Data Sharing in RoPax Ports: Challenges and Opportunities

Bogdan Iancu, Andrei-Raoul Morariu, Yiran Chen, Irina Wahlstrom, Anastasia Tsvetkova, Johan Lilius
Abo Akademi University
Turku, Finland
firstname.lastname@abo.fi

***Abstract*—Aiming to engage additional sectors of the market, numerous companies adopt data-intensive solutions for their customers. Data ecosystems have emerged across a variety of industrial sectors. By design, they create value exploiting data exchange among participating organizations. Establishing such an ecosystem in the maritime sector is accompanied by various impediments. However, exploiting maritime data to the fullest and optimizing the supply chains might become critical for future generations. We investigate, in this article, data sharing in the maritime domain, focusing on data ecosystems in RoPax ports. These ports accommodate maritime vessels which allow the transport of vehicles, passengers and goods. The choice of maritime domain and of RoPax ports specifically is rooted in their crucial role in global trade, transportation, and environmental sustainability. To understand data sharing in this context, we conducted an exploratory case study based on a Finnish RoPaX port. This study was based on a qualitative case-study approach for data collection and analysis. We identified the main challenges and practical implications of data sharing, along with their benefits in the context of RoPax ports. In addition, we provide a conceptual diagram for data sharing for future maritime data ecosystems.**

I. Introduction

A smart port is one that, via the capabilities of its (extended) port community and enabling technologies, ensures safe, resilient, and secure activities while optimizing in-, intra-, and outward movement of goods and information, driving sustainable development [13]. Modern port operations are crucial for a high-functioning economy. While in the past decade, digitalization took by storm these operations, in recent years, they were heavily affected by various unforeseen factors: global political destabilization, dire climatic changes and persistent consequences of the COVID-19 epidemic. A few of the many reasons that push port actors to spend more on digitization include the IMO (International Maritime Organization) greenhouse gas reduction targets, the IMO FAL (Facilitation of International Maritime Traffic) agreement to communicate information digitally on an obligatory basis, and the increasing limits on port growth in terms of space. Moreover, compared to road transport which requires only one document for the transport agreements, maritime transport requires more than ten records listed in an example case of seaborne delivery from the Netherlands to Sweden [14].

These critical circumstances exposed ample deficiencies in the operations of the maritime supply chain. Despite the undeniable advances in the maritime sector in many areas, there is one major hindrance in practical matters: many actors have taken a conservative approach to collaboration and, therefore, have limited particular information flow to their partners.

Many actors in the maritime supply chain hesitate to share data on a large scale with their partners. This reticence stems from various reasons. Some of these actors employ traditional channels of communication (emails, phone, faxes etc.) leaving them susceptible to errors. Others prefer protecting their competitive edge by safeguarding their data. Many maritime actors encounter legal challenges in their pursuit for data privacy and compliance. Moreover, private companies are rather vigilant with regard to data sharing with governmental institutions, which are inherently interlinked with maritime activities. Collectively, data is seldom shared at large scale in the maritime sector, thus hindering digitalization. Contrarily, data sharing is becoming more and more prominent in other domains, with reported frameworks in healthcare [36], urban development [37] or manufacturing [38].

Even though port terminals contribute to over 80% of the global trade [11], data is not nearly exploited to its full potential. To ensure consistency in port operations, the IMO mandated its members an electronic data exchange on cargo, crew and passengers. This would allow public authorities to gather critical data and store it under a unique platform, facilitating an accurate, efficient and secure exchange of information. Moreover, in an attempt to accelerate the digitalization of the maritime sector, the IMO requires its members to adopt the *single window* for data exchange, entering into force in 2024. These requirements are designed to assure the integrity and confidentiality of sensitive data. Even so, few ports actually adhere to the aforementioned requirements.

RoPax is a short term used for categorising vessels with roll-on and roll-off (RoRo) loading and unloading characteristics, often with drive-through design for the carriage of wheeled self- or unpropelled cargo units, those being commercial vehicles (trucks and trailers) and passenger cars with the capability to accommodate a fair number of passengers separately, mainly for short-sea voyages [2]. To simplify the aforementioned, a RoPax is a RoRo vessel that combines the carriage of cargo and passengers. The loading and unloading of vehicles take place via the bow, stern and side ramps or a combination of these. Compared to traditional ships, RoRo vessels and operations offer a number of benefits, such as flexibility, speed and overall time-saving. Short sea shipping is promoted by the EU due to its potential to reduce traffic volumes on roads, thus alleviating traffic congestion and associated emissions [3].

RoRo transport in Finland is preferred over the following reasons:

1) Island-like location, heavy dependence on short sea logistics,
2) Long transport distances, thin cargo volumes, small batches, Just-In-time-Truck-arrival (JIT) deliveries, scarcely populated country.

The paper is organized as follows. Section II depicts a brief literature review of hindrances and drivers of digitalization in port operations. Section III describes the challenges and opportunities brought by digitalization of the martime sector. In Section IV we present the study design and the research questions it was based on. Section V illustrates the results of the qualitative analysis of interviews, workshops, project reports and survey with companies involved. In Section VI, practical implications of data sharing challenges in RoPax ports are summarized and a conceptual diagram for data sharing in a maritime data ecosystem is proposed. Conclusions of the study and future research directions are presented in Section VII.

II. Literature Review

Over recent decades, port operations experienced several generations of digital transformation: from paperless procedures through digitization to partially automated systems and to today's smart intermodal transportation [16].

Digitalization impacts a port's daily operations [6], [7], [16] and reconfigures its business operations in terms of time, place, and forms [18]. Timewise, information sharing in advanced real-time telecommunication enhances the port's overall operational efficiency. Place-wise, digitalization allows real-time communication for multiple organizations instead of only reservations and upon port arrivals. Form-wise, physical document check-in processes are digitized gradually, easing the port's overall security control and reporting process.

