similar characteristics.

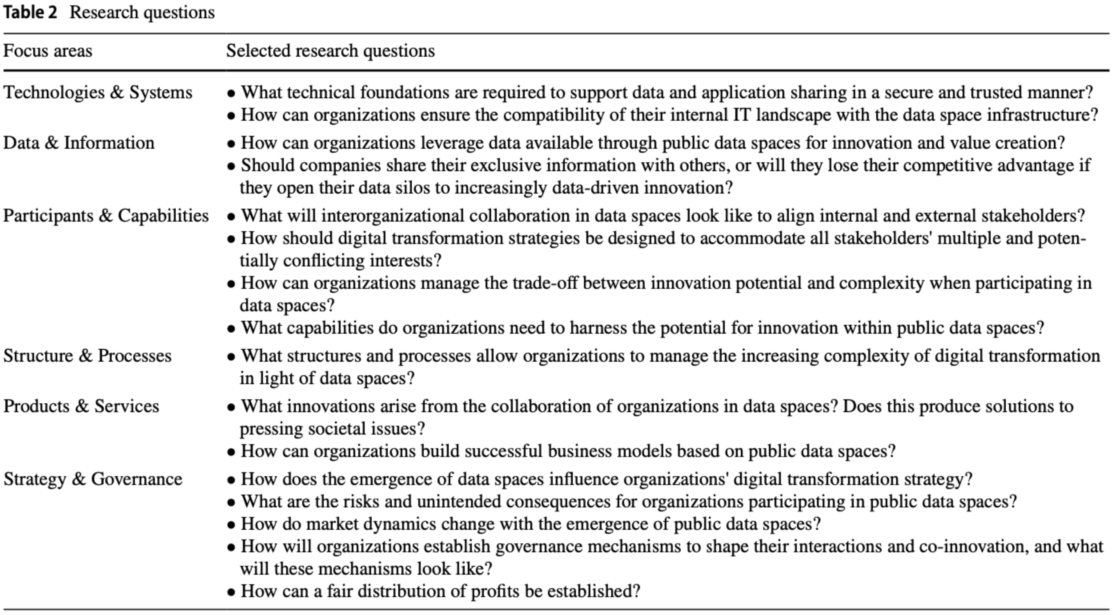


Table 2 Research questions

Focus areas	Selected research questions
Technologies & Systems	• What technical foundations are required to support data and application sharing in a secure and trusted manner? • How can organizations ensure the compatibility of their internal IT landscape with the data space infrastructure?
Data & Information	• How can organizations leverage data available through public data spaces for innovation and value creation? • Should companies share their exclusive information with others, or will they lose their competitive advantage if they open their data silos to increasingly data-driven innovation?
Participants & Capabilities	• What will interorganizational collaboration in data spaces look like to align internal and external stakeholders? • How should digital transformation strategies be designed to accommodate all stakeholders' multiple and potentially conflicting interests? • How can organizations manage the trade-off between innovation potential and complexity when participating in data spaces? • What capabilities do organizations need to harness the potential for innovation within public data spaces?
Structure & Processes	• What structures and processes allow organizations to manage the increasing complexity of digital transformation in light of data spaces?
Products & Services	• What innovations arise from the collaboration of organizations in data spaces? Does this produce solutions to pressing societal issues? • How can organizations build successful business models based on public data spaces?
Strategy & Governance	• How does the emergence of data spaces influence organizations' digital transformation strategy? • What are the risks and unintended consequences for organizations participating in public data spaces? • How do market dynamics change with the emergence of public data spaces? • How will organizations establish governance mechanisms to shape their interactions and co-innovation, and what will these mechanisms look like? • How can a fair distribution of profits be established?

Figure 2.7: *Research questions by D. Beverungen, et al.*

Chapter 3

Theoretical Development

The GAIA-X architecture in combination with the IDS RAM forms a “blueprint” for data space implementation. Starting from the knowledge gained by the literature, this chapter aims to give a contribution to the literature review in order to define the necessary requirements of a public data space in maritime domain. In particular, the research addresses the topic from the two main maritime sector perspectives already presented in the introduction section, as trade sector and as administrative services provider. To achieve these goals, the chapter is divided into a first section in which a framework is outlined to understand the complexity of the data fragmentation problem in the maritime sector and the importance of a public data space to solve it, the second section will go into more detail on the specific requirements of a data space for the maritime domain.

3.1 Framework design to ensure high reliability of maritime data

1. *Current situation*

The maritime industry is characterized by increasing digitization and the increased use of technologies such as big data, artificial intelligence, and the Internet of Things (IoT). This digitization has led to an increase in the amount of data available in the maritime sector, but it has also highlighted the need for improved data sharing and collaboration.

2. *Problem*

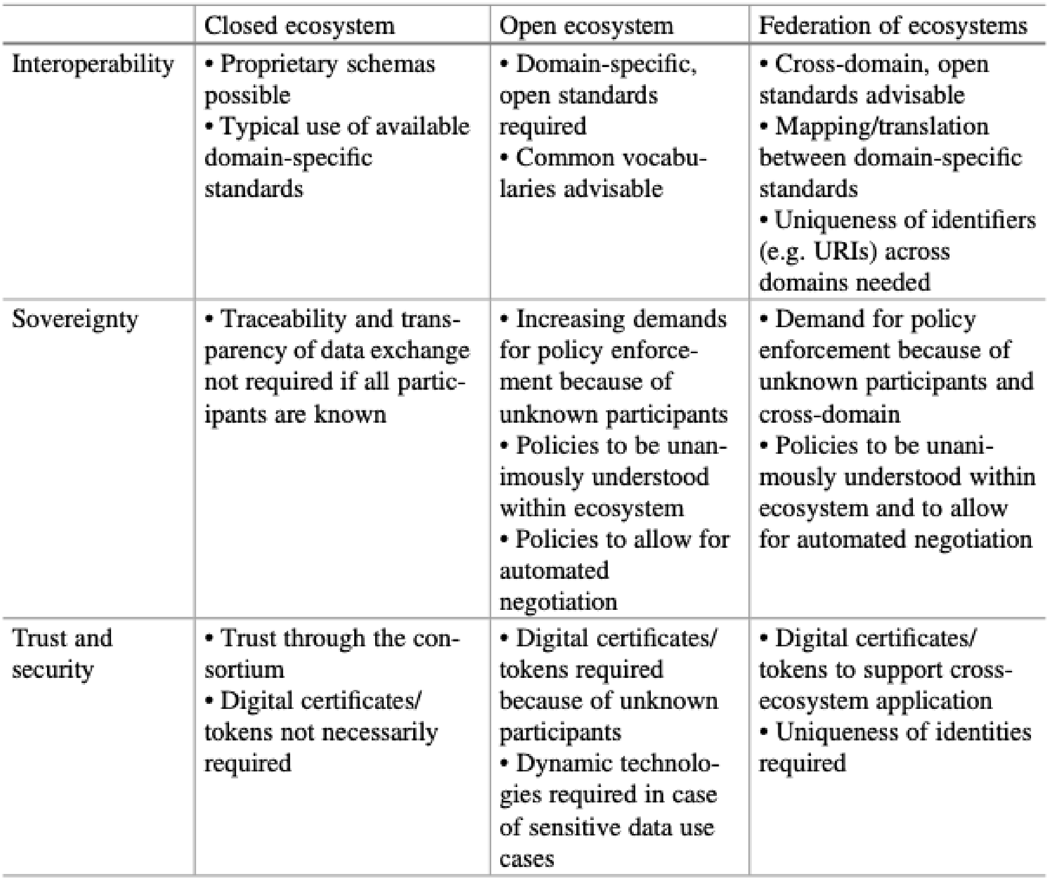
The main problem in the maritime sector is data fragmentation. Currently, maritime data are distributed across a wide range of sources, including ships, surveillance systems, port authorities, and other organizations. This fragmentation makes it difficult to obtain a complete view of the maritime domain and make informed decisions.

3. *Solution*

The proposed solution in this research is the creation of a public data space for the maritime domain. A public data space is a secure and reliable environment in which data can be shared and collaborated by different organizations. This could enable the development of new maritime applications and services that improve safety, efficiency, and sustainability.

4. *Key concepts*

The 3 main aspects to consider are: interoperability, sovereignty, and trust and security. These 3 concepts can be declined in different ways depending on the level of ecosystem involved, as the Figure 3.1 shows (39).



	Closed ecosystem	Open ecosystem	Federation of ecosystems
Interoperability	• Proprietary schemas possible • Typical use of available domain-specific standards	• Domain-specific, open standards required • Common vocabularies advisable	• Cross-domain, open standards advisable • Mapping/translation between domain-specific standards • Uniqueness of identifiers (e.g. URIs) across domains needed
Sovereignty	• Traceability and transparency of data exchange not required if all participants are known	• Increasing demands for policy enforcement because of unknown participants • Policies to be unanimously understood within ecosystem • Policies to allow for automated negotiation	• Demand for policy enforcement because of unknown participants and cross-domain • Policies to be unanimously understood within ecosystem and to allow for automated negotiation
Trust and security	• Trust through the consortium • Digital certificates/ tokens not necessarily required	• Digital certificates/ tokens required because of unknown participants • Dynamic technologies required in case of sensitive data use cases	• Digital certificates/ tokens to support cross-ecosystem application • Uniqueness of identities required

Figure 3.1:
Data space features VS level of ecosystem

5. *Gains*

The benefits that a public data space used in the maritime sector would bring are many in terms of efficiency, security and sustainability, both for navigation data and for the services the maritime sector provides as a public administration to the citizen. For navigation data, a public data space would allow data from different sources, such as automatic identification systems (AIS), satellite systems, and surveillance systems, to be shared and integrated. This would provide a more comprehensive view of maritime traffic, improving the safety and efficiency of operations, improving the real-time monitoring for authorities, developing new maritime applications and services, such as automated navigation systems or navigation assistance systems.

A public data space has the potential to enhance the provision of services by the maritime sector when it operates as a public administration serving citizens. This improvement would result from the increased accessibility and efficiency that such a data space can offer. For instance, it could expedite the delivery of maritime-related information and services to citizens, including data on maritime traffic, weather conditions, and safety alerts. Furthermore, it could pave the way for the development of new maritime services tailored for citizens, such as navigation assistance and maritime transportation services. To gain a deeper understanding of the dynamics between citizens and public administrations, the Public Encounter model, initially introduced by Charles T. Goodsell in 1981, serves as a valuable framework (40). Recently, Lindgren et al. updated this model in 2019 to account for contemporary factors that influence interactions between citizens and public administrations, including the impact of technologies (41). These adaptations highlight the significance of user experiences and interactions when examining the maritime sector engagement with citizens in the digital age.

6. *Challenges*

Finally, the framework recognizes that developing a public data space for the maritime domain is a complex challenge. Some of the challenges might be technical, organizational, in terms of data governance, and last but not least economical. The establishment of European data spaces still faces technical, organizational, and economic challenges. Technical challenges involve specifying and supporting the data-sharing lifecycle with a common framework for data-sharing agreements, managing data ownership policies with semantic and behavioral interoperability, incorporating data provenance mechanisms for data sovereignty and trustworthiness, and defining a decentralized architecture agreed upon by all relevant stakeholders. Organizational challenges encompass supporting practices for trust management, including security, privacy, and assurance management, and adapting data spaces to the specific needs of discrete ecosystems, considering domain, sector, or territorial requirements. Economic challenges include finding the balance between ethical/societal concerns and economics, providing agile support of data monetization models for data spaces' success, and furnishing mechanisms for incentivizing data sharing and exchange, crucial for high-volume data sharing.

3.2 Technical foundations for secure and trusted data

The project goal is to outline a federated environment, ensuring the secure, efficient, and reliable exchange and dissemination of maritime data among endorsed participants in both maritime and terrestrial domains. This effort aligns with the architectural framework and recommendations advocated by the International Data Spaces Association (IDSA). The research endeavors to realize three primary goals: enabling transparent data access, establishing a secure, resilient, and efficient communication channel between maritime vessels and onshore entities, and streamlining the digital provision of dependable services.Therefore the two application cases (maritime sector as trade and public services) will be used to validate these three goals. Due to the frequent occurrence of data generated and possessed by entities distinct from those delivering services reliant on such data, there is an imperative to devise a novel framework for sharing data with trustworthiness. This framework necessitates an arrangement wherein data proprietors maintain direct authority over access privileges, irrespective of the data's physical location. Within the maritime industry, there is an absence of established conventions to fulfill this requirement. To address this void, we can consider the data sharing and exchange procedures outlined within the IDSA methodology, specifically concerning Industrie 4.0. The IDSA approach prioritizes the secure dissemination and sharing of trusted data, empowered by robust data ownership and control measures. It leverages contemporary cryptographic techniques to uphold ownership records of data objects, regardless of their storage locations. Data owners, identified via electronic signatures, can then securely grant access to third-party data consumers. As pointed out in the conclusion of the literature review, the IDSA Reference Architecture Model can be adopted as a blueprint for the implementation of data spaces, aligning with widely accepted system architecture models and standards (e.g., ISO 42010). Indeed, IDSA RAM (Figure 3.2) uses a five-level structure that expresses the concerns and views of various stakeholders at different levels of granularity.The model presents 5 layers dimension and 3 perspective that need to be implemented across all layers, the research focuses on the business and functional layers. In this context, the concept of data spaces for maritime can be defined as an ecosystem for secure and reliable data exchange based on the principles and reference architecture model promoted by IDSA.

