On Enabling the Implementation of a Ship-wide Data Ecosystem

CIMAC Digitalization Strategy Group

This publication is for guidance and gives an overview regarding Enabling the Implementation of a Ship-wide Data Ecosystem. The publication and its contents have been provided for informational purposes only. CIMAC makes no representations or warranties express or implied, regarding the accuracy, adequacy, reasonableness or completeness of the information, assumptions or analysis contained herein or in any supplemental materials, and CIMAC accepts no liability in connection therewith.

The first edition of this Position Paper was approved by the members of the CIMAC Strategy Group Digitalization via written correspondence by November 21, 2023 and by CIMAC Council via written correspondence by December 20, 2023.

Content

1 Motivation

Keeping operational costs to the minimum possible is crucial to any profitable business, and the marine transport industry is no exception. Additionally, the use of carbon neutral synthetic fuels to help demonstrate compliance with the decarbonisation targets set by the Regulatory Authorities (e.g. IMO) will inevitably lead to an increase in operational costs due to higher fuel prices, an expected increase in maintenance costs due to the increased complexity of the fuel system needed to handle such fuels in a safe and reliable way.

It follows that, to remain competitive, ship operators/companies will experience increased pressure to achieve significantly higher levels of optimisation for the entire maritime shipping process than is currently undertaken. Ships have been historically designed with a strong focus on CAPEX but the introduction of measures that aim to reduce the GHG emissions from ships such as EEDI [1], EEXI, and CII [2] increases the focus on delivering vessel designs with fully optimised technical solutions, i.e. designs that perform their function in the most efficient way and provide a means of collecting data to validate their efficiency levels throughout their lifetime.

In this respect, digitalization provides the opportunity to generate optimized technical solutions based on highly integrated intelligent systems, here intended as systems capable of creating added value by continuously monitoring their operational status for assessing their integrity, efficiency and reliability through advanced instrumentation and data analytics. While such solutions have been realised in several industries such as automotive and aerospace, they are implemented only slowly in the maritime industry due to several reasons:

- The requirement to deliver an integrated system that provides a ship owner with an optimised technical solution is not prescribed by the current regulatory instruments (e.g., SOLAS [3]).
- The maritime industry is far less system integrator-driven compared to other industries (e.g., aviation industry DO-178C [4], automotive ISO 26262 [5]).
- There is a high fragmentation in terms of number of system integrators and subsystem suppliers.

Consequently, the full technical potential of highly integrated, optimized technical solutions that support the reduction of operational costs (e.g. minimising fuel consumption for a given voyage, optimising the fuel changeover process so as to limit the usage of costly fuel to the minimum, minimising vessel downtime through predictive maintenance, enabling remote classification society activities so that surveyors do not need to go on board, reducing insurance fees through remote system monitoring and predictive maintenance), identification of hazards, and better transparency on environmental performance is currently not exploited.

As an example, there currently exist various monitoring solutions from sub-suppliers in the market, each with their own protocols for data production, but common, open and secure ways for exchanging data among them have not been generally identified and adopted.

This lack of adequate standards and best practices has resulted in a fragmented approach to realizing the potential benefits that could be achieved through an integrated, ship-wide data ecosystem[1] solution that would enable the whole maritime infrastructure to operate as a highly integrated digital system (see Figure 1, which shows a simplified model of the digital system on the

[1] *A ship-wide data ecosystem is intended here as the complex environment of co-dependent networks and actors on-board and on-shore that contribute to data collection, transfer and use.*

ship (shown in the speech bubble) and the connection of the ships systems with onshore data centres).

Furthermore, the adoption of proprietary protocols and data processing hardware has raised concerns about the costs and difficulties of installing and maintaining several proprietary hardware/software solutions by the end user, who sees digitalisation as a sure source of cost and unsure source of revenue, and concerns about effective protection of know-how and intellectual property of each subsystem vendor, who prefers to develop siloed vertical solutions to avoid the risk of intellectual property loss while interfacing with other parties. In both cases, this results in digitalisation not being leveraged to provide wholly optimised systems.