Three literature streams are relevant regarding port digitalization. The first stream exposes emergent technological solutions and corresponding pros and cons [9], [17]. The second stream reviews common characteristics of port digitalization processes: trajectories, drivers, and hindrances [6], [7]. The last stream gathers empirical reviews regarding port development projects [19], [20]. Some scholars mentioned that technology is the foundation of ecosystem transition [15], [21]. However, regarding ports, more challenges are related to heterogeneous ecosystem participants and complex port operations. Hence, the last two streams were reviewed for this study.

Several frameworks attempted to assess port digitalization levels: [16] classified three generations of ports based on technological development scopes and digital impacts; Philipp [24] defined a digital readiness index for smart port-oriented development assessment based on five dimensions: management, human capital, IT functionality, technology, and information.

Barriers to digitalization unfold from several perspectives: changes in business perspectives [9], resource scarcity, incompatible existing systems, resistance to digital transformations, and security threats [7], [22]. Commonly mentioned hindrances and drivers for port digitalization are mentioned in Table I.

Limited studies took an ecosystem approach to understand port digitalization hindrances and drivers. The origin and potential combinations of these hindrances may provide more insights from an ecosystem perspective. Consequently, adapting an ecosystem perspective is critical for firms to overcome these hindrances, which is the goal of this study.

TABLE I. Hindrances and Drivers for Port Digitalization

Hindrances mentioned in literature	**Description**
Incompatible inter-organizational systems [6], [7]	Extant port operation systems may be incompatible with new technologies and inter-organizational interoperability.
Outdated management perspectives; organizational incapability [7], [12]	Despite digitalization's clear advantages, the ecosystem's participants lack the human resources and skillsets for implementing new systems to adapt to market transitions.
Security threats [7], [23]	Ports are vulnerable to technological changes; digitalization introduces new security threats specifically related to cybersecurity.
Resistance to digital change; unaligned development interests [7], [20]	Resistance to digital changes from different actors often relates to diverse business interests, high investment, and fear of human workforce reduction.
Resource scarcity [7], [20]	Lack of resources, such as financial or physical capacities, are commonly mentioned for small ports.
Drivers mentioned in the literature	**Description**
Corporate and environmental legislation [7], [22], [25]	Standardization for collaboration and data transfer. Environmental legislation related to carbon neutrality.
Business integration or restructuring [9], [12], [22], [23], [26]	Business integration with other actors through digitalization and a structured governance mechanism, such as e-commerce and logistic platforms.
Increased societal impact [22], [26]	Integration of ports into their surroundings, such as cities and significant locations.
Key enabling technologies [9], [12], [17], [22], [25], [27]	Specific technologies for ports' digitalization trajectories, such as 3D printing, AI, big data, IoT and sensor technologies.

III. Digitalization – Challenges and Opportunities

A study from Statista forecasts that global IoT-connected devices will reach more than 31 billion by the year 2025. Large solution providers such as ABB, Airbus, Chevron, and Tesla are a few pioneers in predictive maintenance and real-time performance monitoring of various systems [4]. Therefore, Industry 5.0 that will start from 2030 onward is expected to bring far more advanced systems that will create complete digital ecosystems where customers will be able to interact with systems virtually, monitoring and improving their performance.

A traditional industry like the maritime industry has been hesitant to integrate the latest digital applications into everyday processes and practices. Port connectivity is a major factor considering route planning in transport industry in accordance with the novel European Commission regulations related to pollution and energy consumption [5]. Additionally, in a shared environment with stakeholders from private and public sectors responsible for cities and road infrastructures, collaboration and interaction are an essential factor that help improve the overall network of transport on road and waterborne.

An essential aspect of digitalization is the continuous real-time data flow of information [6]. In the maritime industry, it can be beneficial for the port when trucks have a short delay in the schedule. This way, waterborne vessels can slightly postpone departure time. It is not a direction but a benefit that comes from eliminating the need for rerouting trucks to further locations because other transport vessels can help establish the waterborne connection and consequently reduce pollution.

Compared to other stakeholders, ports participate in projects related to digitalization if they perceive some monetary gain or other benefits such as reduced environmental impact, energy saving or increased social impact. In recent years, ports have been piloting various projects for implementing new technologies (some of which are in collaboration with universities). In [7], Brunila *et al.* discuss the profitable implementation of an energy monitoring model that visualizes real-time energy consumption. The project's success was extended to the connected services in the port area and contributed to decreasing the carbon footprint and energy consumption. Data sharing is one strategy for solving issues related to outdated communication methods, excessive paperwork and dissimilarities in protocols. Moreover, vessel turnaround time in ports will have a substantial advantage from the up-to-date status of all the connected operations to loading and unloading cargo and passengers. Here the four hinterland players (importers, exporters, logistics nodes and land freight) share live information with the port via a common digital platform. This can easily extend to a network of ports that will minimize inefficiencies in the worldwide logistics chains by improving the use of port capacities and ensuring more dependable and shorter transit times [9].

PortXchange [8] is a working platform developed by the Port of Rotterdam where stakeholders present information about port operations. In addition to reducing the total carbon emissions, another advantage that the system brings is improved operational efficiency and reduced idle time in the port area [12]. One example of using the system is when a ship is approaching the port, but the terminal is not free. In such cases, using the platform gives the advantage of informing the vessel to slow down and save fuel and expenses. Additionally, estimated Time of Arrival and Departure (ETA and ETD) accuracy improves from using the system [9]. Notifications about plan changes come close to real-time to stakeholders. Therefore, traffic congestion is minimized.

IV. Research questions and study design

A. Research design

The study uses a qualitative case-study approach for data collection and analysis based on secondary data sources: a survey, interviews, workshops and project reports [10]. An activity diagram depicting the research design for the study is presented in Figure 1. Based on our established research objectives and the literature review, we designed a questionnaire, an interview framework and several workshops. The data collected during these activities was systematically categorized and analysed, resulting into a collection of challenges and implications of data sharing, and complemented by a conceptual model for data sharing, which was subsequently validated by the data ecosystem stakeholders. Empirical data comprises additional notes and observations made by the authors during project meetings, workshops and company visits. Both the survey and the interviews were designed to understand the dependencies, obstacles and benefits of data sharing, considering the perspective of each of the project participants.