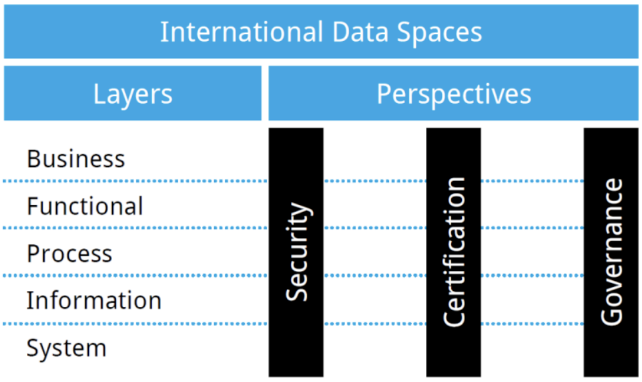


Figure 3.2:
General structure of the IDSA Reference Architecture Model

BUSINESS LAYER

The business layer serves as a delineation of the various roles individuals can assume within a data environment, outlining the key tasks and interactions affiliated with each role. This framework actively contributes to the formulation of inventive business models and digital, data-centric services, which are accessible to participants within the data environment. It presents a conceptual portrayal of these roles within the data ecosystem, essentially serving as a foundational model for the more technically-oriented layers. Consequently, the business layer outlines the prerequisites that the functional layer must satisfy.There are four main categories of roles: core participants, intermediaries, software and service providers, and governance body (already mentioned in section 2.3.1).

The core participants are actively engaged and mandated each time information is shared within the data environment. The data owner establishes policies governing data use and grants access to the data. Typically, a participant serving as a data owner simultaneously takes on the role of a data provider. The data provider ensures data availability for exchange between a data owner and a data consumer. Furthermore, the data provider communicates pertinent information regarding the data owner and their data offerings to a centralized intermediary service provider responsible for housing and presenting metadata collected from all available data providers within the data environment. The process involves the transfer of data from a data provider to a data consumer. From a business process modeling standpoint, the data consumer essentially mirrors the data provider, performing activities similar to those executed by the latter. To access available data offerings, the data consumer can interact with the user interface provided by the broker service provider, allowing them to search for and query data. The data consumer has the option to establish a direct connection with the data provider or connect through the intermediary services of the broker service provider. The data user is the authorized entity with legal permission to utilize the data owned by a data owner, in accordance with the usage policy. This role is analogous to the data owner, who holds legal control over their data. In most instances, the data user and data consumer are one and the same.

Intermediaries play a vital role in the realm of data, instilling trust and furnishing metadata and data catalog functionalities. They cultivate a business model ecosystem centered on the provision of data and services. These intermediaries encompass the identity provider, broker service provider, and clearing house. The identity provider's primary function is to deliver a service that involves creating, maintaining, managing, monitoring, and validating the identity information of participants within the data domain. It comprises a Certificate Authority for the administration of digital certificates and a Dynamic Attribute Provisioning Service tasked with handling dynamic attributes, such as permissions for accessing data offerings. Meanwhile, the broker service provider serves as a centralized entity responsible for storing and overseeing information about participants and their data offerings. It receives and disseminates metadata, conforming to the IDSA Information Model. It's crucial to emphasize that the broker service provider does not partake in the actual data exchange process. Its role concludes once it furnishes the data consumer with metadata pertaining to a specific data provider. Lastly, the clearing house serves as an intermediary offering clearing and settlement services for both financial and data exchange transactions. These clearing activities are distinct from the functions associated with maintaining a metadata repository. The clearing house diligently records all activities carried out during a data exchange. Following the completion of a data exchange, or segments thereof, both the data provider and data consumer validate the data transfer by recording transaction details at the clearing house. This recorded information serves as the basis for billing and conflict resolution.

The data space's service provider function entails offering software and services intended for utilization by participants. Additionally, this role may encompass hosting essential infrastructure for engagement within the data space or delivering supplementary data-related services to enhance the caliber and practicality of data sharing within the ecosystem.

The regulatory entities within the data environment encompass the Certification Body, Evaluation Facilities, and the International Data Spaces Association (IDSA). The Certification Body, along with the Evaluation Facilities, engage in the validation of participants and essential technical elements within the data domain. Their primary role involves guaranteeing access exclusively to organizations that have obtained certification and adhere to the requisite standards, thus maintaining the integrity of the trusted business network. Oversight of Evaluation Facilities' operations, actions, and determinations falls under the purview of the Certification Body. Meanwhile, the IDSA, operating as a not-for-profit organization, lends its support and guidance to the advancement of the Reference Architecture Model and the certification procedure for participants.

So, from the definitions of participants' roles, it is possible to assume generally a flow of information that follows the various interactions (Figure 3.3). In detail, the graph shows the participants colored by role (in blue the core participants, in orange the intermediaries, and in yellow the service provider) and the interactions colored by type of data shared (green when it is data flow and purple when it is metadata flow).

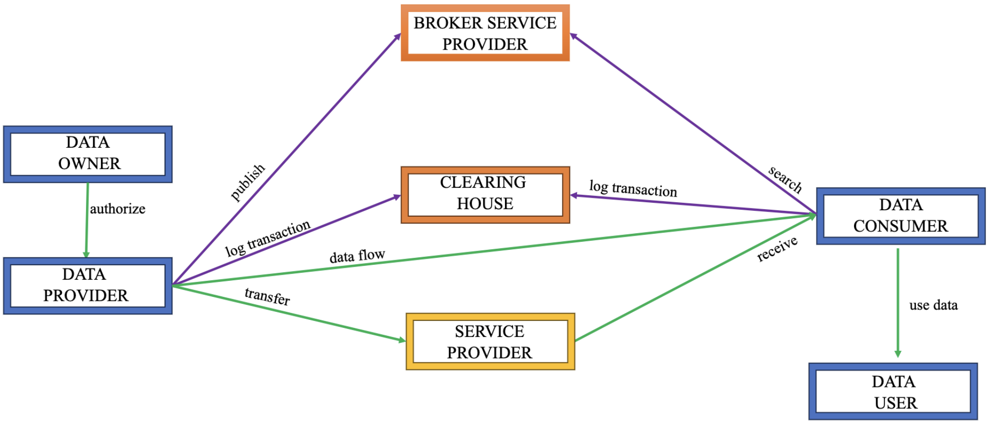


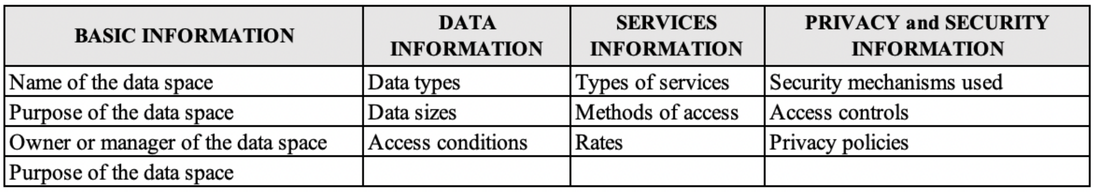
Figure 3.3:
Roles and interactions

Based on these definitions, Chapter 4.2.1 outlines the two maritime domain cases of interest and the corresponding roles and interactions.

FUNCTIONAL LAYER

The Functional Layer within the data domain outlines the necessary criteria, encompassing functionality and features for execution. This segment will explore the potential alignment between the maritime data environment and the functional prerequisites outlined by the IDSA, encompassing trust, security, data sovereignty, data ecosystem, standardized interoperability, value-added applications, and data marketplaces.

The *trust* requirement is achieved primarily by defining roles and responsibilities, with rules in terms of identity management (e.g., each participant must have a unique identifier and a valid certificate). A practical example in response to this need may be a mechanism similar to the Self-Description (SD) document developed in the Gaia-X Trust Framework describing the data and services offered by a data space. It is a fundamental document for the Gaia-X trust framework, as it describes Entities from the Gaia-X Conceptual Model in a machine interpretable format. This includes Self-Description for the participants themselves, as well as the resources and service offerings from the providers. This last aspect is also fundamental for the third requirement which is the creation of the *ecosystem of data*.I Collected in the table below (Figure 3.4) some examples of information that can be included in a Self-Description document:



BASIC INFORMATION	DATA INFORMATION	SERVICES INFORMATION	PRIVACY and SECURITY INFORMATION
Name of the data space	Data types	Types of services	Security mechanisms used
Purpose of the data space	Data sizes	Methods of access	Access controls
Owner or manager of the data space	Access conditions	Rates	Privacy policies
Purpose of the data space			

Figure 3.4:
The Self-Description information

The mechanism can be easily explained in 4 steps:
1. The owner or operator of a data space creates a Self-description document that describes the data and services offered by the data space.
2. The Self-description document is published in a format that conforms to the requirements of the Gaia-X trust framework.
3. The Self-description document is verified by a Gaia-X compliance service.The Gaia-X compliance service verifies the conformity of the Self-description document with the rules of the Gaia-X trust framework.If the Self-description document is compliant, a Gaia-X Certificate of Compliance is issued.
4. The Gaia-X certificate of conformity is used by users to verify the compatibility and security of the data space.

The *security and data sovereignty* requirement can be satisfied through authentication and authorization mechanisms, the use of policies, and a security by design approach. Even in this case the Gaia-x model try to match this point with trustworthy verification mechanisms that in combination with the Self-Description document empower participants in their decision-making processes. Self-description must conform to the set of rules of the Gaia-X trust framework. The rules are designed to ensure that data spaces are secure, private and interoperable. The Trustworthy framework is defined as the process of reviewing and validating the automatically applicable set of rules to achieve the minimum level of compatibility with Self-Description in terms of some main rules:
- correct syntactic insertion
- schema validation
- cryptographic signature validation
- consistency of attribute values
- verification of attribute values

Regarding the security aspect some clarifications need to be made. One of the primary obstacles to data sharing within companies frequently centers on the two most commonly mentioned apprehensions: ”security” and ”data usage control.” The control of data utilization delineates the guidelines governing the transparent sharing of data. Consequently, data becomes linked to clearly defined services established through mutual agreement between partners. Security represents a universal concern not exclusive to data sharing and can be mitigated through the correct implementation of identity and access management, along with the judicious application of data encryption techniques. It is evident that a current problem emerges concerning the influence that privacy and data security aspects have on the willingness of organisations to share data within public data spaces.

Recently, the concept of security by design (42) has received a major boost as an approach to IT security that incorporates security from the very beginning of the development process. This approach is based on the idea that security cannot be added as afterthought, but must be integrated at every stage of the development process, thus from the design phase. Software engineering discipline is about managing complexity that can be incidental or intrinsic. Regarding the intrinsic complexity, there are different methodologies to address the problems. The primary methodologies encompass top-down and iterative design. The initial instance of iterative design emerged with Barry Boehm’s spiral model, where the development progresses through a predetermined series of iterations, during which a prototype is constructed and assessed. Managers have the opportunity to assess risk at each phase, allowing them to make informed decisions about advancing to the subsequent iteration or terminating the process. This model derives its name from its portrayal as a spiral due to the cyclic nature of the procedure as shown in the Figure 3.5.

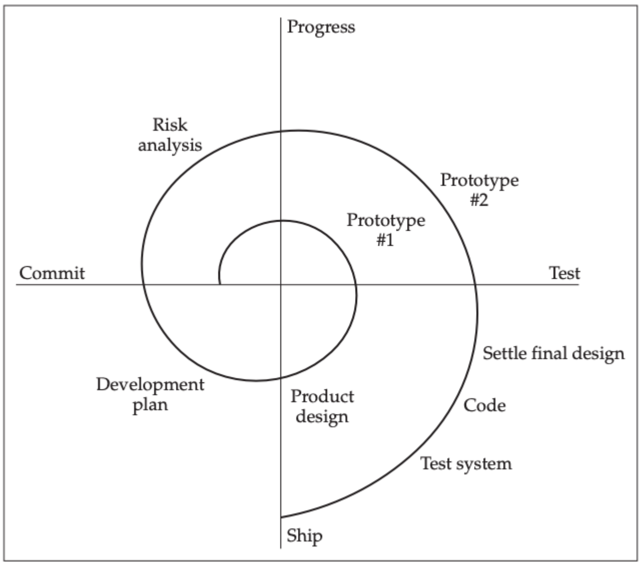


Figure 3.5:
The spiral model

The top-down approach is also called 'waterfall model' (in the Figure 3.6) formalised in the 1960s by Win Royce for the US Air Force.

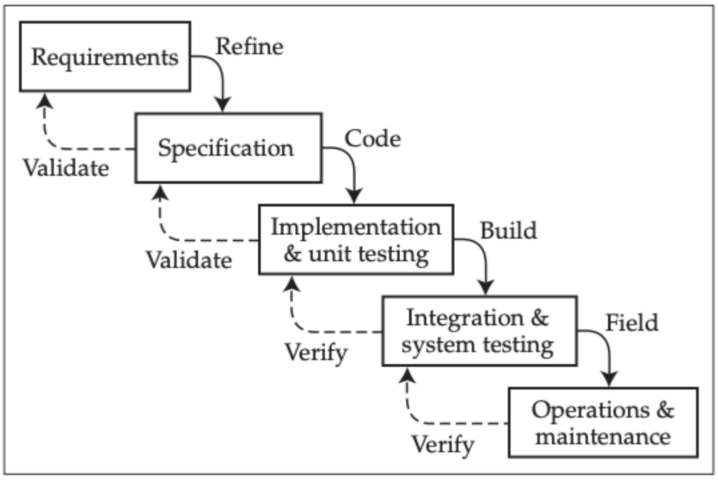


Figure 3.6:
The waterfall model

The approach involves commencing with a succinct outline of the system's demands, which is then expanded into a detailed specification. Following this, the system's components are constructed and tested individually, followed by their integration and comprehensive system testing. Subsequently, the system is prepared for live operation. During the initial two phases of this sequence, feedback is sought to validate whether the correct system is being developed, and during the subsequent two phases, verification is conducted to ensure that it is being developed correctly. The strengths of the waterfall model lie in its ability to establish a clear understanding of system objectives, architecture, and interfaces at an early stage. It simplifies project management by defining distinct milestones and can improve cost transparency by billing for each phase separately, even in cases of late specification modifications. Additionally, it is compatible with various tools, making it a favorable choice when conditions permit. A critical factor to consider is the level of foreknowledge regarding project details. The waterfall model is exceptionally well-suited for projects with well-defined requirements, such as compiler design or cryptographic processor implementation. An example of a framework following this approach is Microsoft's Security Development Lifecycle (SDL), which aims to reduce software vulnerabilities' frequency and severity by integrating security and privacy throughout all development phases. In the context of security, the ISO/IEC 27034 (2011) offers guidance on the specification, design, selection, and implementation of information security controls integrated throughout an organization's SDL.