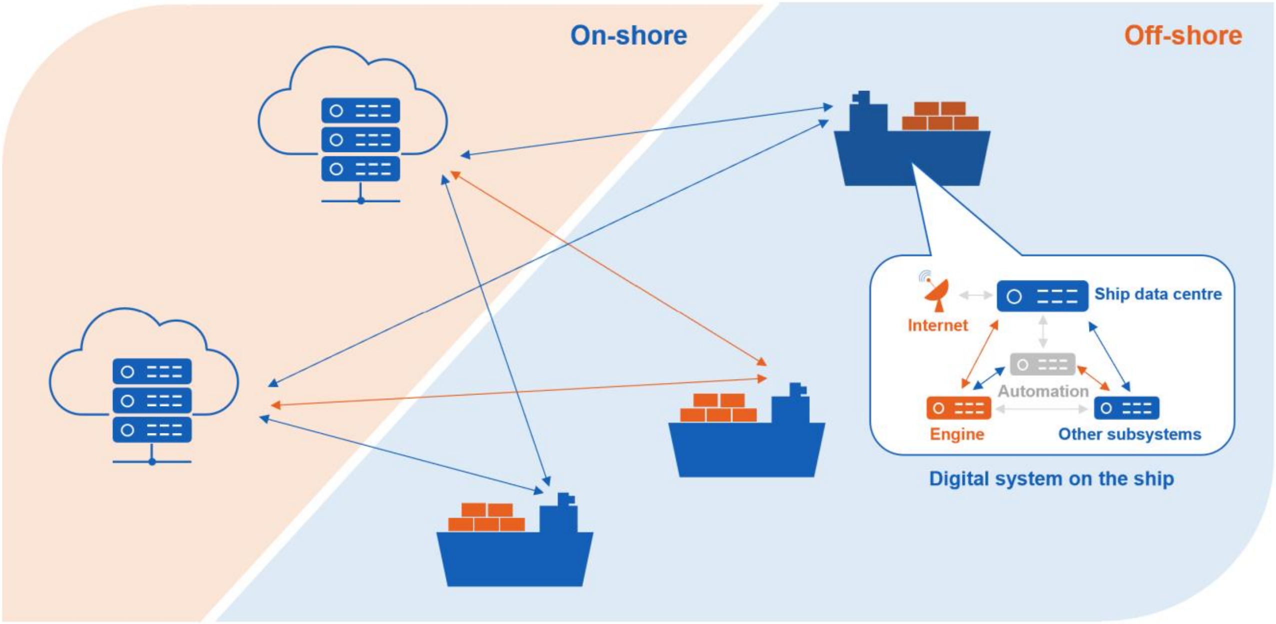


Figure 1 –Maritime infrastructure operated as a highly integrated digital system ©CIMAC

It is therefore desirable to achieve an improved situation in which digitalization can be employed in the best possible way to achieve business benefits (e.g., lower system cost/complexity and higher system efficiency) for all identified stakeholders (e.g. manufacturers, owners, etc.), while at the same time compliance with regulatory bodies, as well as with Classification Society requirements is demonstrated.

To realize the latter in particular, it is key that the output from a highly integrated system provides a structured argument, supported by a body of evidence that, without bias, makes a case for both the ‘safe and ‘unsafe’ characteristics of a system when used for a given application in a given environment. In other words, a specific focus must be placed not only on which overall improvements are conceivable, but also which risks come with a highly integrated system and how they can be adequately addressed.

The first step to achieve an output that satisfies the above, however, will require the industry to enable the release and transfer of data among subsystems well beyond the current minimum necessary to comply with the current rules and regulations, so that globally optimised solutions can be realised, and defined goals can be achieved. To address this, the paper proposes the adoption of a common methodology for data sharing which can be used as the foundation for building digital systems that provide business benefits, improve efficiencies and help demonstration of regulatory compliance of a highly integrated digital system such as the ship, through the implementation of a ship-wide data ecosystem. The focus is here placed on the vessel, because it is the element of the maritime ecosystem where the greatest challenges to enable reliable data sharing exist. However, the approach here presented can be easily extended to the digital equipment on shore that, together with the vessels, makes up the maritime digital ecosystem.

The next chapters present CIMAC's vision and propose possible implementation pathways for a digital data ecosystem. This ecosystem would leverage data and insights from intelligent subsystems, shared at the discretion of the equipment vendors, for monitoring and process optimisation purposes. It is worth emphasising that the approach here presented enables the reliable transmission between stakeholders of all sorts of data for all sorts of purposes. However, once a common methodology for data sharing has been established and adopted, different layers will need to be defined to regulate the process of data utilisation depending on the function to be implemented. In the next chapters, reference is mainly made to monitoring and advisory purposes. Extension of optimisation potential to control and safety systems will be dealt with at a later stage due to safety and compliance implications that will require dedicated legislation, and thus is not in the focus of this paper.