This study involved participants from company representatives (executives, managers and specialists) and academic partners. Data sharing [1] is essential in the operations of a modern maritime supply chain and for this very reason, we analyze its adoption in context. We structured both the questionnaire and the interview questions to correspond to a set of functions, which put data usage within the project into perspective. The functions we guided our interviews by are the following: search and discovery, metadata availability, data exchange, data archiving, data quality, data security and data-driven business opportunities. The interviews framework relied on several functions that promoted understanding, data sharing, and collaboration between partners:

- Search and discovery – to establish data availability and ownership;
- Metadata availability – to provide contextual information regarding data assets;
- Data sharing – to identify existing and potential data flows between organizations;
- Data archiving – to determine key storage management features;
- Data quality – to review implemented solutions ensuring the completeness, reliability and accuracy of data assets;
- Data lineage – to document mechanisms of tracing and tracking data evolution within the project's organizations;
- Security – to elaborate on secure solutions for selective data sharing within the project;
- Business impact – to evaluate profitability based on data governance and usage.

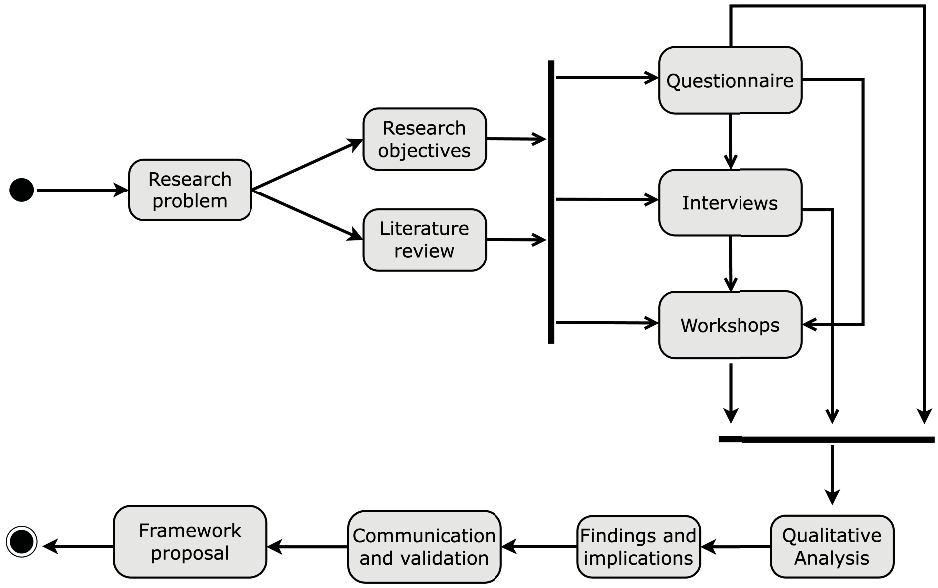

Fig. 1. Research design diagram

First of all, we were interested in search and discovery, i.e. mapping the data sources that were essential for specific use cases. Secondly, metadata availability is crucial to characterize data assets, allowing for a more transparent description and classification. Data exchange among partners is essential in identifying data flows based on different use cases, and enables collaborations and exploring new innovative solutions. Data archiving is essential for reproducibility and interoperability. Another important aspect to consider is data quality since it can drastically affect all data processes from collection to final results to business decisions. Finally, business challenges can be, in many cases, addressed by utilizing data insights.

The study addresses the following research questions:

RQ1: What project-related data does the interviewed organization produce/utilize?

RQ2: Which metadata enriches the standalone data (text data, taxonomies, ontologies)?

RQ3: What data sharing practices does the organization follow?

a) What are the already established data sharing entities in the project?
b) Which platforms/APIs are already established within the project for data sharing?
c) If such platforms are not established, what platforms does the organization use to share data?
d) What data can the organization provide to other project partners? Are such pipelines/APIs readily available?

RQ4: What data archiving practices does the organization follow?

a) How long does the organization store the data for?
b) When is data discarded?

RQ5: What data quality practices does the organization employ?

a) Does the organization collect data?
b) How does the organization process raw data?
c) How does your organization process data retrieved from other organizations in the project?

RQ6: How does the organization manage data lineage?

a) What are the sources for data?
b) Can data be traced from the beginning to the end of a process?

RQ7: Does the company use or implement/plans to implement selective-based access to the data for other parties in the project

RQ8: How does data enhance your business perspective?

a) How different data is used or can be used, e.g., for situational awareness, for optimizing processes and flows, for predicting, etc.?
b) Which data the companies find most valuable, vendable, and to whom?
c) Which available data can be combined for generating useful information/knowledge?

B. Threats of validity

Construct Validity: The semi-structured interview was designed by the author group according to previous experience and knowledge. Before each interview, an introduction was made about the purpose and content of the interview. We began the interview question set with simple questions meant to have an affirmative or negative response on whether interviewees' companies share data with other project-related companies. The interview continued with more in-depth questions on the subject, meant to better understand the process.

Interview Validity: Our goal was not to establish a cause-effect relationship between companies' data and possible connected businesses. Instead, we aimed to document data avenues as a data catalog to create transparency for project members on the activities of each party. Some interview questions had to be further clarified when the interviewees may not have understood them clearly. Our recommendations from the data catalog may not lead to a fast transition to data sharing between service providers in the port area. Some organizations may not have similar technology or the resources necessary for performing data sharing.

Reliability: Some of the questions asked did not have either a clear answer or a negative answer, which, in most cases, did not lead to any further talks. The researchers taking part in this study may have not understood some of the answers to the interview questions as clearly as they were spoken to by the interviewees. Therefore, in order to increase the validation, the responses to the interviews were analyzed by two of the authors of the study. Some answers were edited to provide a better understanding for the reader.

C. Data collection and analysis

The selection of interviewees was made within the author group according to a contact list of project participants. Considering their involvement in the project and partial collaboration with us, we scheduled and interviewed the selected professionals.

The interview process started with a follow-up of a previously organised survey where some of the stakeholders connected to the project did not respond to a set of questions received via electronic post. Consequently, it was decided to approach the situation using another method of having an email agreement of the interviewee of an agreed time frame for the interview.

