Proceedings of 15th International Marine Design Conference (IMDC-2024)
June 2-6, 2024
Amsterdam, the Netherlands

Knowledge Graphs underpinning ship digital twins for decarbonisation options assessment

Bill Karakostas[1,*] and Antonis Antonopoulos[2]

ABSTRACT

We propose the concept of a Knowledge Graph as a data management and inference machinery that underpins digital twins of ships. The Knowledge Graph is a directed graph connecting dependent and independent model variables of interest in the digital twin, where the correlations between variables are continuously updated based on data received from the physical ship. The paper outlines a methodology for constructing the Knowledge Graph and proposes metrics that help to calculate the effectiveness of decarbonization solutions based on changes to the strength of data correlations. The proposed methodology allows for the extrapolation of decarbonization technology potential across specific vessels, fleets, operational patterns, and lifecycle phases.

KEY WORDS

Digital Twin, Knowledge Graph, Correlation Graph, Ship Decarbonisation, Wind Assisted Propulsion

INTRODUCTION

The digitalisation of shipping to accelerate decarbonisation, supported by data gathered from sensors and ship systems (Agarwala et al, 2022), is a promising approach (Ksetri, 2021). Particularly relevant is *Digital Twin* (DT) technology, initially used by NASA, but applied in recent years to smart manufacturing, transport, smart cities and other areas. DTs enable the creation of a digital representation of a ship, which is fed with data acquired by the physical ship via sensors. These digital representations can then be used to analyse ship-related functions and processes, and actuate systems on the physical ship responsible for engine management, navigation and others.

A ship digital twin therefore, can be viewed as a virtual replica of its physical counterpart that is not static but dynamic, with a bidirectional relationship to the physical system. Use of DT in shipping can result in reducing costs and improving time effectiveness and quality.

A broad definition of a knowledge Graph is as a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities (Hogan et al, 2022). Knowledge graphs are knowledge-based models that utilize graph links to connect entities in a particular domain and to augment the existing knowledge by using queries and other types of inferences. A knowledge graph therefore is ideally suited for modelling knowledge in the domain represented by the physical counterpart of the DT. Moreover, the knowledge graph links data about the physical system as these are collected by the digital twin. Such data types may include:

- geometrical data, e.g. CAD models of the ship
- event based data for events generated by the physical ship throughout its operation, and
- time series data such as speed, fuel consumption, etc, collected from the ship, as well as data such as sea state, direction, collected from the ship's environment.

A knowledge graph can be used as the data engine management, and also as the inference engine of the DT, that acts on the datasets received from the physical ship to make inferences such as predictions of future states of the ship, and also to evaluate

[1] Inlecom Systems, Brussels, Belgium
[2] Konnecta, Dublin, Ireland
* Corresponding Author: bill.karakostas@inlecomsystems.com

Submitted: 29 January 2024, Revised: 25 April 2024, Accepted: 1 May 2024, Published: 23 May 2024

Conference paper, DOI: https://doi.org/10.59490/imdc.2024.871 e-ISSN: 3050-4864

the actual current state of the ship and compare with the expected one, thus supporting anomaly detection (the existence of an abnormal state), and error and fault diagnosis.
This paper builds upon foundational work on digital twin technology and knowledge graphs as tools for supporting the decarbonization of the shipping industry (Antonopoulos et al, 2023). In the research presented in (Antonopoulos et al, 2023), a digital twin of maritime vessels was developed to provide a platform for simulating and evaluating the impacts of various decarbonization technologies. Key to that approach is the use of knowledge graphs for maintaining ontologies, storing simulation models, and correlating these models with a combination of measured and estimated variables.
The novel contribution of this paper is on enriching knowledge graphs with dynamic features by introducing a computational framework that establishes quantifiable links between system parameters. By utilizing measures such as correlation, we enable precise comparisons between data-driven ship models and their theoretical counterparts.
The paper is structured as follows: Initially, the paper outlines the methodological framework within which the proposed technique is applied. The next section provides a formal definition of the Knowledge Graph and illustrates it with an example. Then, a procedure for constructing a Knowledge Graph is proposed. Next, we introduce metrics that are based on statistical analyses of the datasets collected by the digital twin. Again, an example is used to illustrate the various proposed metrics. Following that, a case study is shown where a sample data set and Knowledge Graph is used to analyse and quantify the effectiveness of a hypothetical wind-based decarbonisation technology introduced on a ship. We conclude the paper with an overview and critique of the proposed approach and with suggestions for further research.