In order to have establish a security by design approach in the maritime domain, there are some technical requirement and attention to be taken both from the point of view of software security and end-user security. So, here are the main steps:

- **Risk Analysis**

The first step for a security by design approach is to identify and assess the security risks associated with the system or application. This step allows security requirements to be identified early in the design process. This can include defining threat scenarios in order to identify potential vulnerabilities and attack scenarios. This helps prevent and mitigate risks from the outset.

- **Design**

The second step is to design the system or application in a way that mitigates the identified risks. This may include the use of secure design techniques such as encryption, authentication and authorisation. For instance, use encryption to protect sensitive data, or use the cryptographic protocols for secure communications, or implement robust authentication mechanisms to ensure that only authorised users can access the system. The multi-factor authentication can be used to determine which resources and functionalities can be accessed by each user. Attention to privacy protection is crucial. The user should have control over what data is collected, how it is used and with whom it is shared. Tools such as customisable privacy settings and clear notifications on data use can be included. Concerning this only strictly necessary data should be collected and it should be ensured that data are only kept for as long as necessary (as referred to by the GDPR as data minimization). At least, the interface design should comply with the criterion that security should not compromise usability and accessibility.

- **Secure implementation**

The third step is the system or application secure implementation. This means using secure tools and processes and thoroughly testing the system or application to identify and correct any security vulnerabilities. Conduct penetration tests and security tests to identify and correct vulnerabilities as part of the development process.

- **Other security countermeasures**

The attention can be focus on the input data validation to verify and validate input data to prevent attacks such as injection, cross-site scripting and other input-related vulnerabilities, or the logging and monitoring tools integration to detect suspicious activity and intrusions in real time, or to implement detect and response measures by defining incident response procedures and recovery plans, or the implementation of security controls, such as user training and software updates. Also end users should be educated on information security and recommended practices. And should provide users with an easy way to report security problems, bugs or suspicious behaviour.

Chapter 4

Methodology

This chapter discusses the methodology applied for this research and the motivation behind the decisions made, presenting first, the strategy and design of the research and then the methods chosen for data collection and analysis.

4.1 Research Strategy

There are two main research strategies that can be followed, qualitative and quantitative. Each one has its own unique characteristics. Quantitative research is more likely to produce objective findings because it typically involves a larger sample size, fewer potential biases introduced by the researcher, and a more standardized approach to data collection and analysis. Qualitative research, on the other hand, focuses on a more in-depth analysis of a smaller sample of cases/subjects in order to gain a more complete understanding of the phenomena being studied.

For this, a primarily qualitative approach has been preferred. The choice of this strategy is related to the nature of the topic of the thesis and the research question. More precisely, in analyzing which are the new implication of digitalization in the maritime domain and the main benefits and challenges associated to the introduction of data spaces in this field of research and the other related topics, it is important to ensure a wider understanding though in-depth data collection and analysis. For this reason, it is relevant to emphasize the importance of concepts rather than numbers as source of data, which is a characteristic of a qualitative research strategy. This not imply that, at least in the early part of the research such as in the literature section, no weight was given to quantitative data, rather some data were collected analyzed and graphically represented to demonstrate the importance of this study in the current context.

4.2 Research Design

The objective of the research is to apply the notion of public data space to the maritime domain. Even though the inductive method is most commonly used with a qualitative research, the approach followed is mainly deductive. Indeed, the research started from a general theory on public data spaces and then went on to explore a specific topic in order to give a contribution in a new field. For enabling this is required a qualitative method because because it guarantees flexibility and adaptability, which are fundamental elements to investigate a new field.

The deductive approach was the right choice for several reasons. At first, it allowed me to test the theory that public data spaces can improve the efficiency and safety of the maritime sector. Furthermore, it was chosen to be able to cover two specific topics related to public data spaces in the maritime sector: how public data spaces can be used to improve the exchange of real-time navigation data and how public data spaces can be used to simplify and automate processes related to services to citizens. At least, the approach combined with the qualitative method allowed me to discuss with some maritime stakeholders in relation to public data space through interviews.

In particular, I had the opportunity to study the transnational benchmark analysis conducted in the Luiss Business School project which included the Italian, Norwegian, Dutch, French and Spanish Coast Guards. From the results of this study, it seems coherent to me to consider not only the Italian Coast Guard, but also the French Coast Guards, as they showed interesting aspects to focus on.

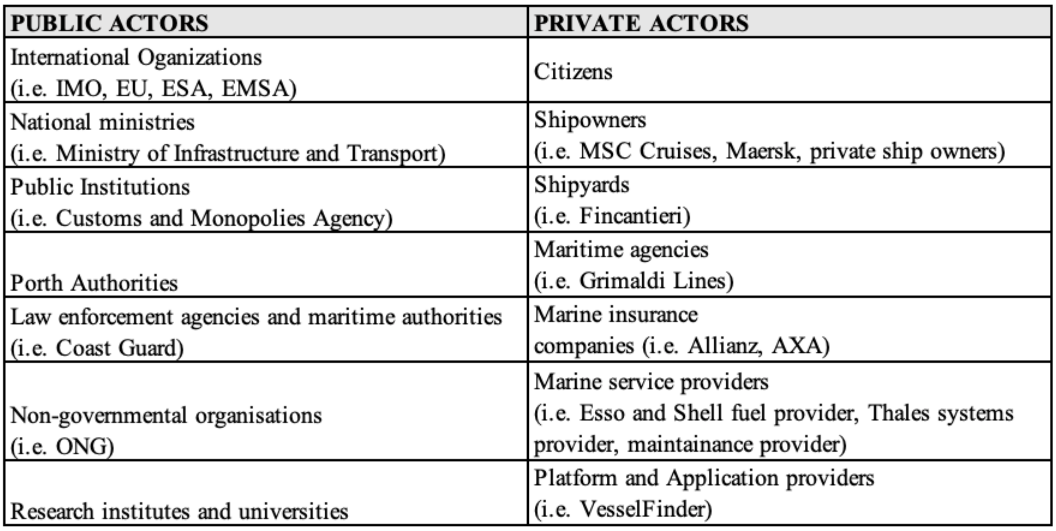
4.2.1 How to build a public data space for maritime domain

The first approach toward the maritime sector ecosystem was to identify with hypotheses what could be the main actors on the public and private side interested in creating a public data space for the maritime sector (summarized in table 4.1)

Furthermore, it is necessary to redefine the two cases of interest within the maritime domain, as being a very large domain it would be too complex to go and analyze all activities involving data exchange. In particular, as already defined the research will deal with:

Case 1 - Real-time maritime data

The maritime sector is a major player in the global economy, accounting for about 90% of the world's trade. However, the sector is facing increasing competition from other modes of transport, such as air and rail. These modes of transport are often faster and more efficient than maritime transport, and they are also benefiting from new investments in infrastructure.The maritime transport industry faces strong competitive pressure from other modes of transport, many of which benefit from entire new infrastructures (e.g. China's Belt and Road, autonomous transport systems and new airports such as those in Berlin, Beijing, Mexico City



PUBLIC ACTORS	PRIVATE ACTORS
International Oganizations (i.e. IMO, EU, ESA, EMSA)	Citizens
National ministries (i.e. Ministry of Infrastructure and Transport)	Shipowners (i.e. MSC Cruises, Maersk, private ship owners)
Public Institutions (i.e. Customs and Monopolies Agency)	Shipyards (i.e. Fincantieri)
Porth Authorities	Maritime agencies (i.e. Grimaldi Lines)
Law enforcement agencies and maritime authorities (i.e. Coast Guard)	Marine insurance companies (i.e. Allianz, AXA)
Non-governmental organisations (i.e. ONG)	Marine service providers (i.e. Esso and Shell fuel provider, Thales systems provider, maintainance provider)
Research institutes and universities	Platform and Application providers (i.e. VesselFinder)

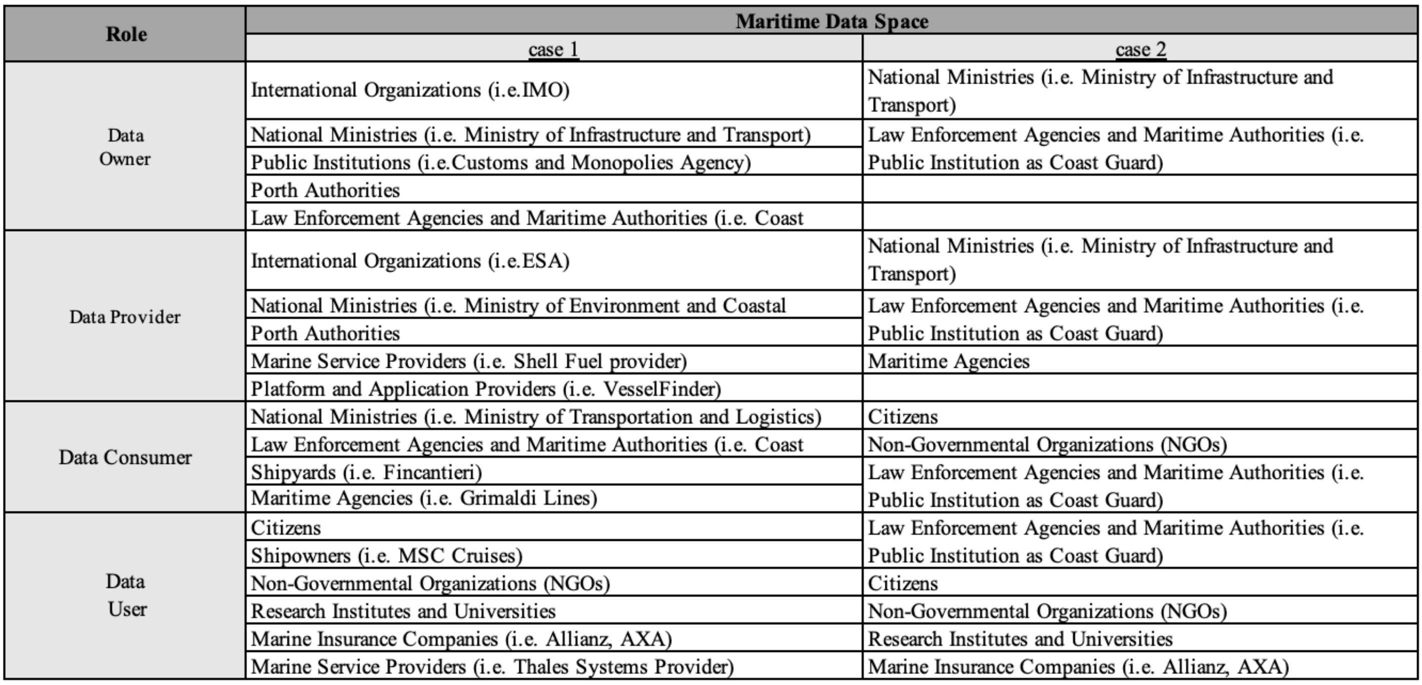
Figure 4.1:
Public and private players in maritime domain

and Sydney). To remain an attractive and contemporary transport option in an increasingly digital economy, the maritime ecosystem needs to create a new digital infrastructure that facilitates the provision of reliable and predictable data on the real-time transit. The availability of real-time data from different sources is fundamental for the trade sector but also for the navigation safety and accident prevention, even for maritime surveillance and national security.

Case 2 - Maritime sector as public administration

The maritime sector acts as a public administration when it has to provide services to citizens. It can be considered an important part of public administration as it includes a number of organizations and government agencies that play a key role in the management, regulation, and delivery of services related to maritime activities. These organizations carry out numerous activities that directly impact the lives of citizens in various ways, such as navigation and user information services, rescue and life-saving services, and management of navigational paperwork. The case of the Coast Guard is a good example of how public administrations are using digital technology to improve their services to the maritime sector.

After briefly summarizing the two cases of interest, I outlined the specific core participants (following the definition of IDSA roles in Chapter 3.2) and how they can be implemented in the two cases studies, keeping in mind that each actor can assume several roles (Figure 4.2).