A total of 12 professionals were interviewed during four weeks between 2022-05-19 and 2022-06-16. The availability of participants made some short changes in the original interview timeline, having extended the total interview period to approximately one month. The interviewees were informed that the interview will lead to a publication that is meant to enhance the practicalities and economical development of their businesses. The time frame scheduled with the professionals was of 60 minutes, with an average of 42 minutes per participant. The transcription of the interviews was done simultaneously as the conversation progressed.

D. Case study

The current research is the outcome of a three year long project that aimed at collectively designing digital solutions for RoPax ports that would allow increasing the efficiency and autonomy in logistics operations while also reducing emissions. The mission of the project was to create replicable models for digitalisation, service innovation and data usage and sharing in harbour environment and prepare for the future by taking steps towards smart and autonomous maritime transportation. Project participants included both the actors from port operation ecosystem such as port authorities and shipping companies, and digital solution providers.

V. Results

In this section, we discuss the main findings of the research. The interview results were categorized systematically according to the questions and connected discussions. It is our goal to develop a model for data sharing that incorporates both practices and obstacles stakeholders must overcome.

Table II characterizes the project consortium, where we interviewed most private stakeholders. We conducted a semi-structured interview to encourage conversation and help us better comprehend the realities of sharing data among companies. Identifiers were used within the first column for better transparency of the interviewees. We found that IT is the most commonly met profile of the companies, with a few in transport, engineering, animation, and port services. According to online sources, many firms have below one hundred employees, with few outliners from ship carriers which have over a thousand in size. Education institutions have different sizes, with few in order of hundreds and others in order of up to three thousand.

The interviews aimed to promote the usage of data in the project, documenting how data was organized, collected and which access modalities were most prominent.

A. Project-related data usage (RQ1)

In this section, we present the variety of data produced or utilized by different companies in the considered data ecosystem. Given the complex network of partners (private and public) that make up the effective operation of a RoPax port, the range of data produced, utilized and/or shared is extensive. This data can be crucial in sustainably adapting port design, flow optimization, process automation, and degree of autonomy in current and future smart ports. In our exploratory case study, we can categorize data use and sharing among the actors in the ecosystem based on three strategic segments of the maritime sector, i.e. operations at sea, in the port area and on land. Several of the companies (both private and public) included in the case study are involved in at least two of the aforementioned segments of the maritime sector, highlighting the scale of complexity of operations in smart ports and the degree of the interdependency of its actors.

Collection and processing of maritime data is heavily affected by the nature of the data, emerging often from continuous monitoring, numerical modeling or maritime market tracking. Hence, the range of data utilized and/or produced in the case study is vast, extending from simple text to voice recording, images for specific use-cases, AIS data, portal history, port receipts, train locations, winter navigation data, port gate events, vehicle location inside the port, traffic data, meteorological data, etc.

B. Metadata enrichment of the standalone data (RQ2)

Most interviewees acknowledge the significance of metadata in maintaining historical aspects of datasets and in the robust

TABLE II. PROJECT PARTNER DETAILS, ANONIMISED

ID	Size	Profile	Sector	Operation scope
D1	180	IT consultancy in software products for multilingual business communication	Private	Language management
D2	22	IT technology provider in software and hardware for remote and autonomous systems	Private	Remote and autonomous solutions
D3	28	Engineering company developing and implementing automation in multiple industries	Private	Automation solutions
D4	28	IT consultancy and services	Private	Shipping software design
B1	32	IT technology provider in software and hardware surveillance systems	Private	Identification and management of vehicle, cargo, people
B2	45	Engineering and design for machineries	Private	Solution provider for infrastructure environments
B3	19	Animation design and effects	Private	Virtual training platform
B4	26	Transport services	Private	Logistics
P1	863	Technology group	Private	Telecommunication services
P2	80	Port	Private	Port infrastructure management
P3	83	Port	Private	Port infrastructure management
P6	1555	Ro-Pax passenger ship operator	Private	Cargo and passenger shipping service
S11	1155	Ro-Pax passenger ship operator	Private	Cargo and passenger shipping service
S12	3314	Education	Public	Research and education
S13	1100	Education	Public	Research and education
S14	320	Education	Public	Research and education
S15	717	Education	Public	Research and education
S16	2600	Education	Public	Research and education

management of the data life-cycle. Moreover, metadata is crucial in governing various aspects of data quality management: availability of data, inconsistencies regarding data types or formats, missing information, biased data, insufficient information regarding deficient datasets, etc. To address these challenges, several of the interviewed companies mentioned various data avenues they used to enrich standalone data: taxonomies, ontologies, simulation environments, digital twin solutions and mappings of public datasets.

C. Practices of data sharing within the company (RQ3)

In this section, we present various practices for data sharing employed by the companies involved in our case study. First and foremost, more than half of the participants reported that data was not shared with other project participants at the time of the interview. Our interview platform served also as a tool to analyze and reflect on these practices and promote data sharing while considering its risks and limitations.

One of the central limitations that seems to permeate data exchange practices in the maritime sector is data privacy, especially brought about by GDPR guidelines. Since this is still a somewhat new regulation relative to the history of maritime transportation, various actors treat data sharing with a certain wariness. Many of these actors are small companies or individual contractors, who often do no benefit of an in-house GPDR expert, and for this very reason they prefer to safeguard data against any possible liabilities in the simplest of manners: they just do not share it. This often stems from a deeper problem in the maritime sector, these companies do not have the incentive to share data and they have to put forward great efforts into building physical IT infrastructures and software pipelines to share data, with only an expectation they would secure long-term benefits by doing so. Of course, this is in practice unsustainable and many actors decide against data sharing.

Of the interviewed project partners a few mentioned they already had specific platforms, either commercial or in-house developed for data sharing. The solutions were varied, some benefited from their own APIs that can be used by clients to connect to their platforms, some mentioned using Google Cloud based platforms to offer selected access to different data streams. Others used Microsoft Azure or Oracle Cloud Infrastructure. One company presented their own MLOPS solutions promoting smart maritime logistics. One project participant mentioned a video stream-based access based on VPN. Other mentioned Sharepoint and Teams as possible means of data access.