ENCODING DOMAIN-SPECIFIC KNOWLEDGE OF DIGITAL TWINS USING KNOWLEDGE GRAPHS

Within the maritime sector, a DT is conceptualized as a real-time digital counterpart of a physical entity, such as a ship, port, or maritime operation [Madusanka et al, 2023]. Specifically, a ship's DT represents an exact digital reflection of the physical vessel, encapsulating its structural design (like hull type and layout, hull parameters), onboard equipment (engines, propellers, rudders), and operational functionalities (propulsion, navigation, cargo handling). This comprehensive digital model integrates technical specifications, component designs, and operational data within a fleet management framework, facilitating the computerized simulation and optimization of ship operations across various performance metrics with a focus on environmental sustainability.
Digital Twins engage in a two-way communication with the ship's Information and Communication Technology (ICT) infrastructure [Tao et al, 2018], utilizing data collection techniques and devices to continuously update the digital model with real-time information from the physical vessel. This dynamic learning process allows the DT to accurately reflect the physical ship's lifecycle, offering insights into its present condition and forecasting future states through simulations and predictive analytics.
A ship DT is built around measurable parameters and can be structured into a network of interconnected modules, such as propulsion systems, voyage and fleet management systems. These modules may also be integrated with broader digital twins that encompass ship operations management or the entire spectrum of ship design, construction, and lifecycle management. The primary goal of employing a ship DT is to facilitate both reactive and predictive decision-making processes, aiming to achieve objectives that include reducing environmental impact. The insights gained from one ship's DT can often be applied to others, enabling data and knowledge sharing in advancing green shipping practices. This prerequisite a degree of interoperability between digital twin instances, which can be achieved by using a common ontology, as suggested by (Hiekata et al, 2010).
Digital Twin technology has found application across various maritime domains, including shipbuilding, offshore oil and gas exploration, marine fisheries, and renewable marine energy production (Zhihan et al, 2023). For instance, (Coraddu et al, 2019), have developed a data-driven digital twin for ships, capitalizing on vast datasets from onboard sensors to estimate speed loss due to marine fouling, showcasing the practical utility of digital twins in enhancing maritime operations.

KNOWLEDGE GRAPH FORMAL DEFINITION

Rationale

Knowledge graphs are sophisticated semantic networks acting as knowledge bases, structured like directed graphs (Qiu et al, 2017). They function by organizing data into triples of (subject, predicate, object) from semi-structured or unstructured sources, thereby enabling advanced knowledge retrieval and reasoning capabilities (Wei et al, 2018). Knowledge graphs are adept at modelling complex relationships within domains, linking disparate pieces of information, and supporting a wide range of applications including knowledge retrieval, question and answer (Q&A) systems, recommendations, and visualization.
Initially developed for extracting knowledge from extensive datasets, knowledge graphs are now a cornerstone in the semantic web, setting a benchmark for efficient information retrieval and usage. Their utility spans various sectors, notably in industry

for tasks such as maintenance planning of sophisticated equipment (Xia et al, 2023), and predictive maintenance for hydraulic systems (Yan et al, 2023).

Within the maritime and shipping sector, knowledge graphs have found applications in analyzing ship collision accident reports to enhance maritime traffic safety. (Zhang et al., 2020), have developed a knowledge graph for maritime dangerous goods, streamlining the knowledge retrieval of hazardous materials, automating the assessment of cargo stowage and segregation, and advancing the intelligent transport of dangerous goods. Similarly, (Langxiong et al, 2023) crafted a Ship Collision Accident Knowledge Graph, aiding in uncovering accident correlations and streamlining the judicial and investigative processes for marine accidents.

Crucially, knowledge graphs have become integral to augmenting the functionality of Digital Twins. By mapping entities, their relationships, and attributes in an organized fashion, they offer a systematic approach to collating and interlinking data within the DT framework. In maritime contexts, they enable comprehensive representations of vessel ecosystems, covering equipment, maintenance histories, and compliance with regulations. These interconnected networks support predictive analytics, risk evaluations, and the simulation of various scenarios, enhancing operational decision-making. By bridging real-time and historical data, knowledge graphs also underpin predictive maintenance strategies, facilitating early detection of potential failures, thus ensuring operational reliability and safety.

A related modelling formalism, *dependency graphs*, have been studied as part of engineering design (Rötzer et al, 2022). There is a variety of such graphs based on their formal underpinning and role in the design/manufacturing cycle. Effect Graphs, for instance are qualitative models built to produce early qualitative statements about the system behaviour. Directed Acyclic Graphs (DAG) or causal diagrams as described by (Pearl, 1995).

The Knowledge Graph we propose is a dependency network of dependent and independent model variables of interest, where the links show the strength of correlation associations between the variables. The Knowledge Graph utilizes the ship data collected by the DT to learn the strength of the correlations between system variables and to represent them as rule nodes. Association rule learning is a rule-based machine leaning method for discovering interesting relations between variables in large databases. (Agrawal et al, 1993). Association rule mining has been studied since the 1990s in domains such as supermarket transactions. However, in such domains there are not many theoretical models to cross validate against the empirically mined association rules. In contrast, the association rule we propose utilizes both theoretical and data driven ship models.