Role	Maritime Data Space	
	case 1	case 2
Data Owner	International Organizations (i.e.IMO)	National Ministries (i.e. Ministry of Infrastructure and Transport)
	National Ministries (i.e. Ministry of Infrastructure and Transport)	Law Enforcement Agencies and Maritime Authorities (i.e. Public Institution as Coast Guard)
	Public Institutions (i.e.Customs and Monopolies Agency)	
	Porth Authorities	
	Law Enforcement Agencies and Maritime Authorities (i.e. Coast	
Data Provider	International Organizations (i.e.ESA)	National Ministries (i.e. Ministry of Infrastructure and Transport)
	National Ministries (i.e. Ministry of Environment and Coastal	Law Enforcement Agencies and Maritime Authorities (i.e. Public Institution as Coast Guard)
	Porth Authorities	
	Marine Service Providers (i.e. Shell Fuel provider)	Maritime Agencies
	Platform and Application Providers (i.e. VesselFinder)	
Data Consumer	National Ministries (i.e. Ministry of Transportation and Logistics)	Citizens
	Law Enforcement Agencies and Maritime Authorities (i.e. Coast	Non-Governmental Organizations (NGOs)
	Shipyards (i.e. Fincantieri)	Law Enforcement Agencies and Maritime Authorities (i.e. Public Institution as Coast Guard)
	Maritime Agencies (i.e. Grimaldi Lines)	
Data User	Citizens	Law Enforcement Agencies and Maritime Authorities (i.e. Public Institution as Coast Guard)
	Shipowners (i.e. MSC Cruises)	
	Non-Governmental Organizations (NGOs)	Citizens
	Research Institutes and Universities	Non-Governmental Organizations (NGOs)
	Marine Insurance Companies (i.e. Allianz, AXA)	Research Institutes and Universities
	Marine Service Providers (i.e. Thales Systems Provider)	Marine Insurance Companies (i.e. Allianz, AXA)

Figure 4.2: *IDSA roles for maritime domain*

Case 1 - Real-time maritime data

I identified the potential data owner in the International Maritime Organization (IMO), which owns and manages global data on safety of navigation and international rules for maritime traffic; National Ministries as the Ministry of Infrastructure and Transport because it owns data on port infrastructure and navigation in national territorial waters; Public Institutions such as the Customs and Monopolies Agency that may own data on the control of goods and vessels in national ports; Port Authorities which owns data on port operations, vessel traffic and port management; and Law Enforcement Agencies and Maritime Authorities such as the Coast Guard, who owns critical data on maritime safety, water surveillance and rescue operations at sea. Then, I categorized in data provider International Organizations such as ESA which provides satellite data, including satellite images, for global maritime surveillance; National Ministries for instance the Ministry of Environment and Coastal Protection that could provide environmental data and water quality information; Port Authorities that provides real-time data on port traffic, local weather conditions and other relevant information; Marine Service Providers such as the Shell Fuel Provider that provides data on fueling of ships and vessels in ports; and Platform and Application Providers like VesselFinder which provides an application that offers real-time data on the location of ships and marine traffic. Following, the data consumer could be identify in the National Ministries (i.e.Ministry of Transportation and Logistics) that use navigation data and port information for national maritime transportation planning and management; Law Enforcement Agencies and Maritime Authorities such as the Coast Guard which works with real-time navigation data to monitor and ensure the safety of territorial waters and shipping lanes; Shipyards (i.e. Fincantieri) that uses data related to scheduling ship deliveries and planning shipbuilding activities; and Maritime Agencies (i.e. Grimaldi Lines) that need data to monitor its fleet operations and optimize shipping routes. Finally, in this exchange of information, the data user can be citizens that use applications such as VesselFinder to track cruise or merchant ships, plan sea voyages, or simply monitor maritime traffic; shipowners who uses navigation data and maritime traffic information for cruise ship route planning and operations management; Non-Governmental Organizations (NGOs) because for instance an environmental organization might use maritime data to monitor marine pollution or illegal fishing activities; research institutes and universities to conduct scientific research; Marine Insurance Companies that uses navigation data and maritime safety information to assess risks and set insurance rates; and marine Service Providers that works with maritime data to provide advanced safety and communication solutions to ships and offshore platforms.

Case 2 - Maritime sector as public administration

I identified as potential data owners National Ministries such as the Ministry of Transportation and Infrastructure because it is the owner of data regarding port infrastructure management, maritime development projects, and maritime safety plans; and public institutions as Coast Guard that could be considered a data owner of data related to sea rescue operations, ship inspections, citizens' maritime documentation and maritime control activities. I then identified as a data provider again the Ministry of Transportation and Infrastructure because it provides data on the status of port infrastructure, shipping routes, and maritime traffic information to citizens; again the Coast Guard which provides real-time data on maritime safety, weather conditions and rescue operations to ships and citizens at sea; and Maritime agencies such as the Local harbor master's office that provides data on vessel registration and license requirements. Further, I have listed as data consumer the citizens that mainly use digital services provided by the Coast Guard or other maritime organizations to obtain information on water safety, marine weather forecasts, or emergency rescue services; Non-Governmental Organizations (NGOs) that might consume data on the location of ships at sea to participate in rescue operations or to monitor water safety; and Law Enforcement Agencies and Maritime Authorities that need data on nautical licenses for the enforcement of maritime laws and to ensure that only those who are eligible can operate vessels. Last but not least, I suppose that in this context the data user can be the citizens themselves who use the data to apply for boat license renewals, verify eligibility requirements, and monitor expiration dates but also who use online services provided by the maritime government to obtain real-time information on navigation, weather conditions, and rescue services; Non-Governmental Organizations (NGOs) that might use data on the validity of boating licenses to promote boater safety and training or they use data provided by the Coast Guard or public institutions to participate in environmental monitoring activities and assistance operations at sea; research institutes and universities; Marine Insurance Companies that could use data on nautical licenses as part of insurance risk assessments for boat owners; and Law Enforcement Agencies and Maritime Authorities as the Coast Guard which is involved in the marine license renewal process use data to provide efficient services to citizens.

4.2.2 Comparing Italian and French Coast Guards

For the empirical context, I had the opportunity to study the transnational benchmark analysis conducted in the Luiss Business School project, and I decided to focus on the mandate, activities, and technologies of the Italian and French Coast Guards.

Italian Coast Guard

Institutional set-up: Navy Corps
Geographical Area: Mediterranean

The functions of the Italian Coast Guard are very broad and are best represented by its various institutional referents, for instance:

- Ministry of Infrastructure and Sustainable Mobility (i.e. maritime and navigation safety through S.A.R. operations and the Global maritime distress safety system (GMDSS))
- Ministry of Ecological Transition (i.e. Surveillance, protection and prevention of the marine and coastal environment)
- Ministry of Defence (in peacetime i.e. contribution to State maritime and coastal defence and support for the State Cartographic Institute IIMM)
- Ministry of Culture (i.e. protection of historical, artistic and archaeological heritage)
- Ministry of the Interior (i.e. combating drug trafficking and illegal immigration)
- Civil Protection Department

As part of its services to shipping, the Italian Coast Guard carries out several activities: Search And Rescue activities (S.A.R); safety of navigation and maritime security; Port State control; boating administrative functions; and maritime mercantile administration.

It can be concluded that the Italian Coast Guard has to fulfil a lot of functions, but what are the main supporting technologies?

- ARES (Automated Search and Rescue) is naval reporting system designed to provide up-to-date information on vessel movements.

• LRIT (Long Range Identification and Tracking) is a satellite-based vessel traffic monitoring system capable of automatically transmitting certain information such as identity, location (latitude and longitude), date and time. The VTMIS platform is used by the Coast Guard integrates LRIT information with available AIS information so that ships of interest can be monitored.

• AIS (Automatic Identification System). The 'national network' for the reception of AIS (Automatic Identification System) information transmitted by ships consists of 63 base stations installed to ensure complete coverage of the national coastal profile up to the entire Italian Search and Rescue (SAR) area. The information acquired is centralised at the General Command and made available by it to other services under the General Command's responsibility and to other State Administrations. The national AIS network not only receives the information transmitted by ships, but it also transmits messages concerning the main Aids to Navigation - AtoN; messages concerning navigation safety; basic station report messages through which ships can synchronise their AIS transmissions; in certain geographical areas, in repeater mode, the AIS information received in that area.

• VTS (Vessel Traffic Service) is a radar system with the purpose of increasing the safety and efficiency of maritime traffic, facilitating the intervention of the Authorities in the event of an accident or in the presence of potentially dangerous situations at sea, including search and rescue operations, and providing an aid to improve the prevention and detection of pollution caused by ships. The service has the ability to interact with traffic and respond to its evolution in VTS areas. VTSs can be of two types: Port VTS and Coastal VTS. The 11 VTS Centres are organised and equipped to provide INS, NAS and TOS. The Information Service (INS) is provided by radio via the VHF channels reserved for the VTS Centre, transmitting at regular intervals or when deemed necessary, information regarding traffic, weather or navigability conditions. The Navigation Assistance Service (NAS) assists on-board decision-making and to monitor its effects. Using the NAS, the captain receives navigation suggestions. The Traffic organisation service (TOS) concerns the operational management of traffic and the preventive planning of vessel movements to avoid congestion and the occurrence of dangerous situations.

• VMS (Vessel Monitoring System) is a satellite monitoring system adopted in compliance with Regulation No. 1224 of 20.11.2009 and Regulation No. 404 of 8.04.2011 aimed at controlling the exploitation of fishery resources.

• SafeSeaNet (SSN) is a monitoring and information system on maritime traffic set up in compliance with Directive 2002/59/EC and subsequent amendments and additions, with the aim of increasing the safety of navigation and port safety, the protection of the marine environment and the efficiency of maritime traffic and transport.

• CleanSeaNet (CSN) is a European service adopted in compliance with Directive 2005/35/EC of the European Parliament and of the Council of 7 September 2005 on ship-source pollution, i.e. the detection of marine oil pollution through satellite monitoring provided to coastal states.

• SEG (SafeSeaNet Ecosystem GUI) is a platform through which data from EMSA applications (AIS, LRIT, SAT-AIS, Fishing) can be combined and processed. It is used in conjunction with the Pelagus system and in the absence of the latter becomes the reference system for tracking vessels and covering much of Europe. It is operated by EMSA (European Maritime Safety Agency).

• NAVTEX (Navigational Text Warning) is a fully automated, direct-printing service for the real-time transmission of notices to mariners, weather information and other urgent warnings to vessels at sea. It is an integral part of the Global Maritime Distress Safety System (GMDSS), is a low-cost, simple and automatic service for broadcasting urgent warnings and weather reports to mariners, developed by the International Maritime Organisation (IMO) as part of the 1988 amendments to the International Convention for the Safety of Life at Sea (SOLAS). This involves the dissemination of text messages, also known as Maritime Safety Information (MSI), which are broadcast by coastal stations on dedicated frequencies and received by mariners by means of special equipment that also allows them to be displayed and/or printed.

• National maritime Single Window - PMIS system Pursuant to Law 221/2012 art.8 paragraphs 10 to 17 implementing Directive 2010/65/EU, the PMIS (Port Management Information System) system represents the single national interface for sending the declaration formalities for ships arriving in and departing from Italian ports (National Maritime Single Window). The single interface constituted by the PMIS must ensure interoperability with the Safe Sea Net system, the Customs information system, and with the platforms set up by the port authorities for the better performance of the functions of direction and coordination of the logistical nodes that report to them.

French Coast Guard - Gendarmerie Maritime

Institutional set-up: Gendarmerie Structure
Geographical Area: Ocean and Mediterranean

The Gendarmerie Maritime (Gendarmerie Maritime - GMAR), deals with tasks related to maritime security, control of territorial waters, monitoring of maritime traffic, prevention of illegal activities at sea and rescue at sea. In particular, the institutional mandate of the Maritime Gendarmerie in France is defined in the Defence Code and the Transport Code. The institutional mandate includes several aspects and responsibilities in the maritime context, which can be identified in 3 macro-areas: maritime security; criminal investigations; and coastal infrastructure protection.

The Gendarmerie Maritime carries out these tasks in close cooperation with other maritime agencies and authorities, including the National Navy, maritime administration and port authorities, to ensure the safety, security and efficient management of French coastal waters. The Gendarmerie Maritime carries out these tasks in accordance with French and international laws and regulations, with the aim of ensuring the safety, order and protection of French territorial waters and EEZ.The level of infrastructural integration of the Maritime Gendarmerie in France can be described as unitary governance rather than separate silos. The Maritime Gendarmerie operates as a specialised division within the national Gendarmerie, with a specific mandate for maritime security. This means that, despite its focus on maritime issues, the Maritime Gendarmerie is part of a single organisational structure with a unified chain of command and integrated governance. This unitary governance enables the Maritime Gendarmerie to collaborate and coordinate with other units and divisions of the National Gendarmerie, as well as with other agencies and authorities involved in maritime security, such as the National Navy and port authorities. This promotes effective resource management, information sharing and synergy between the different entities involved in maritime security. This unified infrastructural integration helps to ensure a coherent and coordinated approach in the management of maritime police and security activities, enabling the Maritime Gendarmerie to carry out its mandate efficiently and effectively.

Activities include maritime security and general police duties in territorial waters and the EEZ (exclusive economic zone under the United Nations Convention on the Law of the Sea, UNCLOS). Specifically, the activities are divided into:

- the control of territorial waters
- the safeguard of human life at sea
- the maritime security
- the maritime traffic control
- the protection of marine resources
- the activities related to the Department for criminal investigations (i.e. investigation of maritime criminal activities, gathering of evidence, co-operation with the public prosecutor, legal aid, protection of coastal naval establishments, security and surveillance, access control, investigating and countering threats, maritime security and counter-terrorism)

As for the Italian case, the Maritime Gendarmerie carries out a breadth of the most varied activities, but what technologies is it currently supported by to fulfill its functions?