D. Data archiving practices (RQ4)

Regarding data archiving among the interviewed stakeholders, one common practice that emerged was the permanent storage of all data collected or used by the company. Whether data is stored in the cloud or on-premises servers, this arose as a popular solution with almost half of the participants in the interviews reporting it. This might seem counter-intuitive in the era of cloud-computing, but maintaining their data onsite gives shareholders more control, and ensures that fewer individuals have access to it when a third-party is not involved, which is perceived as a safer option when it comes to critical data.

E. Data quality procedures (RQ5)

In this section, we describe multiple aspects of the data that the project stakeholders use. According to the interview information, data collection is scarce, with only a few vague mentions. Some stakeholders use open data, previously collected or already processed data. Customers have design-specific requirements depending on their needs, where pre-processing and normalising are sometimes part of the processing pipeline. Some stakeholders mentioned collecting raw surveillance data used for optical character recognition. In one case, vehicle license plate recognition and imagining are performed. The usage of Microsoft Azure was mentioned, as well. The last question in this category was meant to encourage the interviewees to share about the practices of using data from other project partners. One mentioned estimating the ship departure according to the truck loading status. In sum, the practices are disparate and there was little system-level coordination of various processes.

F. Data linage practices (RQ6)

In order to understand how stakeholders can track system errors or debug the system, the author's group decided to ask a question about data traceability [28]. Provenance is sometimes ambiguous, considering filtering and data reduction performed before transmission to cloud systems. Some applications are designed to interact with data by performing different methods such as collecting key performant indicators.

Some perceived the question as similar to Section V-D, regarding archiving methods. Others mentioned using their cloud-based platforms or external services such as Microsoft Azure or Oracle for traceability. The most important remark on this question is that most interviewees responded affirmatively to the data lineage question. One mentioned the capability of regenerating their data in case a storage node fails.

G. Data access for project members (RQ7)

Data access and security are sensitive topics for most of the stakeholders. In many cases, a multiauthority cloud scheme is in use which handles efficient encryption/decryption, and different access rights. Therefore, a company representative, a contract-based collaborator, and a port authority may withhold different attributes that were issued by several authorities. Generally, all access rights are based on a set of regulations defining conditions for every partner.

Few respondents noted giving selective-based access rights to universities and port authorities. It is generally more convenient for stakeholders to give access to universities due to project-based collaboration, and increased chances of new discoveries from research. One mentioned offering access to a traffic database for the company's customers.

H. Business data sharing opportunities (RQ8)

A majority of the interviewed companies mentioned data-enabled business enhancement in different stages of the port operation. The highest-valuated data among the companies differ for cargo and passenger traffic. For cargo traffic, companies appreciate operational data from infrastructure (gate data and terminal situations), and ships (AIS and port calls). The data could be collected in real-time or historically, which can enable situational awareness of multiple parties, supports terminal operation flow modelling and worker training, and hence, leads to smooth intermodal transport and stress-free port operation. The outcome benefits numerous parties, such as cargo owners, ship operators and logistic companies. For passenger traffic, companies aim to collect data related to improving passenger user experience. For instance, knowing the needs of the passengers alongside the journey, not only inside the terminal. Data is used to optimize passengers' intermodal travelling experience, reduce waiting time, alleviate urban congestion, and prevent safety or security threats. Shipping companies, ports, public transport providers and municipalities showed interest in this data type. Several data combination proposals were mentioned during the interviews. Nevertheless, the commercialization of such a proposal is currently challenging due to diverse impediments, for instance, low quality of existing data, lack of data collection facilities, or limitations due to data sharing regulations for privacy protection (GDPR, etc.).

VI. Discussion

Adopting the maritime data ecosystem paradigm has many benefits, which were identified in Section III. At the heart of a maritime data sharing ecosystem is the port, which has the unique ability to impose regulatory data exchange incentives through port authorities as illustrated in Fig. 2. Port authorities can establish a data sharing timeline for the participants in in the ecosystem. Moreover, the port can demand involved actors to find a consensus mechanism for data sharing. One finding from our interviews and workshop was that many companies find it difficult to adapt their business model to integrate into a data ecosystem. This reflects also in their data sharing practices and openness. The port can also impose data sharing policies, facilitating in this way a smooth flow of data among the data ecosystem participants.

Based on research questions RQ1-RQ8, we identified and summarized the main challenges we encountered in our case study with regards to data sharing, see Fig. 2. The challenges can be classified in three non-exhaustive categories: legal, economic, and technical, which we briefly describe as follows. Some companies lack the infrastructure to decisively establish the ownership, provenance and trustworthiness of data, which prevents them from opening their data resources to the ecosystem players at large scale. One solution proposed was the exchange of data through selective-based access and providing usage rights for specific individuals or groups of individuals. A second aspect (see Fig. 2) often mentioned in our discussions with ecosystem partners was of economic nature. Organizations abide by certain economic responsibilities and aim to preserve or improve their profitability maintaining a competitive advantage. This often means that companies have a preference for preserving data sovereignty, in an attempt to maintain a perceived cutting edge over competitors. Frequently, in many organizations data is stored indefinitely, impairing in this way awareness. The absence of data sharing agreements in many cases expose companies to various liability risks, which could result into high financial losses. By designing specific regulatory networks and distribution mechanisms for data sharing we can incentivize companies to share data. A third aspect (see Fig. 2) is of technical nature and is related to the actual exploitation of shared data. Some data is unavailable, other can be of a quality difficult to establish.