The Knowledge Graph can therefore, be considered as a qualitative statistical summarizer of the ship's models parameters of interest, reflected in the strength of associations between the key variables that represent the system and its environment.

As shown in Figure 1, the knowledge graph models dependency rules such as statistical correlations of the data collected from the physical ship and its environment. The rules describe statistical dependencies (correlations) between independent variables (factors) and dependent variables (factors) of interest. For instance, independent factors include ship speed and direction as well as environment factors such as ship state. Dependent factors include ship fuel consumption and emissions.

The exact relationships between independent and dependent variables are complex and mostly nonlinear. Complexity means that the independent variables may be correlated with each other as well as with the dependent variables

To calculate the strength of the influence of each independent variable on a dependent variable, standard correlation analysis techniques such as Person for linear, and Spearman, for nonlinear correlations, can be employed. For instance, although ship sped is positive related to fuel consumption for very low speeds the polarity of the relation is reversed as in such speed, the friction is increased and requires more power to overcome.

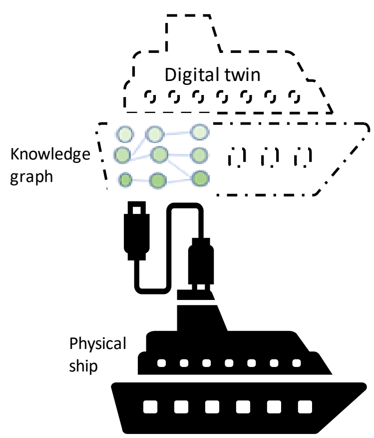

Figure 1: Knowledge Graph in the context if the digital twin

Formal Definitions

Let X be a finite set of independent variables ('factors') $x_1, \ldots x_2, \ldots$ and Y a finite set of dependent variables $y_1, y_2, \ldots$

Each factor x in X draws values from a finite countable set of ordered values called the domain of X and denoted as $Dom(x)$
We use $val(x_i)$ to refer to a value drawn from $Dom(x_i)$
Similarly, each dependent variable y in Y draws values from a finite countable set of ordered values demoted as $Dom(y)$
We define a *record type r* of variables $x_1, \ldots, x_k$ $k > 1$, as a tuple $(val(x_1) \ldots, val(x_k))$ with $k >= 1$ where $x_i \neq x_j$ for all i, j.
We define a set of correlation rules C where $c \in C$ *is a tuple* $(x_1, \ldots, x_n, y)$, and the strength of the rule c $corr(c)$ as a function *corr* returning a value between 0 and 1.
We define a Knowledge Graph as a directed graph $<V,E>$ where $v \in V$ is a variable from $X \cup C \cup Y$ and $e \in E$ is a tuple (x, c) where $x \in X$ and $c \in C$ or $\{c, Y\}$ where $c \in C$ and $y \in Y$.
Informally, factors are connected to rules and rules are connected to dependent variables. A rule $(x_1, \ldots, x_n, y)$, shows the correlation of factors $x_1, \ldots, x_n$ with the dependent variable y, while the weight on the edge between rule and dependent variable corresponds to the strength of the correlation . The correlation strength can be the result of a theoretical formula or an empirical (e.g. regression) formula.

Example

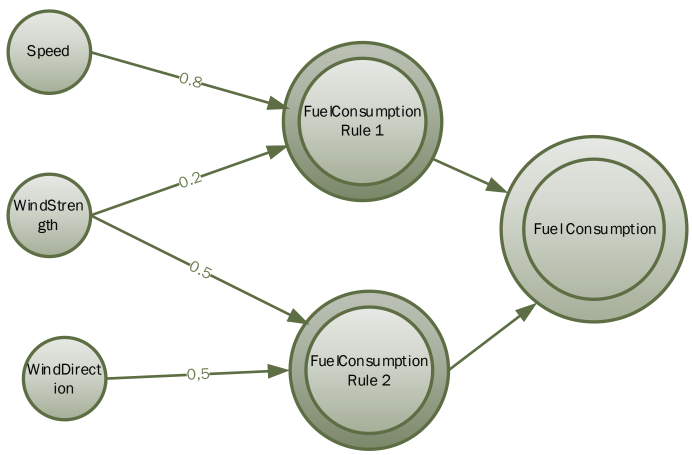

Figure 2: Knowledge Graph for Fuel Consumption

Figure 2 shows a small sample Knowledge Graph. The modelled factors are *Speed*, Wind *Strength* and *Wind Direction*. All these factors are quantized as discussed in the next section, and draw values from their respective domains. For instance, *Wind Strength* draws values from the domain of integers between 0 and 9 following the Douglas Sea Scale units of measurement. The three factors are connected to rule nodes. Consequently, the rule nodes connect to dependent variables of interest, in this case to *Fuel Consumption*. The labels on edges between rules and dependent variables show the strength of the correlation.