The Maritime Gendarmerie uses various digital technologies and services to support its maritime security and territorial waters control activities but also its services to citizens, such as:

• digital communication systems, such as VHF/UHF radio networks and satellite communications, to enable fast and secure communication between operational units and coordination centers. These systems facilitate the exchange of crucial information during maritime operations

• digital platforms for information management and data sharing between operational units and coordination centers. These platforms enable better collaboration and more efficient management of operations. One example is SCADA (Supervisory Control and Data Acquisition), a large-scale remote management system used to process large numbers of telemetry readings in real time and to remotely control technical installations. SCADA systems are not specific to the maritime sector and are used in many critical infrastructures

• digital monitoring and surveillance tools, such as high-definition cameras and advanced sensors, to gather information, detect suspicious behavior, and identify dangerous situations. These tools help improve situational awareness and response capability. One example is the Electronic Chart Display and Information System (ECDIS), a system for displaying and managing chart information and navigation data used in the maritime domain. This system provides shipboard operators with an electronic representation of nautical charts and information necessary for the safe navigation of vessels. ECDIS integrates data from a variety of sources, such as electronic charts, Global Positioning System (GPS) data, and other navigational information, enabling operators to accurately display the vessel's position, maritime obstacles, recommended routes, navigation signals, weather forecasts, and other relevant navigational information

• Brigade numérique, new unit of the French National Gendarmerie that from 2018 is active 24 hours a day, seven days a week to request information, solicit action or send a report. The French initiative is among the first of its kind in Europe, following police in London and the Netherlands. The goal is to dispose in less than 24 hours of requests received worldwide

• Interreg Italy-France Maritime (2021-2027). It is a cross-border Program cofinanced by the European Regional Development Fund (ERDF), under the European Territorial Cooperation (ETC) objective of the EU Cohesion Policy 2021 - 2027. The main objective of the program is to help strengthen cross-border cooperation between participating regions and make the cooperation area a competitive and sustainable one in the European and Mediterranean landscape

• Project HAROPA. It is a collaboration between the ports of Le Havre, Rouen and Paris, located in the Seine and Yonne region of France. The main objective of the HAROPA project is to consolidate and develop port and logistics activity in this region, creating an integrated and internationally competitive port platform. Some key objectives of the HAROPA project include the development of intermodality; the modernization of port infrastructure:; the environmental sustainability; the development of industrial and logistics clusters; the digitization and innovation.

4.3 Research Method

Based on the selected research design, the research method chosen for this work concerns a first section dedicated to the literature review and theory development. In particular, since it is a very sector-specific and specific application, I decided first to outline the status quo in terms of digitization in the maritime sector. I started from the introduction of AIS system as the first major innovation to explain its evolution and today's challenges. For this first part I have the opportunity to attend a seminar on satellite technologies and security 'space as the new cyber frontier' seminar held by experts of the sector. A key concept emerges from the analysis of the innovative trends in the maritime industry: the strategic role of data in the maritime ecosystem. Hence, the research problem is to identify the specific digital technologies that permits to address the data fragmentation problem ensuring the reliability, integrity, and confidentiality of these data.

Therefore, in the literature review chapter I find out that while previous research on digital transformation has mainly focused on digital transformation within organisations, it is increasingly important to understand the new trend towards greater interconnection between different stakeholders to co-create services based on the provision and use of data.In particular, the literature review had a threefold purpose: to summarize the main findings of the research field in the existing literature, to validate the relevance of the research objective and research question, and to provide a theoretical background to be used in the interpretation and analysis of the empirical study. The answer founded was not fulfilled related to the maritime domain, so from this chapter I develop my research question to understand the technical requirements and the benefits and challenges associated with introducing data spaces into this domain. The research question identify is:

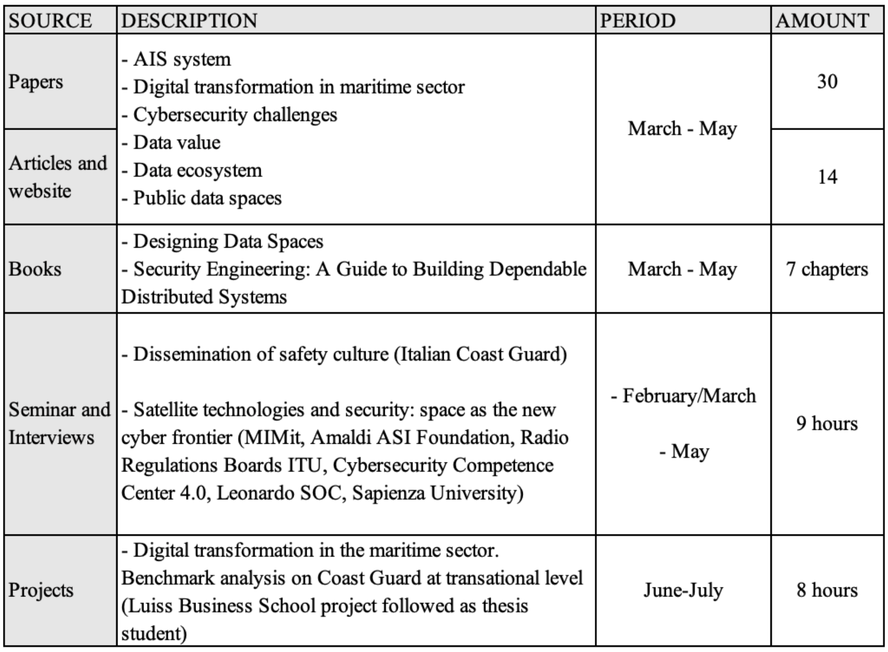
How to build a public data space for the digital transformation of maritime activites? What technical foundations are required to support secure and trusted data and application sharing in the maritime sector?

Thus, I decided to try to give a theoretical contribution in this sense. In particular, the material relevant for the literature review section and the theoretical development was searched in two main books:
- Designing Data Spaces: The Ecosystem Approach to Competitive Advantage by B. Otto, M. t. Hompel and S. Wrobel. (39)

- Security Engineering: A Guide to Building Dependable Distributed Systems by R. Anderson (42)

The second part of the method consists on studying an empirical context for the findings analysis. Hence, in order to demonstrate the proposed solution's credibility I identified the Coast Guard case study. In this regard, the Coast Guard was chosen as a case study to provide an end-user perspective on the research question in order to make this research credible, transferable, dependable, and confirmable.The case was chosen mainly as a result of a Luiss Business School project that involved interactions with consultants in contact with the Coast Guards (that I followed as a thesis student) and an interview I was able to obtain during a seminar held by the Italian Coast Guard at the Luiss University.

The main data sources are summarized in the table below (Figure 4.3).



SOURCE	DESCRIPTION	PERIOD	AMOUNT
Papers	- AIS system - Digital transformation in maritime sector - Cybersecurity challenges - Data value - Data ecosystem - Public data spaces	March - May	30
Articles and website			14
Books	- Designing Data Spaces - Security Engineering: A Guide to Building Dependable Distributed Systems	March - May	7 chapters
Seminar and Interviews	- Dissemination of safety culture (Italian Coast Guard) - Satellite technologies and security: space as the new cyber frontier (MIMit, Amaldi ASI Foundation, Radio Regulations Boards ITU, Cybersecurity Competence Center 4.0, Leonardo SOC, Sapienza University)	- February/March - May	9 hours
Projects	- Digital transformation in the maritime sector. Benchmark analysis on Coast Guard at transational level (Luiss Business School project followed as thesis student)	June-July	8 hours

Figure 4.3:
Data sources

Chapter 5

Empirical Findings

This chapter consists of the empirical side of the research. In particular, to accomplish this goal, I chose the Coast Guard as a case study. The first section focuses on the overview of the Coast Guard context, and then moves forward with the analysis of the findings with respect to the theoretical development chapter.

5.1 The Coast Guard case overview

The maritime industry is currently undergoing a rapid digital revolution that has a major impact on the industry's operations and sustainability, including advances in automation, data analytics, autonomous ships, and the Internet of Things. However, the industry faces several risks, including cybersecurity and integration of existing systems. To further explore the topics covered so far, the Coast Guard was chosen as a case study to provide an end-user perspective on the research question and the theoretical contribution.

Like other components of Italian and European law enforcement agencies, the Body of the Harbour offices, the Coast Guard, needs to adapt its services by leveraging the digital transformation taking place in the maritime domain. The Coast Guard carries out tasks related to civil uses of the sea and is framed within the Ministry of Infrastructure and Transportation, as well as working on behalf of various ministries, including the Ministry of Environment and the Ministry of Agriculture, Food Sovereignty and Forestry. These competencies include, first and foremost, the safeguarding of human life at sea, the safety of navigation and maritime transport, as well as the protection of the marine environment, its ecosystems, and the supervision of the entire maritime fishing chain, from the protection of resources to that of the end consumer. These are complemented by inspections of domestic merchant, fishing and recreational vessels, also conducted on foreign merchant vessels calling at domestic ports.

To address digital transformation from a long-term perspective, the Coast Guard must look at processes and technologies outside the perimeter of the individual organization, but it can and should work in synergy with the public and private sectors that make up the ecosystem of digital transformation in the maritime sector. As a result, the Italian Coast Guard must activate itself to the adoption of new technologies with a proactive and predictive technology adoption strategy, thus avoiding the risk of finding itself in the near future with obsolete technologies or chasing trends that could quickly run out of steam.

The main element that emerges, both from trends in technological development and from the behavior of other entities (such as Public Administrations), is the need for greater integration among all Entities and Authorities operating in the maritime sector and the relationship between these Institutions and Citizen-Users that is increasingly proactive, simple, integrated, and efficient in terms of costs and time, taking advantage of the new digitisation, communication and automation technologies for this. To best act on this technological lever, however, it is essential to thoroughly analyze how these synergies and relationships develop. Their identification is then followed by a study activity to highlight possible technological upgrades through:

- analysing good practices already in use in the European context,

- highlighting emerging technologies that would most improve the Institution-Institution and Institution-Citizens synergies in the Italian case

- outlining possible implementation paths of these technologies in the medium and long term

The objective of the study is to identify models for the Coast Guard's use of technology in order to improve, evolve, and make efficient the activity of the Italian Coast Guard, thanks to the upgrade and enhancement of the digital component. In particular, this research analyzes the two aspects previously presented for the Italian Coast Guard: how the real-time exchange of data for navigation takes place and what is the technology that supports the services towards citizens, and compares them with other European models, including mainly the French Coast Guard which is the most significant one, in order to outline a set of guidelines from which the Coast Guard, and why not other players of the maritime ecosystem, can be inspired.

5.2 From theory to exploratory case

With reference to the object of interest, the digital transformation of the maritime sector, for the Coast Guard, strategic planning passes through a reflection regarding the area of activity that is proposed to be changed. Thus, the design of a coherent digital strategy passes through the comparison with the practices of other state-of-the-art agencies, the identification of a future model of service delivery that intersects with the forecasts of technological development and its possible effect on the creation of new types of services. The definition of the new model of public value creation thus makes it possible to define the plan of actions to be implemented to reorganize the Entity, its tools and its organization.

Therefore, the starting point of the analysis is the concept of strategic alignment, which has been an essential reference in IT governance for the past 30 years (43). The Strategic Alignment Model (SAM) can be adapted to the context on the assumption that the institutions' strategy is defined by the institutional mandate and therefore the optimal fit must be sought between activities and technology. Otherwise, in the last 30 years, technology has gone from being seen as the optimal fit between mandate and activity to a new view, influenced by the generalised imposition of digital technology, whereby IT instead of supporting existing activities digital technology becomes a transformative activity. We refer to the concept of 'digital transformation' (44) to distinguish it from 'IT-enabled organisational transformation' because:

- digital transformation leverages digital technology to (re)define an organisation's value proposition, whereas IT transformation leverages technology to support the existing value proposition

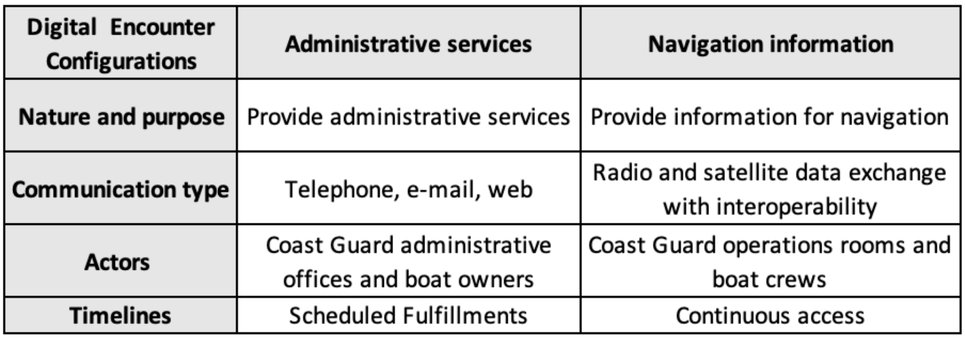
- digital transformation brings out a new organisational identity, while IT transformation enhances the existing organisational identity.

A redefinition of the organisational identity of the Coast Guard will not be hypothesised here as the objective is to improve, evolve and make efficient the experience of existing services. Therefore, the focus is on the Entity-Entity and Entity-User experience, with a view to adopting the most suitable technology to support the activities and consequently strengthen the organisational identity. The two theories highlight different aspects of the organisation-technology relationship. The strategic alignment theory provides the dimensions for case descriptions. The digital transformation theory, on the other hand, emphasises the structuring power of technology, which is able to shape or strengthen the institutional mandate as a result of business transformation.

5.3 Emerging findings on the Italian Coast Guard

Coastguard organisations are structured very differently according to geographical contexts, institutional mandates and contingent situations that impose different priorities for action. Historical events and different locations in the international context also contribute to changes in personnel composition, deployed means and operational doctrines. Technology and the need/possibility of ever greater integration of data flow and coordinated interventions in extremely difficult situations, as is typical in the maritime environment, contribute to making these particular public administrations complex and varied. Organisations with different institutional mandates and operating in different geographical contexts were selected for a comparison of the impact of digitisation in relation to the technology used for real-time monitoring of vessels and to the relationship between the Coast Guard and users/citizens.

From the analysis of the identified cases, two configurations of digital encounters between institutions-institutions and institutions-citizens emerge in the context of services currently offered by Coast Guard organizations in Europe to citizens/users (Figure 5.1).



Digital Encounter Configurations	Administrative services	Navigation information
Nature and purpose	Provide administrative services	Provide information for navigation
Communication type	Telephone, e-mail, web	Radio and satellite data exchange with interoperability
Actors	Coast Guard administrative offices and boat owners	Coast Guard operations rooms and boat crews
Timelines	Scheduled Fulfillments	Continuous access

Figure 5.1:
Digital Encounter Configurations

Considering the two Digital Encounter Configurations and the technological developments that characterize the maritime sector, the analysis highlighted the following innovative experiences of the Italian Coast Guard.