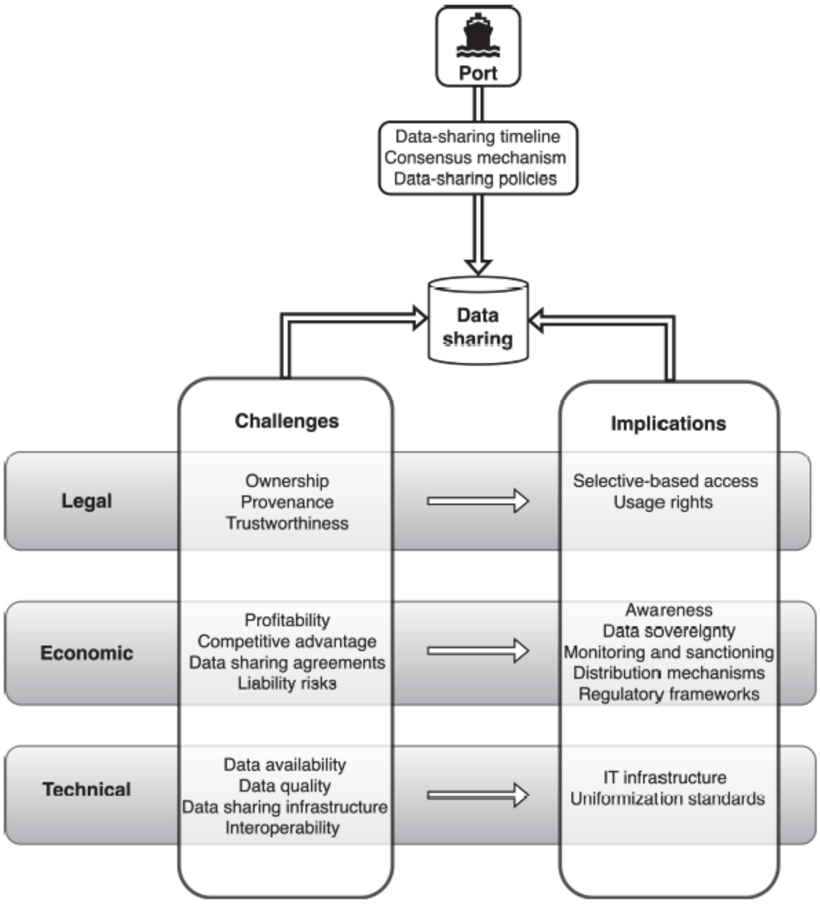

Fig. 2. Challenges and implications for data sharing in a maritime data ecosystem

Many SMEs develop efficient solutions which can only be scaled up and shared through a drastic infrastructure change. Another important aspects that emerged in our workshops was interoperability, which is absolutely essential to be able to harness the benefits from data sharing. Investing resources to scale IT infrastructures and uniformize standards of data artifacts is necessary for future data-intensive solutions in ports.

One prevalent concern raised during our interviews and workshops was data privacy and security, specifically how can data be safeguarded within a big data ecosystem. A few privacy and security related solutions addressing data sharing, which could be investigated in future maritime data ecosystems, include:

- Homomorphic encryption [30]: promotes executing computations on encrypted data directly, the output being identical had they been performed on unencrypted data.
- Differential privacy [31]: enables sharing information concerning specific group patterns in a dataset avoiding the disclosure of information about individual samples in the given dataset.
- Federated (collaborative) learning [32]: facilitates the training of AI algorithms across multiple decentralized devices, on locally-managed datasets, without any explicit exchange of data.
- Zero-knowledge encryption [33]: is achieved through a unique user-key (only known by the end user, who get the sole access to the encrypted data).
- Privacy-preserving computation [34]: promotes the distribution of operations across multiple devices in such a manner that not one device has access to the entire array of inputs.

A data governance shift in ports requires not only a change in attitudes but also robust data management systems that are easily scalable and adaptable. We provide such a conceptual model in Fig. 3. The platform for data sharing we propose in the conceptual diagram in Fig. 3 is designed to be aligned with the Gaia-X initiative and its federated data infrastructure [39]. Data streams in Fig. 3 originate from various sources: sensor arrays, navigation and communication, condition monitoring, business intelligence, etc. Data is shared through various dynamic networks, which are exposed to different levels of uncertainty. The classification and validation of data through quality monitoring is a prerequisite to building valuable recommendations for port operators and it starts by simply understanding how maritime actors plan their operations and how they execute them. To achieve this, a first