PROCEDURE FOR CONSTRUCTING KNOWLEDGE GRAPH

The main steps in constructing a Knowledge Graph are as follows

1. Identify subsystems and components of the ship system.
2. Identify additional systems that the current system interacts with/is part of.
3. Identify model variables (endogenous as well as exogenous) that describe the behaviour/performance of the system/subsystem/component.
4. Define the Knowledge Graph vertices (nodes) in terms of the variables in Step 3.
5. Discretise continuous or numerical variables into categorical variables.
6. Determine dependencies between independent variables and dependent ones using theoretical models, experimental data and expert knowledge/Create one rule vertice per rule in the Knowledge Graph.

Discussion on the proposed Knowledge Graph construction procedure.

In general, quantizing/binning of continuous variable introduces non-linearity and tends to improve the performance of the model. It can be also used to identify missing values or outliers. We quantise the values of the independent factors using natural (e.g. physics dictated, as well as business related criteria, using ordered categorical variables.

Example

Although vessel speed can be modelled as a continuous variable, we discretise it according to speed categorisation that is dictated by physics/ operational and legislation constraints So, speed is discretised according to the categories of normal, slow steaming and extra slow steaming. Table 1 shows examples of quantizing ship and environment variables.

Table 1: quantizing variables

Record id	Avg. Speed (knots)	Speed quantized	Avg wave height (m)	Waves in Douglas Sea Scale	Fuel consumption rate (lt/min)	Fuel consumption rate quantized
T1	22	*n* (normal)	6.5	7	14	*h*(high)
T2	19	*ss* (slow steaming)	3	5	10	*m* (medium)
T3	21	*n* (normal)	2	4	8	*l*(low)
T4	15	*ess* (extra slow steaming)	1	3	6	*vl* (very l(ow)

Metrics

In this section we define metrics related to the numbers, frequency of occurence and correlations in the dataset vailable to the DT and consisting of measurements taken from the physical ship, discretised and grouped into records. We define a record instance of a record *r* an appearance of a record of type *r* in the dataset.

Expected Frequency

Expected frequency *Ef(r) is* a measure of how frequently instances of record of type *r* are expected to appear in the dataset. It is based on the strength of correlation between variables appearing in the record according to the existing knowledge (theoretical and/or empirical) of the domain.

Example

Assuming high (0.9) correlation between ship speed and fuel consumption and the following quantization of speed and fuel consumption:

Dom(*Speed*):{ *ss*,*n*} with *ss* < *n*

Dom(FuelConsumption) : {*l*, *h*} with *l* < *h*.

If speed and fuel consumption were totally uncorrelated we would expect to find all combinations (*ss, l*),(*ss*,*h*),(*n*,*l*),(*n*,*h*) with equal probability of 0.25.

However because of the assumed correlation we expect to find record types (*ss*,*l*),(*n*,*h*) with probability 0.9 and record types (*n*,*l*) , (*ss, h*) with probability 0.1

However, not all combinations are equally represented in the dataset due to measurement limitations. For instance, it may be possible to measure some particular combinations of factors due to technical and physical limitations such as very strong winds co-occurring with high vessel speeds. Therefore, we introduce the concept of *support* below.

Coverage

Coverage for a record of type *r* is the number of record instances in the dataset divided by the number we could theoretically find in a dataset of that size. For example, because of the strength of the correlation between speed and fuel, we could expect on average 10 instances of record (*low speed*, *high fuel*) in a sample of 100 instances, If we find instead 5 instances of the record the coverage is 5/10 or 0.5

When *coverage* < 1 for a record type it means that the record type is *underrepresented* in the dataset while > 1 means the record type is *overrepresented*

Counterrecord

A counterrecord of record type $r{:}(x_1, x_2, .., x_k, y)$ with respect to some correlation rule *c is a record type* $\check{r}{:}(x_1, x_2,..,x_k, y')$ where $y \neq y'$ has expected frequency $\mathbb{E}f(\check{r}) < \mathbb{E}f(r)$

For instance, if the association between speed:(*l,h*) and fuel consumption:(*l,*h) is strong positive (with for example, 0.9 strength) we expect to find record instances of type (*l, l*) and speed: (*h, h*) with high frequency and records of type (*l,h*) and (*h,l*) with lower frequency in the dataset. Thus record types (*l,h*) and (*h,l*) are counterrecords of (*h, h*).

We use $\hat{R}_c$ for the set of all counterecord types of *r* under correlation rule *c*.