Case 1 - Real-time maritime data

The purpose of this section is to highlight the current configuration of information of the Coast Guard regarding the first case previously identified in the theory chapter, namely real-time maritime data. In detail from the two cases presented on the Coast Guard (Italian and French) it appears that currently, activities related to this type of information are primarily based on radio systems. But the analysis shows a trend toward interoperable real-time data exchange systems. As evidence of this, the Italian Coast Guard certainly shows the most propensity. In fact, in addition to the aforementioned technologies, it demonstrates close cooperation in data exchange with ESA through various platforms. A concrete example is the Pelagus platform that they use daily to monitor marine traffic. This is integrated with other data, defined as 'data fusion' by the Italian Coast Guard, besides AIS such as radar data, meteorological/environmental data, and search and rescue operations. Through this platform, the Italian Coast Guard receives alerts even if the signal is off, and as already mentioned this is one of the major limitations of the ais system. In particular, the Pelagus platform is also integrated with the SafeSeaNet (SSN) system of the European Maritime Safety Agency (EMSA). Integration with the SSN system allows the Pelagus platform to access AIS data from ships of all types, including small ships that are not required to have an AIS system installed. This expands the geographic coverage of the Pelagus platform and allows more ships to be monitored. A final key point of this platform is the ability to perform risk analysis tasks on historical data The Pelagus platform is an evolving system. The Coast Guard is working to improve integration with other data and to expand the platform's functionality such as interactions with ASI and ESA for information sharing through imagery, (of course this information is only related to civilian use, however, for military use it is taken care of by NATO). The Italian Coast Guard is constantly looking for ways to improve its use of real-time maritime data. The Coast Guard is working to improve the integration of different data sources, and to develop new platforms and applications that can be used to share and collaborate on this data.This willingness to share and collaborate has been evident especially in the relationship between the French and Italian Coast Guards, who have a long history of cooperation in the field of maritime safety. The purpose of the Italian and French Coast Guards to share and collaborate on real-time maritime data is a model for other Coast Guards around the world. By sharing data and working together, Coast Guards can improve their ability to prevent maritime accidents by ensuring safety and security.

Case 2 - Maritime sector as public administration

Today, more than ever, the Public Administration must work constantly to improve the services provided to citizens and businesses, showing a level of responsiveness comparable to that of the private sector. To achieve this and avoid chasing the resolution of contingent situations, PA must shift from a reactive to a proactive approach, which is essentially expressed in two dimensions: strategic planning and synergistic management.Currently, the administrative services offered by the Coast Guard are scheduled fulfillments delivered online and supported by traditional forms of communication. The analysis highlights the need to make access to services smoother and easier.

A good practice on this front has been outlined in the French Coast Guard's Brigade Numerique service already explained. The Brigade Numerique has helped to significantly improve accessibility to French Coast Guard services. In 2021, the service received more than 1 million contacts, a 20% increase over 2020. The service has also helped reduce Coast Guard response times, with an average of 24 hours for a first response. The Italian Coast Guard could implement a similar service to improve accessibility to their services. A 24/7 online response service would allow citizens and businesses to contact the Coast Guard at any time and from anywhere. The service could also offer a number of features that simplify access to services, such as the ability to send messages, upload documents, and initiate administrative procedures. A 24/7 online response service would improve the Coast Guard's efficiency and its ability to respond to the needs of citizens and businesses. The service would also help improve Coast Guard transparency and accountability by allowing citizens and businesses to monitor service activities.

Chapter 6

Discussion and Conclusions

The purpose of the chapter is to analyze the empirical results in light of the theory that inspired the research. In this chapter, the empirical findings are compared with theoretical concepts, highlighting correspondences and differences. The purpose of this final chapter of the research is to understand what emerges from the empirical context with respect to the theory to point out its contribution.

The analysis allows the potential impact of the technologies being deployed in the maritime sector to be declined in the Coast Guard context. It was seen how geographic and territorial distribution greatly influences this Entity, but each of the two reported cases showed leadership in one aspect in particular: the Maritime Gendarmerie in terms of digital user services, and the Italian Coast Guard in terms of interoperability of navigation data. The research has shown that Coast Guards at the transnational level do an exemplary job in terms of effectiveness, but there is certainly room for improvement in terms of efficiency that can be achieved through the potential of digital technology. For each configurations of digital encounters it is possible to identify the most appropriate technologies for that service to become more efficient.

The digital encounter in Coast Guard administrative services can be improved through the use of innovative technologies, such as blockchain. In this context, it can be used to improve document management, providing greater security against fraud and counterfeiting. In addition, blockchain can be used to ensure the traceability of supply chain and financial transactions, providing greater transparency and security. Furthermore, it would allow the creation of databases so that larger volumes of vessel information could be managed while ensuring the privacy of private operators. The latter would ensure the creation of a data ecosystem as insurance companies would also have a way to access and share data with the Coast Guard and other maritime authorities more efficiently. Another technology that can be used to improve the digital encounter in Coast Guard administrative services is generative AI. Generative AI can be used to create chatbots that can provide information and assistance to citizens and businesses. Chatbots can be used to automate some tasks, freeing up Coast Guard employees to focus on more complex tasks.

Technological developments in generative AI and blockchain suggest deepening experiences of other PA entities. The Coast Guard can learn from the examples of other PA realities that have already implemented these technologies. For example, the Italian PA has already begun experimenting with the use of blockchain for document and payment management.

The digital encounter in Coast Guard navigation information can be improved through the use of innovative analysis technologies for Big Data. This type of data can be used to improve the sharing and use of data among the various maritime stakeholders, including maritime authorities, shipping companies, ports, and businesses operating at sea. Furthermore, the use of Artificial intelligence (AI) and machine learning (ML) in this context and application could bring several benefits in the maritime sector. In particular AI can be used to develop algorithms that can be used to analyze data more efficiently and accurately. While the ML can be used to develop models that can predict the behavior of maritime traffic and risks to navigation. The implementation of these technologies would allow for better understanding of maritime trade traffic, improve risk prediction, improve route planning and thus management of resources such as fuel and personnel, and allow for the refinement of anomalies detection techniques.

From the research question, the literature has shown how in today's context digital transformation is evolving from private digital platforms to the emergence of public data spaces, and in particular we have demonstrated through the use of the Gaia-x case the possibility of applying this theory to the maritime sector as well. From what emerged from the empirical context, analyzing the case of the Coast Guard, it is evident that we are moving in the right direction, also because all the technologies highlighted so far can be implemented within a public data space. For instance, in the Gaia-x use cases there are many example of other sectors that could be applied to the need for the maritime sector. Considering the real-time monitoring case there is an example about earth observations that perfectly fits with the maritime case. The benefits of this system include:

- Federated access. This means that data users would not need to download data from multiple sources, but could instead access it from a single, centralized location. This would save time and bandwidth.
- Stable and comprehensive access. The system would provide stable and comprehensive access to maritime data, even if the number of sources or the size of the individual repositories increases over time.
- Efficient access. The system would be designed to efficiently deliver and process data, even for large datasets.

- Simple access. The system would be easy to use, even for non-technical users.
- Reduced carbon footprint: The system would help to reduce the carbon footprint of maritime data processing by deduplicating data and retaining it instead of regenerating it if necessary.
- Open and transparent digital ecosystem. The system would create an open and transparent digital ecosystem for maritime data, where data and services could be made available, collected, and shared in a trustworthy environment.

Considering the second case, which regards the services for citizens, the example of the medical records framework in the healthcare sector perfectly fits. The database could be used to store and manage all of the documentation related to a ship, such as the ship's registration, the crew's qualifications, and the ship's safety certificates. This would make it easier for authorities to access this information and to ensure that ships are complying with regulations. Moreover, the database could be used to provide information to citizens, such as the location of ships, the weather conditions, and the latest safety regulations. This would help citizens to make informed decisions about their travel plans. Additionally, the database could be used to provide information to insurance companies, such as the history of a ship's accidents and the condition of its equipment. This would help insurance companies to assess the risk of insuring a ship and to set appropriate premiums. And at the same time, insurance companies data could be available for authorities' controls. The Figure 6.1 represents the healthcare sector framework applied to maritime data.

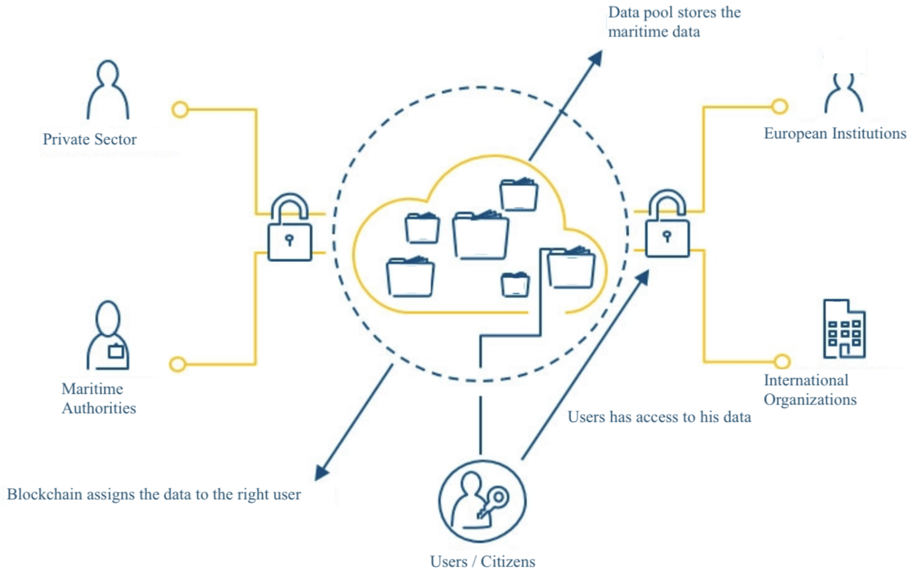

Data pool stores the maritime data
Private Sector
European Institutions
Maritime Authorities
International Organizations
Users has access to his data
Blockchain assigns the data to the right user
Users / Citizens

Figure 6.1:
Maritime data framework

6.1 Limitations and Future Studies

The limitations of the research also allow for the identification of possible future developments and can be indentified mainly in:

- the empirical case can be traced back to the sample size; in particular, a small sample of two coast guards is referred to. This means that the results may not be generalizable to all coast guards, although samples were selected so as to observe geographic diversity;

- the research focuses on the potential impact of digital technologies on the Coast Guard and the requirements needed by providing guidelines for achieving greater security, reliability and interoperability of data. However, it is important to note that the implementation of these technologies will require significant investment and resources;

- the introduction of the technologies being analyzed is a complex operation that needs to be planned on a long-term basis, providing for intermediate stages of introduction, training, and settling of use procedures;

- it could be necessary to design a framework for assessing the ethical and social implications of the development and implementation of digital technologies in the maritime context;

- a quantitative approach to quantify the efficiency improvement in the maritime sector with the introduction of the outlined technologies;

- explore the feasibility of implementing a public data space for the maritime sector one would need to interact with the key stakeholders that have been identified in the research and not just one entity;

- future discussion could be based on the standards and requirements necessary to broaden the focus globally and not within European borders.

6.2 Conclusions

Digitalization has transformed various industries, and the maritime sector is no exception. As technology continues to advance, this sector faces new challenges and opportunities. The emergence of maritime informatics in the era of advancing technology presents a critical area of study within the maritime domain. This field seeks to harness digitalization to enhance the efficiency, safety, and security of maritime sector.

The research provides a foundation for addressing the challenges and opportunities presented by digital transformation in the maritime domain. It underscores the importance of collaboration, data security, and the development of public data spaces to unlock the full potential of the maritime sector in the digital age. The research's goal is to contribute to defining the requirements of a public data space for the maritime sector, with a focus on inter-organizational collaboration and data security. While existing European projects like Gaia-x provide valuable insights, the maritime sector presents unique challenges due to its scale and specific requirements. The choice of technologies to integrate into this space will depend on specific activities and desired outcomes, the research focuses on real-time ship monitoring and administrative service provision.

The Coast Guard case highlighted the need for a public data space for the maritime sector to ensure confidentiality, integrity, and availability of data and the creation of an ecosystem aimed at high reliability and interoperability. In particular, the possibility of integrating different technologies within a public data space makes it fits for the case of maritime data, which have specificities depending on their use.

Indeed, it was pointed out that for the case of real-time navigation data it would be optimal to integrate Artificial Intelligence to develop algorithms that can be used to analyze data more efficiently and accurately and Machine Learning to develop models that can predict the behavior of maritime traffic and risks to navigation. The implementation of these technologies would allow for better understanding of maritime trade traffic, improve risk prediction, improve route planning and thus management of resources such as fuel and personnel, and allow for the refinement of anomalies detection techniques.

On the other hand, for the case of users/citizens data which involve the interaction with the maritime sector in its capacity as a public administration, blockchain could be integrated to improve document management, providing greater security against fraud and counterfeiting and to ensure the traceability of supply chain and financial transactions, providing greater transparency and security. Furthermore, it would allow the creation of databases so that larger volumes of vessel information could be managed while ensuring the privacy of private operators. The latter would ensure the creation of a data ecosystem as insurance companies would also have a way to access and share data with the Coast Guard and other maritime authorities more efficiently. Another technology that can be used to improve the digital encounter in Coast Guard administrative services is generative AI that can automate some tasks and provide information and assistance to citizens and businesses.

References

[1] M. Balduzzi, A. Pasta, K. Wilhoit, ”A security evaluation of AIS automated identification system” in *the 30th Annual Computer Security Applications Conference*, 436-445, 2014, doi: 10.1145/2664243.2664257.