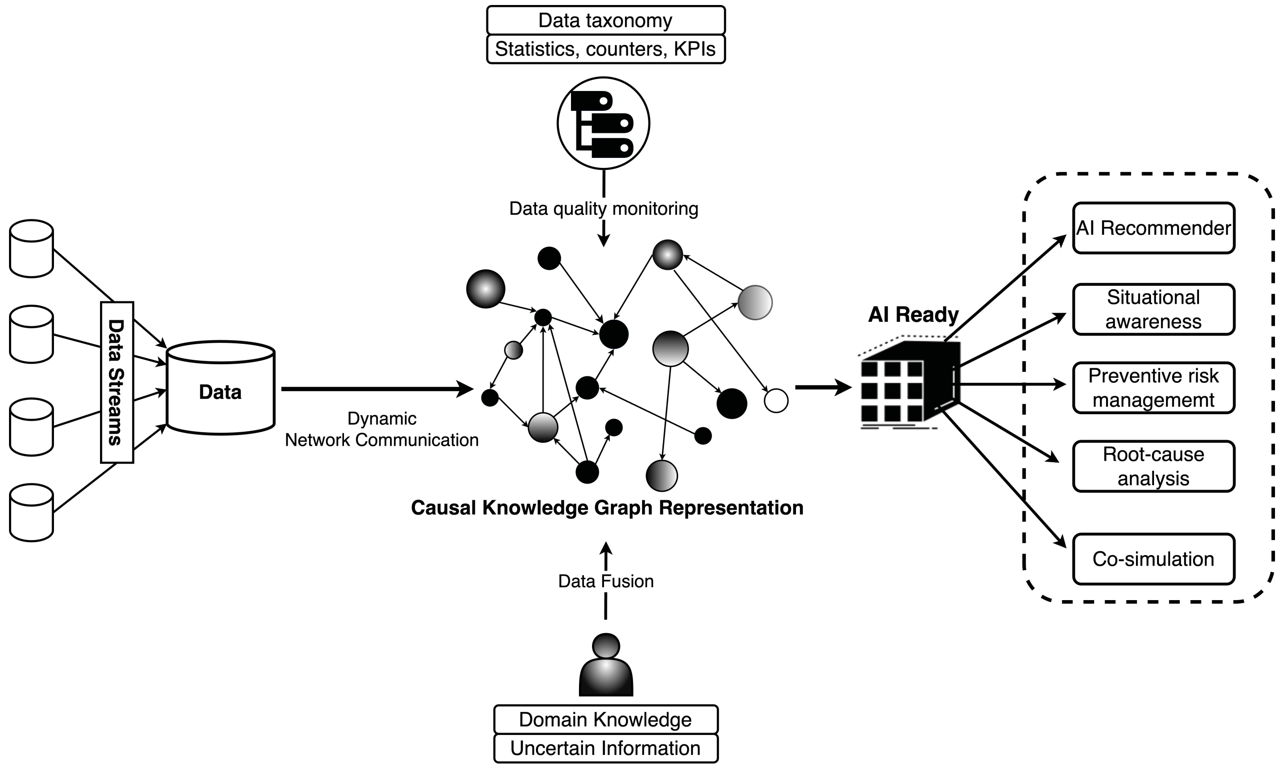

Fig. 3. Conceptual model for data management in a maritime data ecosystem

step is to establish data taxonomies and perform exploratory analysis to extract initial counters and evaluate possible KPIs. Domain knowledge can determine causal relationships among various data streams ([29]) and act as a catalyst by promoting various data fusion solutions to create a holistic perspective of the data sharing ecosystem. By integrating causal knowledge with collected data, AI models can learn causal relationships responsible for inaccurate inferences, eventually improving the performance of algorithms. Such algorithms contribute to situational awareness for (semi-)autonomous navigation, preventive maintenance for maritime machinery, root-cause analysis to aid fault tolerance, AI recommenders to address various challenges in harbor environments and co-simulation to holistically model and test different functional scenarios in the data ecosystem. This approach also promotes the integration of explainable components, which can be developed utilizing tools that support continuous data quality monitoring. Similar architectures have been proposed for employing data ecosystems in other domains, i.e. agriculture [35].

VII. Conclusion and Future Work

We illustrated the challenges, benefits and opportunities induced by adopting the data ecosystem paradigm to promote data sharing in the maritime sector, focusing on a RoPax port. We considered various aspects of this paradigm shift, establishing first data management practices used by various stakeholders in our ecosystem, revealing constraints and requirements that are necessary for organizations to migrate towards this model. In addition, we determined the limitations of this ecosystem paradigms for medium sized RoPax ports, identifying how these challenges reflect in practical implications of the port data ecosystem. Furthermore, we proposed a data management system based on literature and following upon our discussions with various stakeholders during interviews and workshops.

To summarize, by sharing and aggregating data, stakeholders can evaluate the financial impact of adopting the data ecosystem paradigm in a continuous manner. Data sharing has several advantages, including expanded market potential, improved company and investment stability, more productive workflows, and lower, even more predictable expenses. Stakeholders can learn from past decisions through a data lens and develop solutions by using data from previous projects to help them make better decisions. On the other hand, ports can become the key players in establishing data platforms with controlled access for connected businesses in the maritime sector.

The maritime domain imposes various restrictions and limitations on data sharing arising from its inherent complexity. The challenges and their implications have been summarized in Fig. 2. Defining a strict list of requirements for the construction of a sustainable maritime data ecosystem for RoPax ports can be challenging given the highly unique context of every RoPax port. However, there are some aspects that encompass the prerequisites for creating a sustainable data ecosystem which we summarize below. First of all, all actors included in the data ecosystem must commit to managing potential liabilities. When they do not commit, data is generally not shared by the uncommitted actors. Secondly, participating stakeholders should reach a consensus regarding data sharing paradigms and policies. Often, stakeholders have discordant views with regard to data sharing policies. Lastly, clear protocols must be established for managing sensitive information. This brings trust among shareholders and, thus, promotes the development of data sharing strategies.

Regarding future work, we will study the viability of data sharing solutions from other domains (healthcare [36], urban development [37] or manufacturing [38]) to determine which novel approaches are suitable to address the particularities introduced by the maritime domain.

Acknowledgment

This work was supported by projects funded by Business Finland.

References

[1] Morariu, A., Iancu, B. & Björkqvist, J. Practices for Data Sharing: An Empirical Survey. *Data Analytics 2022*. pp. 27-32 (2022)

[2] Sauri, S., Morales-Fusco, P., Toledano, M. & Martin, E. Empirical Analysis of Resiliency of Terminal Operations for Roll-On–Roll-Off Vessels. *Transportation Research Record*. **2273**, 96-105 (2012)

[3] Douet, M. & Cappuccilli, J. A review of Short Sea Shipping policy in the European Union. *Journal Of Transport Geography*. **19**, 968-976 (2011)

[4] Statista Industry 4.0 – In-depth Market Insights & Data Analysis. (https://www.statista.com/)(2023)

[5] Raza, Z., Svanberg, M. & Wiegmans, B. Modal shift from road haulage to short sea shipping: A systematic literature review and research directions. *Transport Reviews*. **40**, 382-406 (2020)

[6] Inkinen, T., Helminen, R. & Saarikoski, J. Port digitalization with open data: Challenges, opportunities, and integrations. *Journal Of Open Innovation: Technology, Market, And Complexity*. **5**, 30 (2019)

[7] Brunila, O., Kunnaala-Hyrkki, V. & Inkinen, T. Hindrances in port digitalization? Identifying problems in adoption and implementation. *European Transport Research Review*. **13**, 1-10 (2021)

[8] Suvadarshini, P. & Dandapat, P. Digitalizing the maritime supply chain: The case of Rotterdam's port call operations. *Journal Of Information Technology Teaching Cases*. (2022)

[9] Hirata, E., Watanabe, D. & Lambrou, M. Shipping Digitalization and Automation for the Smart Port. (IntechOpen,2022)

[10] Yin, R. Case study research: Design and methods. *Sage*. (2009)

[11] Sarabia-Jácome, D., Palau, C., Esteve, M. & Boronat, F. Seaport data space for improving logistic maritime operations. *Ieee Access*. **8** pp. 4372-4382 (2019)