Association rule strength

The strength of an association rule *c* for a record of type *r* is defined as

$$\Sigma_{\check{r} \in \hat{R}_c}\left(\mathrm{E}f(\check{r})\right) * \frac{|r|}{|\check{r}| * \mathrm{E}f(r)} \qquad [1]$$

Where $|\check{r}|$ is the cardinality of the set of all counterexamples found in the data set.

The summation of expected frequencies of counterrecords in formula [1] can be explained with this example: If we have 3 types of counter records and we find one instance per counter record and the expected frequency of each counter record is 1/30 and 5 record instances with frequency 0.9 ,we obtain the association rule strength as 3 x 1/30 x (5/3*0.8= 0.13

CASE STUDY

Background

We illustrate our approach with a hypothetical example about the introduction on a ship of a fuel consumption reduction technology that is based on wind assist propulsion (WASP). It is well known today that the use of wind is one of the solutions to substitute existing fossil-based propulsion technologies. The sizing of wind assists devices such as sails and kites needs to take into account the propulsion system. Therefore, vessels equipped with variable pitch provide a greater range of applicability. Moreover, analytical models for kite have been developed (Leloup et al, 2016). Models of the operation of wing sails have calculated the Energy Efficiency Design Index (EEDI) change for specific commercial routes, identifying a potential reduction of 18% (Yong et al, 2919).

Estimating the correlation strengths

The correlation graph of Figure 3 is small as it is only used to illustrate the principles of our approach. A real life scale knowledge graph would also include variables such as fuel consumption/mile, wind direction, wind speed, vessel speed and propeller pitch. The new WASP technology is expected to negatively correlate fuel consumption with wind power, i.e. as the wind power increases, the new technology will utilize it to reduce required engine power, and hence, fuel consumption. The size of our synthetic data is small compared to a real life scenario where the data set could contain thousands or millions of records collected over large time periods, and is used to illustrate the proposed approach. It is envisaged that the Knowledge Graph would periodically carry out the calculations described below as the ship conditions change, on very large datasets that are updated over the ship's lifecycle. The impact of both speed and wind strength on fuel consumption has been investigated both empirically and theoretically. More specifically, the weather a ship encounters during voyage has significant influence on her fuel consumption, in particular relating to prevailing wind and waves (Bialystocki and Konovessis, 2016). Accordingly, prior to introducing the new technology the correlation between wind strength, speed and fuel consumption rate for the particular vessel was calculated experimentally and theoretically as shown in the Knowledge Graph of Figure 3. As per Figure 3, there is a very strong positive correlation between speed and fuel consumption rate (0.9) and a strong (0.7) positive correlation between wind strength and fuel rate consumption.

The factors' corresponding domains after discretization are:
Speed: {(*n*)ormal, (*s*)low steaming. *m*(anouvring)}
Wind Strength: {*0..9*}
Fuel Consumption Rate: {*n*(ormal), *l*(ow), *h*(igh)}

The sample dataset received by the digital twin is shown in Table A1 in appendix A.

To assess the effectiveness of the decarbonisation technology based on the collected data we need to pose several questions and analyse the dataset using the metrics that we defined previously. The different analyses are further discussed below.

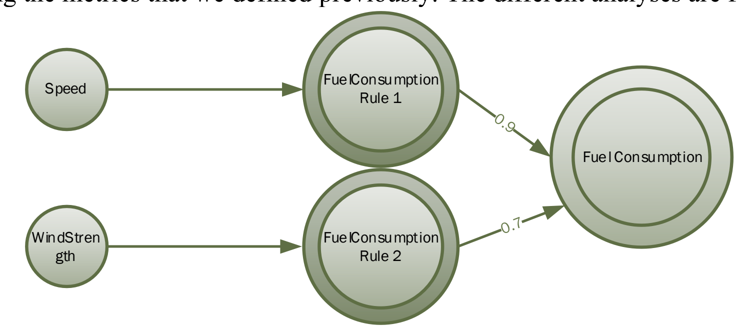

Figure 3: Correlation graph for Speed, Wind Strength and Fuel Consumption

The coverage of the dataset for every speed-wind strength combination

The total number of record types is ||*Dom(Speed)*|| x ||*Dom(WindStrength)*|| i.e. 30. The coverage of the different record types *Speed* x *WindStrength* is shown in Table 2.

Table 2: Coverage of record types

Record type	(n,0),(n,2),(n.3),(n,4),(n,5),(n,6),(n,7),(n,8),(n,9) (ss,0),(ss,1),(ss,2),(ss,5),(ss,6),(ss,7),(ss,8),(ss,9) (m,0),(m,1),(m,2),(m.3),(m,4),(m,5),(m,6),(m,7),(m,8),(m,9)	(n,1), (ss,3), (ss,4)
Expected Coverage	2/3	2/3
Actual Coverage	0	2

The reasons that most record types are underrepresented on this occasion is the small size of our dataset. In real life underrepresentation could be interpreted as caused by:

- Data corresponding to underrepresented record types were difficult to collect due to limitations of the measuring apparatus, Measuring or recording errors
- Due to the rarity of the physical events corresponding to the record type. For instance, sea scales of strength between 3-5 are more commonly experienced in open seas than those of strength 0, 1, 8, 9.