[2] K. Wolsing, et al., “Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches,” *Journal of Marine Science and Engineering*, vol. 10, no. 1, p. 112, Jan. 2022, doi: 10.3390/jmse10010112.

[3] K. Wang, et al., ”Maritime Traffic Data Visualization: A Brief Review,” in *Proceedings of the International Conference on Big Data Analytics* (ICBDA), 2019, pp. 1-5, doi: 10.1109/ICBDA.2019.8713227.

[4] European Space Agency (ESA), “ESAIL maritime satellite launched”, 2020.Available: https://www.esa.int/Applications/Telecommunications_Integrated_Applications/ESAIL_maritime_satellite_launched.

[5] L. Thomas, E. Autio, “Innovation ecosystems in management: An organizing typology”, 2020, doi: 10.1093/acrefore/9780190224851.013.203.

[6] D. Beverungen, T. Hess, A. Köster, et al. ”From private digital platforms to public data spaces: implications for the digital transformation”, *Electron Markets*, 32, 493–501, 2022. https://doi.org/10.1007/s12525-022-00553-z

[7] M. Lind, R.T. Watson, M. Bergmann, R. Ward, N. Bjørn-Andersen, et al., ”Digitizing the Maritime Eco-system: Improving Door-to-door Coordination via a Digitized Transport Chain”, in *STM - Sea Trafic Management*, 2018. Available at: http://fathom.world/wp-content/uploads/2018/05/STM-concept-note-11.pdf

[8] R. T. Watson et al., “Physical and Digital Innovation in Shipping: Seeding, Standardizing, and Sequencing”, in *Hawaii International Conference on Systems Science*, 2017, doi: 10.24251/HICSS.2017.579.

[9] European Maritime Safety Agency (EMSA), “Annual Overview of Marine Casualties and Incidents”, 2022.

[10] M. W. Bockmann, “Iran Oil Tankers Said by Zanzibar to Signal Wrong Flag”, in *Bloomberg*, 2012. https://www.bloomberg.com/news/articles/2012-10-19/iranian-oil-tankers-said-by-zanzibar-to-be-signaling-wrong-flag.

[11] Windward, " AIS Data on the High Seas: An Analysis of the Magnitude and Implications of Growing Data Manipulation at Sea2", 2014.

[12] F. S. Alqurashi et al., "Maritime Communications: A Survey on Enabling Technologies, Opportunities, and Challenges", 2022.

[13] E. Lee, et al., "The Maturity of Automatic Identification Systems (AIS) and Its Implications for Innovation," *Journal of Marine Science and Engineering*, vol. 7, no. 9, p. 287, Aug. 2019, doi: 10.3390/jmse7090287.

[14] K. Wang, et al., "Maritime Traffic Data Visualization: A Brief Review," in *Proceedings of the International Conference on Big Data Analytics* (ICBDA), 2019, pp. 1-5, doi: 10.1109/ICBDA.2019.8713227.

[15] Italian Space Agency, " Sullo spazio l'Italia si conferma protagonista in Europa", in CM22, 2022. Available: https://www.asi.it/?p=36865.

[16] S. Li, S. et al., "Data Reception Analysis of the AIS on board the TianTuo-3 Satellite",*The Journal of Navigation*, 70(4), 761-774, 2017, doi:10.1017/S0373463316000916.

[17] M. Srivastava, et al., "China building cyber weapons to hijack enemy satellites, says US leak", *Financial Time*, 2023. Available: https://www.ft.com/content/881c941a-c46f-4a40-b8d8-9e5c8a6775ba.

[18] Aviation Week Network, "Viasat KA-SAT Satellite In Europe Still Under Attack In 2023", 2023. Available: https://aviationweek.com/aerospace/commercial-space/viasat-ka-sat-satellite-europe-still-under-attack-2023.

[19] M. Manulis, et al., "Cyber security in New Space: Analysis of threats, key enabling technologies and challenges", *International Journal of Information Security* 20, 2021, doi: 10.1007/ s10207-020-00503-w.

[20] Safety4sea, "Cyber attacks on maritime OT systems increased 900% in last three years", in *The Editorial Team in Cyber Security*, 2020. Available:https://safety4sea.com/cyber-attacks-on-maritime-ot-systems-increased-900-in-last-three-years/.

[21] F. Akpan, G. Bendiab, S. Shiaeles, S. Karamperidis, and M. Michaloliakos, "Cybersecurity Challenges in the Maritime Sector," *Network*, vol. 2, no. 1, pp. 123–138, Mar. 2022, doi: 10.3390/network2010009.

[22] L. Jensen, "Challenges in Maritime Cyber-Resilience", *Technology Innovation Management Review*, 2015, 35-39, doi:10.22215/timreview/889.

[23] Y. Jo, et al., "Cyberattack Models for Ship Equipment Based on the MITRE ATT&amp;CK Framework," *Sensors*, vol. 22, no. 5, p. 1860, Feb. 2022, doi: 10.3390/s22051860.

[24] European Space Agency (ESA), “Integrated Applications Promotion - IAP”, 2012.Available: https://www.esa.int/About_Us/Ministerial_Council_2012/Integrated_Applications_Promotion_IAP.

[25] R. Adner, “Ecosystem as Structure: An Actionable Construct for Strategy”, *Journal of Management*, 2016, vol. 43, doi:10.1177/0149206316678451.

[26] G. L. Romme, et al., “Mapping, analyzing and designing innovation ecosystems: The Ecosystem Pie Model,” *Long Range Planning*, 53(4):101850, 2020.

[27] M. Jacobides, et al., “Towards a Theory of Ecosystems”, *Strategic Management Journal* 39, 2018, doi: 10.1002/smj.2904.

[28] C. Alaimo and J.Kallinikos and A. Aaltonen, “Data and Value,” *Handbook of Digital Innovation*, edited by Nambisan, S, Lyytinen, K. and Yoo, Y. (pp.162-178), chapter 9, 2020.

[29] M. Jovanovic, et al., “Co-evolution of platform architecture, platform services, and platform governance: Expanding the platform value of industrial digital platforms,” *Technovation*, 2021.

[30] E. B. Swanson, ”When Data Becomes Infrastructure and our Lives Depend on it,” 2021.

[31] B. Otto, M. Jarke, ”Designing a multi-sided data platform: Findings from the International Data Spaces case”, in *Electronic Markets*, 29(4), 561–580, 2019. https://doi.org/10.1007/s12525-019-00362-x

[32] H. Richter, P.R. Slowinski, ”The data sharing economy: On the emergence of new intermediaries”, in *IIC - International Review of Intellectual Property and Competition Law*, 50(1), 4–29, 2019. https://doi.org/10.1007/s40319-018-00777-7

[33] F. Hunke, D. Heinz, G. Satzger, ”Creating customer value from data: foundations and archetypes of analytics-based services”, in *Electronic Markets*, 32(2), 2022. https://doi.org/10.1007/s12525-021-00506-y

[34] European Commission, ”A European Strategy for Data”, Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions, 2020. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0066&from=EN

[35] International Data Space Association. Available: https://internationaldataspaces.org/

[36] Global Architecture for Interoperable Automated Data Spaces. Available: https://www.gaia-x.eu/

[37] Gaia-X Franco-German Position Paper. Available: https://www.bmwk.de/Redaktion/DE/Downloads/F/franco-german-position-on-gaia-x.pdf?__blob=publicationFile&v=10

[38] N. Kazemargi, P. Spagnoletti, P. Constantinides, A. Prencipe, "Data control coordination in cloud-based ecosystems: The EU GAIA-X Ecosystem", in *Research Handbook on Digital Strategy*, 2023. doi: 10.4337/9781800378902.00024.

[39] B. Otto, M. Hompel, S. Wrobel, "Designing Data Spaces", chapter: 1-4,13, 2022. doi: 10.1007/978-3-030-93975-5.

[40] C.T.Goodsell, "The Public Encounter: Where State and Citizen Meet", Indiana University Press, 1981.

[41] I. Lindgren, I., C. Ø. Madsen, S. Hofmann, U. Melin, "Close encounters of the digital kind: A research agenda for the digitalization of public services" in *Government Information Quarterly* , 36(3), 427, 2019.

[42] R. Anderson, Ross, "Security Engineering: A Guide to Building Dependable Distributed Systems", chapter: 23,27, Third Edition 2020.

[43] J.C. Henderson, N. Venkatraman,"Strategic Alignment: A model for organizational transformation through information technology", in *T.A. Kocham and M. Useem (eds.) Transforming organizations*, 1st edn, New York: Oxford University Press, pp. 97–117, 1992.

[44] L. Wessel, A. Baiyere, R. Ologeanu-Taddei, J. Cha, T. Blegind Jensen, "Unpacking the Difference Between Digital Transformation and IT-Enabled Organizational Transformation", in *Journal of the Association for Information Systems*, 22(1), 102-129, 2021. doi: 10.17705/1jais.00655.

[45] International Maritime Organization (IMO), "Guidelines for the onboard operational use of shipborne automatic identification systems (AIS)", 2019.

[46] SOLAS, 1. SOLAS, Safety Of Life At Sea. "Carriage Requirements for Shipborne Navigational Systems and Equipment". s.l. : SOLAS, 2000. Vol. Chapter V.

[47] M. Franklin, A. Halevy, D. Maier, "From databases to dataspaces" in *ACM SIGMOD Record*, 34(4), 27–33, 2005. https://doi.org/10.1145/1107499.1107502

[48] C. Bartelheimer, P. zur Heiden, H. Lüttenberg, D. Beverungen, "Systematizing the Lexicon of Platforms in Information Systems: A Data-Driven Study" in *Electronic Markets* , 32(1), 2022. https://doi.org/10.1007/s12525-022-00530-6

Appendix

Appendix 1

Figure 5.3 illustrates the functioning of information sharing within the AIS system, as documented in (1). AIS operates by capturing GPS coordinates and facilitating real-time information exchange with both vessels and maritime authorities. This exchange occurs through VHF radio transmission, specifically utilizing two distinct radio channels operating at frequencies of 161.975 MHz and 162.025 MHz. These frequencies align with the services provided by online AIS providers. Data suppliers primarily gather information via geographically distributed AIS gateways located along coastlines and administered by port authorities, as well as through VTS systems managed by maritime authorities. VTS systems serve as a monitoring framework similar in concept to air traffic control systems in aviation.

In particular, even individual entities, such as a ship's captain, possess the capability to share AIS data with their chosen suppliers through mobile applications and specialized forwarding applications. These applications duplicate and promptly transmit the data as it becomes available. AIS data transmission occurs regularly, with intervals ranging from a few seconds to minutes, contingent on the type of information being transmitted and prevailing station conditions. For example, vessels equipped with Class B transponders and traveling at speeds exceeding 23 knots are mandated to broadcast their positions every 5 seconds. Conversely, AtoN entities like lighthouses or buoys send hazard notifications at 3-minute intervals. Compliant with regulations, each communicating station, including ships, must duly register and obtain valid AIS identifiers. These identifiers consist of the Maritime Mobile Service Identity (MMSI) number and the call sign, both of which are issued by recognized maritime authorities such as the Coast Guard or the Italian Ministry of Economic Development.

The MMSI, comprising nine digits, serves as a unique station identifier, with the initial three digits, referred to as ”Maritime Identification Digits (MID),” indicating the station's country of origin (e.g., 247 for Italy and 338 for the United States, in accordance with ITU-R's Table of Maritime Identification Digits). Conversely, call signs are radio designations for AIS stations designated to communicate with a diverse range of personnel, including those in maritime, aeronautical, military, and space sectors, as well as amateur radio operators. The International Maritime Organization (IMO) introduced AIS through a series of regulatory directives and guidelines, as outlined in (45), detailing the system's adoption and the associated technical requirements. These technical requirements encompass performance standards and data transmission protocols, encompassed within the IMO Convention on the Safety of Life at Sea, Chapter V, as referenced in (46).