[12] Tsvetkova, A., Gustafsson, M. & Wikström, K. Digitalizing maritime transport: digital innovation as a catalyzer of sustainable transformation. *A Modern Guide To The Digitalization Of Infrastructure*. pp. 123-148 (2021)

[13] Boullauazan, Y., Sys, C. & Vanelslander, T. Developing and demonstrating a maturity model for smart ports. *Maritime Policy & Management*. pp. 1-19 (2022)

[14] Hvolby, H., Steger-Jensen, K., Bech, A., Vestergaard, S., Svensson, C. & Neagoe, M. Information exchange and block chains in short sea maritime supply chains. *Procedia Computer Science*. **181** pp. 722-729 (2021)

[15] Adner, R. Ecosystem as structure: An actionable construct for strategy. *Journal Of Management*. **43**, 39-58 (2017)

[16] Heilig, L., Schwarze, S. & Voß, S. An analysis of digital transformation in the history and future of modern ports. (2017)

[17] Sanchez-Gonzalez, P., Diaz-Gutierrez, D., Leo, T. & Nunez-Rivas, L. Toward digitalization of maritime transport?. *Sensors*. **19**, 926 (2019)

[18] Lusch, R., Vargo, S. & Tanniru, M. Service, value networks and learning. *Journal Of The Academy Of Marketing Science*. **38** pp. 19-31 (2010)

[19] Feng, L. & Notteboom, T. Peripheral challenge by small and medium sized ports (SMPs) in multi-port gateway regions: The case study of northeast of China. *Polish Maritime Research*. **20**, 55-66 (2013)

[20] Gerlitz, L. & Meyer, C. Small and medium-sized ports in the tent network and nexus of Europe's twin transition: The way towards sustainable and digital port service ecosystems. *Sustainability*. **13**, 4386 (2021)

[21] Autio, E. Orchestrating ecosystems: a multi-layered framework. *Innovation*. **24**, 96-109 (2022)

[22] Inkinen, T., Helminen, R. & Saarikoski, J. Technological trajectories and scenarios in seaport digitalization. *Research In Transportation Business & Management*. **41** pp. 100633 (2021)

[23] Langen, P. Seaports as Clusters of Economic Activities. *International Encyclopedia Of Transportation: Volume 3. Freight Transport And Logistics*. pp. 310-315 (2021)

[24] Philipp, R. Digital readiness index assessment towards smart port development. *Sustainability Management Forum— NachhaltigkeitsManagementForum*. **28**, 49-60 (2020)

[25] Chuprina, E., Zahorodnia, Y., Petrenko, O., Britchenko, I. & Goretskyi, O. Specific characteristics of seaports development in the context of digitalization: international experience and conclusions. *International Journal Of Agricultural Extension*. (2022)

[26] Haraldson, S., Lind, M., Breitenbach, S., Croston, J., Karlsson, M. & Hirt, G. The port as a set of socio-technical systems: A multi-organisational view. *Maritime Informatics*. pp. 47-63 (2020)

[27] Henriquez, R., Osés, F. & Marin, J. Technological drivers of seaports' business model innovation: An exploratory case study on the port of Barcelona. *Research In Transportation Business & Management*. **43** pp. 100803 (2022)

[28] Olaya, P., Kennedy, D., Llamas, R., Valera, L., Vargas, R., Lofstead, J. & Taufer, M. Building Trust in Earth Science Findings through Data Traceability and Results Explainability. *IEEE Transactions On Parallel And Distributed Systems*. **34**, 704-717 (2022)

[29] Jaimini, U. & Sheth, A. CausalKG: Causal Knowledge Graph Explainability using interventional and counterfactual reasoning. *IEEE Internet Computing*. **26**, 43-50 (2022)

[30] Yi, X., Paulet, R., Bertino, E., Yi, X., Paulet, R. & Bertino, E. Homomorphic encryption. *Springer*. (2014)

[31] Dwork, C. Differential privacy: A survey of results. *Theory And Applications Of Models Of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings 5*. pp. 1-19 (2008)

[32] Li, L., Fan, Y., Tse, M. & Lin, K. A review of applications in federated learning. *Computers & Industrial Engineering*. **149** pp. 106854 (2020)

[33] Yang, C., Zhang, M., Jiang, Q., Zhang, J., Li, D., Ma, J. & Ren, J. Zero knowledge based client side deduplication for encrypted files of secure cloud storage in smart cities. *Pervasive And Mobile Computing*. **41** pp. 243-258 (2017)

[34] Kerschbaum, F. Privacy-Preserving Computation: (Position Paper). *Privacy Technologies And Policy: First Annual Privacy Forum, APF 2012, Limassol, Cyprus, October 10-11, 2012, Revised Selected Papers 1*. pp. 41-54 (2014)

[35] Schoenke, J., Aschenbruck, N., Interdonato, R., Kanawati, R., Meisener, A., Thierart, F., Vial, G. & Atzmueller, M. Gaia-AgStream: An Explainable AI Platform for Mining Complex Data Streams in Agriculture. *Smart And Sustainable Agriculture: First International Conference, SSA 2021, Virtual Event, June 21-22, 2021, Proceedings 1*. pp. 71-83 (2021)

[36] Kiourtis, A., Mavrogiorgou, A., Kyriazis, D., Graziani, A. & Torelli, F. Improving Health Information Exchange through Wireless Communication Protocols. *2020 16th International Conference On Wireless And Mobile Computing, Networking And Communications (WiMob)*. pp. 32-39 (2020)

[37] Makhdoom, I., Zhou, I., Abolhasan, M., Lipman, J. & Ni, W. PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities. *Computers & Security*. **88** pp. 101653 (2020)

[38] Manogaran, G., Alazab, M., Shakeel, P. & Hsu, C. Blockchain assisted secure data sharing model for Internet of Things based smart industries. *IEEE Transactions On Reliability*. **71**, 348-358 (2021)

[39] Tardieu, H. Role of Gaia-X in the European Data Space Ecosystem. *Designing Data Spaces: The Ecosystem Approach To Competitive Advantage*. pp. 41-59 (2022)