Examine the sensitivity of the new technology to wind strength i.e. whether it works better with higher or lower wind strengths

To analyse that, we compare the frequency of occurrence of record types (S*:*any*, WindSrength:0-4, FuelConsumption:h*), with those of record types (S*:*any*, WindStrength:5-9, FuelConsumption:h*)

According to the dataset of Table 1, the new technology tends to be more strongly associated with low fuel consumption at low wind strengths. Of course, that can be explained that other confounding factors such as ship wind resistance have a stronger effect on fuel consumption than the new technology at high wind speeds.

Identify potential situations where the technology causes increased fuel consumption results

For this type of analysis we need to enumerate all counterrecord types of records where fuel consumption is medium or low and compare the frequency of the record with that of the counterrecords.
Counterecord types of (*m,l*) are (*m,m*),(*m,h*). From the association rule strength we expect to find 2 counterecord instances, however 5 instances where found in the dataset. This means that the new technology can have unintended increase in fuel consumption at maneuvering speeds.
Counterecord types of (*s,m*) are (*s,h*),(*s,l*). From the association rule strength we expect to find 2 counterchord instances, however 4 instances where found in the dataset. This means that the new technology can have unitentended increase in fuel consumption at slow steaming speeds.

However, all above analysis results should be interpreted in the context of the dataset size and coverage of each record type.

Calculate the overall fuel improvement potential of the new technology

We need to test whether the new tech reduces the strength of the association between speed and fuel consumption and increases the strength of the association between wind strength and fuel consumption
By applying equation [1] to the dataset we obtain the revised correlation strengths as shown in Table 4.

Table 3 revised correlation strengths after new decarbonization technology

$\|R_{Speed}\|$	$\|\hat{R}_{speed}\|$	c(Speed→Fuel Consumption)	$\|R_{\text{Wind Strength}}\|$	$\|\hat{R}_{\text{Wind Strength}}\|$	c(Wind→Fuel Consumption)
4	5	~0.11	4	11	~0.65

From the results shown in table 3 we observe that the strength of the correlation between speed and fuel consumption has been reduced from 0.9 to 0.11. This means that speed is no longer the main determinant of fuel consumption. Similarly, the strength of the wind and fuel consumption has increased from 0.1 to ~0.65, meaning that wind strength is now the significant determinant of fuel consumption. This of course assume that all other factors have remained unchanged after the introduction of the new technology.

CONCLUSIONS

In this paper we presented the theoretical foundation of an approach to encode quantified domain-specific knowledge, followed by practical examples, and demonstrate its application in a use case focused on wind-assist technology. While this work primarily explores the methodology's principles and provides a preliminary report, it lays the groundwork for future development and implementation of a holistic decision-support framework aimed at decarbonizing the shipping industry with the use of DTs and knowledge graphs. The proposed Knowledge Graph correlates independent and dependent variables that model the physical system in a digital twin. Knowledge Graphs in general, are an important technology for data representation and knowledge inference in many industrial domains. (Abu-Salih, 2021). The Knowledge Graph connects theoretical and empirical ship knowledge with the data received from the ship throughout its operation.
The purpose of the proposed Knowledge Graph is to compare the theoretical with empirical model of the ship in order to identify discrepancies. Such discrepancies can then be interpreted as anomalies, or as changes to physical ship parameters, and used to compare 'before' and 'after' scenarios. This is also the approach employed in this paper where the Knowledge Graph is used to analyse the impact of decarbonization technologies on the ship's operational parameters. This dynamic characteristic of the knowledge graphs enables the optimization of predictive capabilities for computational efficiency, and facilitates the encoding of knowledge patterns that can be transferred across different instances of digital twins. By leveraging quantifiable associations, we can effectively analyze and quantify the effectiveness of newly introduced decarbonization technologies within an existing ship's system. Such analyses allow for the drawing of insightful conclusions regarding system changes brought about by the integration of new subsystems and their impact on system parameters. Through this novel integration of dynamic knowledge graphs and digital twin technology, our work lays a foundation for a more informed and

effective application of decarbonization technologies across the maritime industry, ensuring a strategic approach to mitigating environmental impact while maintaining operational efficiency.

The research presented in this paper aims to support the development of a knowledge graph for encoding domain-specific knowledge generated by the digital twin. The knowledge graph is not just a static repository of data but a dynamic system that integrates, processes, and standardizes knowledge for easy access and application. The proposed knowledge graphs catalogs simulation outcomes, operational insights, and environmental data, transforming raw data into actionable intelligence.