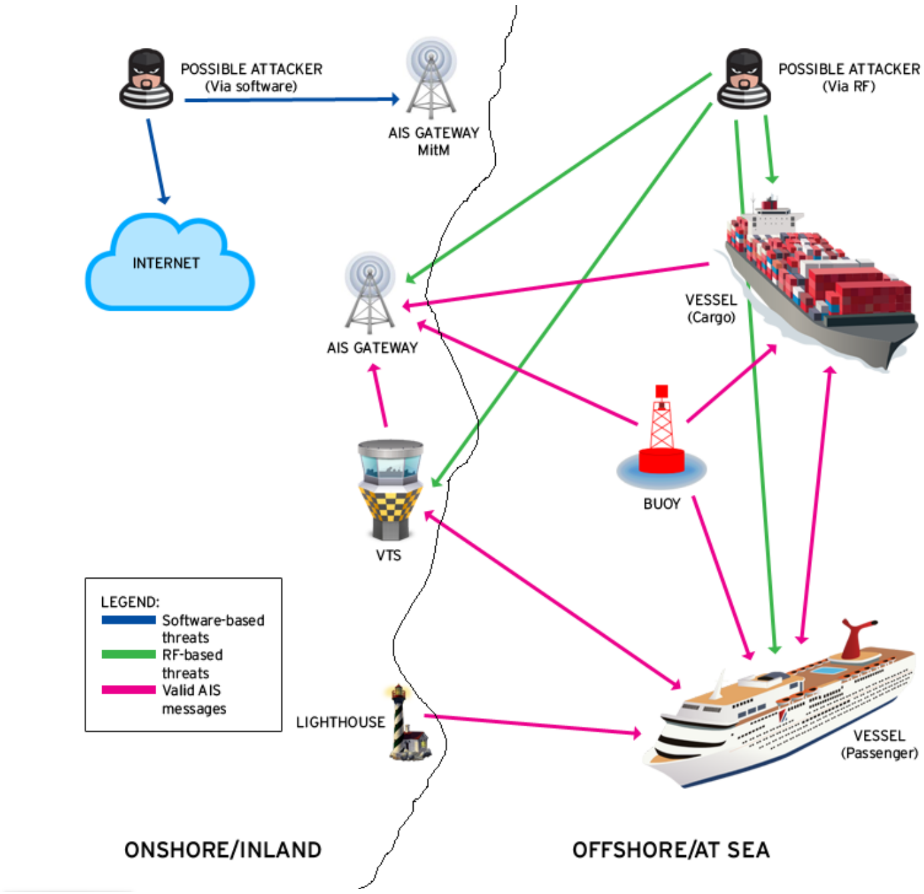


Figure 6.2:
Possible AIS attack scenarios by M. Balduzzi et al.(1)

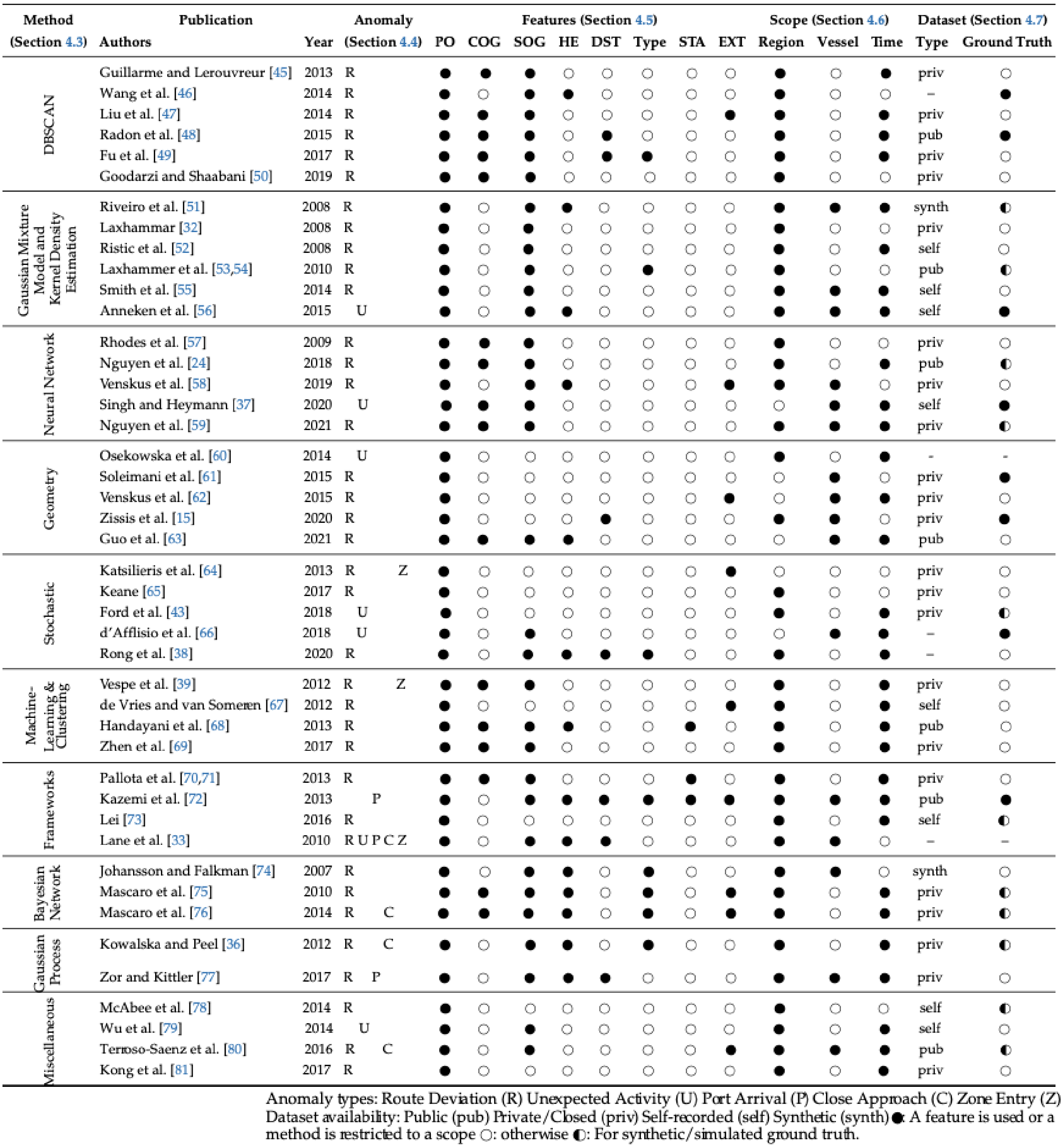
AIS information can be divide in 3 categories: static, dynamic and travel-related, as listed below

- Boat name
- Call sign
- MMSI number and international radio call sign
- IMO Number
- Type of vessel (pleasure, tug, cargo ship, oil tanker, passenger ship, SAR)
- Dimensions of the boat

- Boat position (LAT, LON)
- Speed over ground (SOG)
- Course over ground (COG)
- Boat position time in seconds (UTC)
- Navigational status
- Course heading (Headling)
- Rate of course change (ROT)

- Maximum static draught in dm
- Port of destination (UN/LOCODE)
- Estimated time of arrival (ETA)
- Specification of cargo category (class of dangerous goods)

Appendix 2



Method	Publication		Anomaly	Features (Section 4.5)								Scope (Section 4.6)			Dataset (Section 4.7)	
(Section 4.3)	Authors	Year	(Section 4.4)	PO	COG	SOG	HE	DST	Type	STA	EXT	Region	Vessel	Time	Type	Ground Truth
DBSCAN	Guillarme and Lerouvreur [45]	2013	R	●	●	●	○	○	○	○	○	●	○	●	priv	○
	Wang et al. [46]	2014	R	●	○	●	●	○	○	○	○	●	○	○	–	●
	Liu et al. [47]	2014	R	●	●	●	○	○	○	○	●	●	○	●	priv	○
	Radon et al. [48]	2015	R	●	●	●	○	●	○	○	○	●	○	●	pub	●
	Fu et al. [49]	2017	R	●	●	●	○	●	●	○	○	●	○	●	priv	○
	Goodarzi and Shaabani [50]	2019	R	●	●	●	○	○	○	○	○	●	○	○	priv	○
Gaussian Mixture Model and Kernel Density Estimation	Riveiro et al. [51]	2008	R	●	○	●	●	○	○	○	○	●	●	●	synth	◐
	Laxhammar [32]	2008	R	●	○	●	○	○	○	○	○	●	○	○	priv	○
	Ristic et al. [52]	2008	R	●	○	●	○	○	○	○	○	●	○	●	self	○
	Laxhammer et al. [53,54]	2010	R	●	○	●	○	○	●	○	○	●	○	○	pub	◐
	Smith et al. [55]	2014	R	●	○	●	○	○	○	○	○	●	●	●	self	○
	Anneken et al. [56]	2015	U	●	○	●	●	○	○	○	○	●	●	●	self	●
Neural Network	Rhodes et al. [57]	2009	R	●	●	●	○	○	○	○	○	●	○	○	priv	○
	Nguyen et al. [24]	2018	R	●	●	●	○	○	○	○	○	●	○	●	pub	◐
	Venskus et al. [58]	2019	R	●	○	●	●	○	○	○	●	●	●	○	priv	○
	Singh and Heymann [37]	2020	U	●	●	●	○	○	○	○	○	○	●	●	self	●
	Nguyen et al. [59]	2021	R	●	●	●	○	○	○	○	○	●	●	●	priv	◐
Geometry	Osekowska et al. [60]	2014	U	●	○	○	○	○	○	○	○	●	○	●	-	-
	Soleimani et al. [61]	2015	R	●	○	○	○	○	○	○	○	○	●	○	priv	●
	Venskus et al. [62]	2015	R	●	○	○	○	○	○	○	●	○	●	●	priv	○
	Zissis et al. [15]	2020	R	●	○	○	○	●	○	○	○	●	●	○	priv	●
	Guo et al. [63]	2021	R	●	●	●	●	○	○	○	○	○	●	●	pub	○
Stochastic	Katsilieris et al. [64]	2013	R Z	●	○	○	○	○	○	○	●	○	○	○	priv	○
	Keane [65]	2017	R	●	○	○	○	○	○	○	○	●	○	○	priv	○
	Ford et al. [43]	2018	U	●	○	○	○	○	○	○	○	●	○	●	priv	◐
	d'Afflisio et al. [66]	2018	U	●	○	●	○	○	○	○	○	○	●	●	–	●
	Rong et al. [38]	2020	R	●	○	●	●	●	●	○	○	●	○	●	–	○
Machine-Learning & Clustering	Vespe et al. [39]	2012	R Z	●	●	●	○	○	○	○	○	●	○	●	priv	○
	de Vries and van Someren [67]	2012	R	●	○	○	○	○	○	○	●	●	○	●	self	○
	Handayani et al. [68]	2013	R	●	●	●	●	○	○	●	○	●	○	●	pub	○
	Zhen et al. [69]	2017	R	●	●	●	○	○	○	○	○	●	○	●	priv	○
Frameworks	Pallota et al. [70,71]	2013	R	●	●	●	○	○	○	●	○	●	○	●	priv	○
	Kazemi et al. [72]	2013	P	●	○	●	●	●	●	●	●	●	●	●	pub	●
	Lei [73]	2016	R	●	○	○	○	○	○	○	○	●	○	●	self	◐
	Lane et al. [33]	2010	R U P C Z	●	○	●	●	●	○	○	○	●	●	○	–	–
Bayesian Network	Johansson and Falkman [74]	2007	R	●	○	●	●	○	●	○	○	●	●	○	synth	○
	Mascaro et al. [75]	2010	R	●	●	●	●	○	●	○	●	●	○	●	priv	◐
	Mascaro et al. [76]	2014	R C	●	●	●	●	○	●	○	●	●	○	●	priv	◐
Gaussian Process	Kowalska and Peel [36]	2012	R C	●	○	●	●	○	●	○	○	●	○	●	priv	◐
	Zor and Kittler [77]	2017	R P	●	○	●	●	●	○	○	○	●	●	●	priv	○
Miscellaneous	McAbee et al. [78]	2014	R	●	○	○	○	○	○	○	○	●	○	○	self	◐
	Wu et al. [79]	2014	U	●	○	●	○	○	○	○	○	●	○	●	self	○
	Terroso-Saenz et al. [80]	2016	R C	●	○	●	○	○	○	○	●	●	●	●	pub	◐
	Kong et al. [81]	2017	R	●	○	○	○	○	○	○	○	●	○	●	priv	○

Anomaly types: Route Deviation (R) Unexpected Activity (U) Port Arrival (P) Close Approach (C) Zone Entry (Z) Dataset availability: Public (pub) Private/Closed (priv) Self-recorded (self) Synthetic (synth) ●: A feature is used or a method is restricted to a scope ○: otherwise ◐: For synthetic/simulated ground truth.

Figure 6.3:
Anomaly detection approaches for maritime AIS tracks by K. Wolsing et al. (2)

Appendix 3

The concept of data spaces (47), which was introduced in computer science around 15 years ago, represents a novel approach to data integration. Unlike traditional central data integration methods that require physical consolidation of data and adherence to a common database schema, data spaces allow data to remain stored at its source and rely on shared vocabularies for semantic integration. This flexibility enables data redundancies and the coexistence of data within these spaces. Moreover, data spaces can be nested and overlapping, allowing individual participants to engage in multiple data spaces simultaneously.

The emergence of the Internet and information and communication technologies has transformed the landscape of data exchange, enabling cost-effective data sharing and fostering innovation beyond organizational boundaries. This shift has led to increased interorganizational cooperation within supply chains and business networks, exemplified by concepts such as the Internet of Things and Cloud Computing, which have accelerated digital transformation.

In recent years, the concept of "public data spaces" has gained prominence in scientific literature. These spaces aim to facilitate open and accessible data sharing among diverse stakeholders. While the terminology may vary, the core idea centers on creating environments where data and information can be openly shared and accessed.

The trend of organizational interconnection for data exchange is closely linked to value co-creation, transcending individual organizations. Digital platforms have played a pivotal role in this evolution, serving as generative IS artifacts that combine technology and organizational arrangements to invite third-party contributions and foster digital communities or markets. As mentioned by C. Bartelheimer et al. (48), a digital platform as a whole is, thus, 'a generative IS artifact that provides a mutual core of technology and organizational arrangements, inviting compatible and complementary resources (e.g., hardware, software, or content) from third parties to enable the emergence of digital online communities or markets'. However, public data spaces introduce a new perspective to this landscape, offering unique opportunities for data sharing and collaboration.

LIST OF ACRONYMS:

ABS: Analytics-Based Services
AIS: Automatic Identification System
ASI: Italian Space Agency
CPA: Closest Point of Approach
DDS: Digital Data Streams
DoS: Denial of Service
EDI: Electronic Data Interchange
EMSA: European Maritime Safety Agency
ESA: European Space Agency
FRAND: Fair Reasonable and Non-Discriminatory
GAIA-X: Global Architecture for Interoperable Automated Data Spaces
GIS: Geographic Information Systems
GMES: Global Monitoring for Environment and Security
GNSS: Global Navigation Satellite System
GPS: Global Positioning System
HF: High Frequency
IDS: International Data Spaces
IAP: Integrated Applications Promotion
IMO: International Maritime Organization
LRIT: Long-Range Identification and Tracking
MMSI: Maritime Mobile Service Identity
MRF: Radio Frequency
OT: Operational Technology
OSINT: Open-Source Intelligence
SAT-AIS: Satellite Automatic Identification System
UHF: Ultra High Frequency (UHF)
VHF: Very High Frequency
VTS: Vessel Traffic Services
VSAT: Very Small Aperture Terminal