The encoding process involves several key steps: Firstly, simulation models within the digital twin generate data reflecting the performance and environmental impact of various decarbonization technologies under different operational scenarios. This data, along with measured data, is then contextualized within the knowledge graph, which correlates it with existing operational parameters, environmental conditions, and technology performance metrics. Through semantic tagging and linkage, the knowledge is not only stored but also interconnected in meaningful ways, facilitating complex queries and analysis (Fonseca et al, 2022).

Using ontologies to maintain a shared vocabulary and structure, allows for the consistent interpretation of data across different digital twin instances. This standardization is crucial for enabling the transfer of knowledge between instances, ensuring that insights gained from one vessel or fleet can inform decisions on others, even if they operate under differing conditions.

The ultimate goal of this methodology is to enable the extrapolation of decarbonization technology potential across various maritime contexts. By analyzing data from specific vessels or fleets, our approach can predict the effectiveness of decarbonization technologies in different operational patterns (e.g., speed, route, and cargo type) and throughout different phases of a vessel's lifecycle (e.g., design, mid-life retrofitting, and decommissioning). This predictive capability allows ship owners, operators, and industry stakeholders to make informed decisions on adopting decarbonization technologies, tailored to their specific needs and circumstances.

To conclude, our work seeks to bridge the gap between theoretical decarbonization potential and its practical application, offering a scalable and adaptable tool for accelerating the shipping industry's transition towards a more sustainable future.

It must be emphasized that a knowledge graph is never complete or entirely encompassing the modelling perspectives for a ship. It is instead modelled from the perspective of the stakeholders who use the Knowledge Graph/ digital twin in order to study and understand the physical ship. Also, some types of data that may be of interest may not be represented in the digital twin due to technology limitations or even due to non technical reasons (e.g. confidentiality issues).

Additionally, the increasing utilization of Knowledge Graphs as parts of digital twins raises questions about their quality and robustness. (Abu-Salib, 2021). Both model quality and the quality of the underlying data needs to be present for the inferences and predictions made with the use of the Knowledge Graph to be trustworthy. The size and dynamicity of the data sets handled by the DT requires data quality assurance techniques to be automated and integrated in the overall Knowledge Graph infrastructure. As part of future research, we propose techniques of self-reflection and self-correction to be utilized by the Knowledge Graph in order to always remain a reliable and up to date representation of the physical ship.

DATA ACCESS STATEMENT

No research datasets or repositories were used in this paper.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN WRITING

Statement: During the preparation of this work the author(s) did not use any generative AI and AI-assisted technologies

CONTRIBUTION STATEMENT

Author 1: Conceptualization; data curation,; writing – original draft. **Author 2:** methodology**,** review and editing.

ACKNOWLEDGEMENTS

Research described in this report has received funding from the European Union’s Horizon Europe research and innovation program under grant agreement no. 101056799 (Project ‘DT4GS’)

REFERENCES

1. Abu-Salib , B. (2021). Domain-specific knowledge graphs: A survey. Journal of Network and Computer Applications. 185(1).

2. Agarwala P., et al. (2021). Using digitalisation to achieve decarbonisation in the shipping industry, Journal of International Maritime Safety, Environmental Affairs, and Shipping 5(4).

3. Agrawal, R., Imieliński, T.; Swami, A. (1993). Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93. p. 207.

4. Aggarwal, CC and Philip S. Yu. (1998). A New Framework For Itemset Generation. PODS 98, Symposium on Principles of Database Systems

5. Antonopoulos, A., Karakostas, B., Katsoulakos, T., Mavrakos, A., Tsaousis, T., Zavvos, S. (2023). A Digital Twin Enabled Decision Support Framework for Ship Operational Optimisation Towards Decarbonisation. In: Yang, XS., Sherratt, R.S., Dey, N., Joshi, A. (eds) Proceedings of Eighth International Congress on Information and Communication Technology. ICICT 2023. Lecture Notes in Networks and Systems, vol 694. Springer, Singapore. https://doi.org/10.1007/978-981-99-3091-3_38.

6. Bialystocki, N. and Konovessis, D. (2016). On the Estimation of Ship's Fuel Consumption and Speed Curve: A Statistical Approach March 2016Journal of Ocean Engineering and Science 1(2).

7. Coraddu Andrea, Luca Oneto, Francesco Baldi, Francesca Cipollini, Mehmet Atlar, Stefano Savio. Data-driven ship digital twin for estimating the speed loss caused by the marine fouling. Ocean Engineering. 2019 ; Vol. 186.

8. Fonseca Ícaro Aragão & Henrique Murilo Gaspar (2021) Challenges when creating a cohesive digital twin ship: a data modelling perspective, Ship Technology Research, 68:2, 70-83, DOI: 10.1080/09377255.2020.1815140.

9. Hiekata Kazuo , Hiroyuki Yamato b , Sho Tsujimoto. Ontology based knowledge extraction for shipyard fabrication workshop reports. Expert Systems with Applications 37 (2010) 7380–7386
.

10. Hogan, Aidan, Eva Blomqvist, Michael Cochez, Clau dia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. 2022. Knowledge graphs. ACM Computing Surveys, 54(4):1–37.

11. Kshetri, N. (2021). "The Economics of Digital Twins" IEEE Computer, 54(4): 86-90.

12. Langxiong Gan, Beiyan Ye, Zhiqiu Huang, Yi Xu , Qiaohong Chen, Yaqing Shu. Knowledge graph construction based on ship collision accident reports to improve maritime traffic safety. Ocean & Coastal Management Volume 240, 1 June 2023, 106660

13. Leloup R, K. Roncin, M. Behrel, G. Bles, J.-B. Leroux, C. Jochum, Y. Parlier, A continuous and analytical modeling for kites as auxiliary propulsion devoted to merchant ships, including fuel saving estimation,Renewable Energy, Volume 86, 2016, https://doi.org/10.1016/j.renene.2015.08.036

14. Madusanka, Nuwan Sri, Yijie Fan, Shaolong Yang, and Xianbo Xiang. 2023. "Digital Twin in the Maritime Domain: A Review and Emerging Trends" Journal of Marine Science and Engineering 11, no. 5: 1021. https://doi.org/10.3390/jmse11051021

15. Mauro, F. and Kana, AA. (2022). Digital twin for ship life-cycle: A critical systematic review. Ocean Engineering.

16. Pearl J. (1995). Causal diagrams for empirical research. Biometrika, 82:669–710..

17. Qiu Lirong and Zhang Huili. 2017. Review of development and construction of uyghur knowledge graph. In Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering and IEEE/IFIP International Conference on Embedded and Ubiquitous Computing. Institute of Electrical and Electronics Engineers Inc., 894–897.

18. Rötzer S., Schweigert-Recksiek , S., Thoma, D. and Markus Zimmermann. (2022). Attribute dependency graphs: modelling cause and effect in systems design. Published online by Cambridge University Press.

19. Tao, Fei & Sui, Fangyuan & Liu, Ang & Qi, Qinglin & Zhang, Meng & Song, Boyang & Guo, Zirong & Lu, Stephen & Nee, Andrew. (2018). Digital twin-driven product design framework. International Journal of Production Research. 57. 1-19. 10.1080/00207543.2018.1443229.

20. Wei Yan, Yu Shi, Zengyan Ji, Yuan Sui, Zhenzhen Tian, Wanjing Wang, Qiushi Cao, Intelligent predictive maintenance of hydraulic systems based on virtual knowledge graph, Engineering Applications of Artificial Intelligence, Volume 126, Part A, 2023,106798, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2023.106798

21. Yan, J, Lv, T, Yu, Y. Construction and recommendation of a water affair knowledge graph. Sustainability 2018, 10, 3429.

22. Yong Ma, Huaxiong Bi, Mengqi Hu, Yuanzhou Zheng, Langxiong Gan, Hard sail optimization and energy efficiency enhancement for sail-assisted vessel, Ocean Engineering, Volume 173, 2019, https://doi.org/10.1016/j.oceaneng.2019.01.026.

23. Zhang Yimeng, Yuanqiao Wen, Fan Zhang, Chunhui Zhou, Lei Du, Liang Huang,Changshi Xiao. Semantic model of ship behaviour based on ontology engineering. The 2nd 2018 Asian Conference on Artificial Intelligence Technology(ACAIT 2018).

24. Zhang Qi, Yuan Q Wen, Dong Han, Fan Zhang, Chang S Xiao. Construction of knowledge graph of maritime dangerous goods based on IMDG code. The 3rd Asian Conference on Artificial Intelligence Technology (ACAIT 2019) 22 May 2020 https://doi.org/10.1049/joe.2019.1147

25. Zhihan Lv, Elena Fersman, 2022, Digital Twins: Basics and Applications, Springer Cham, https://doi.org/10.1007/978-3-031-11401-4

APPENDIX A

Table A.1: Sample dataset

Record id	Speed	Wind Strength	Fuel consumption rate
T1	n	1	n
T2	n	2	n
T3	n	5	l
T4	s	2	l
T5	s	3	n
T6	s	4	n
T7	s	3	l
T8	m	8	h

T9	m	6	h
T10	m	4	h
T11	n	1	n
T12	s	4	l
T13	s	7	n
T14	s	6	h
T15	n	8	h
T16	n	3	h
T17	m	7	l
T18	m	3	h
T19	n	4	l
T20	n	7	h