2 Vision

Any approach aimed at fully leveraging the potential of digitalisation for process optimisation and overcoming the limiting factors outlined in section 1 needs to address the following key aspects:

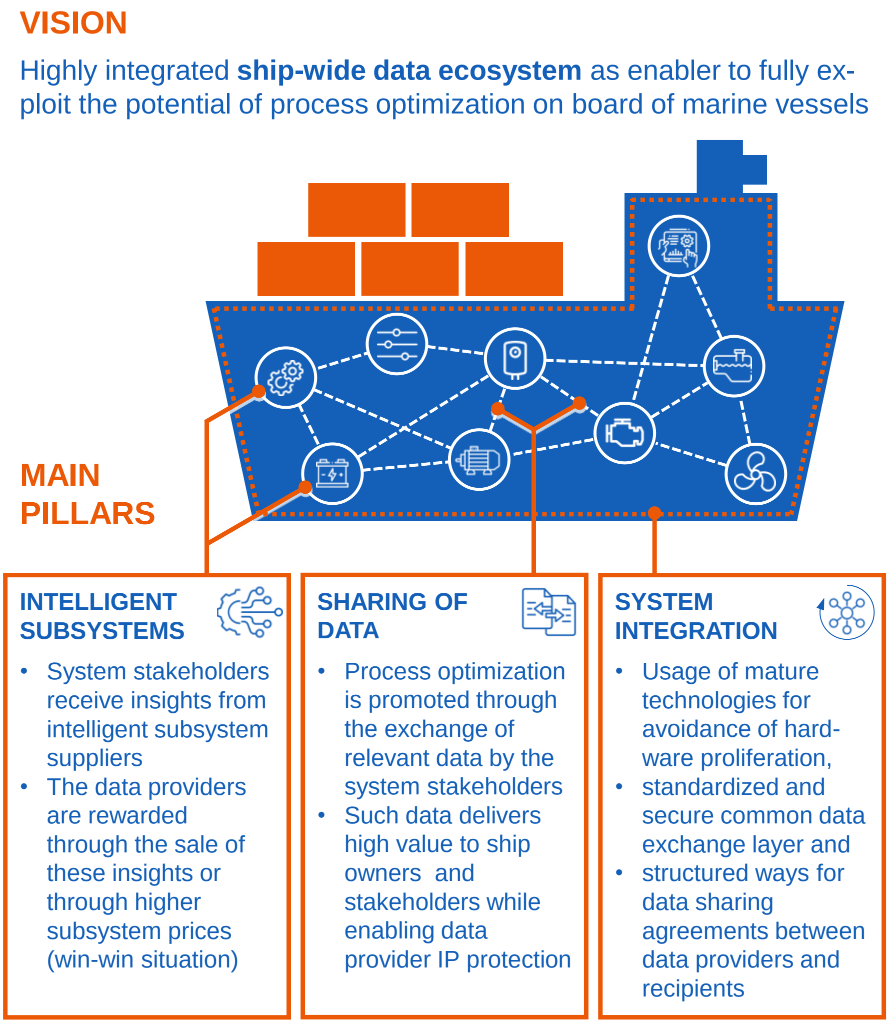
- Protection of IP
- Business model freedom
- Secure data exchange between authorised parties
- Low implementation cost and operation overhead
- Vendor neutrality

As a result, CIMAC's vision of a ship-wide data ecosystem that addresses the above issues is presented in Figure 2. It relies on seamless integration of intelligent subsystems that share data according to standard data exchange technologies.

Intelligent subsystems

The first main pillar to support the ship-wide data ecosystem concept illustrated in Figure 2 relates to intelligent subsystems. In order to reach a global optimisation of the shipping process, i.e. going from port A to port B with the minimum OPEX (e.g. minimising fuel consumption and emissions, avoiding unscheduled maintenance, fully exploiting the available lifetime of each component etc.) while complying with all regulatory requirements, it is beneficial to add intelligence to all the main subsystems that make up the ship ecosystem, so that insights from subsystems can be made available to overall system optimisation algorithms, as well as information regarding functional anomalies and residual useful life of the major components.

Providing valuable insights into the operation and status of subsystems requires an investment on the part of the subsystem suppliers to add sensors to their components and to develop software in accordance with a defined standard that makes use of their specific domain knowledge. Such investment can only be promoted if the vendor can turn it in a source of additional revenue. As a recent example, the EU Data Act [6] offers a perspective for subsystem suppliers to be rewarded for their effort in providing insights. While this regulation imposes the obligation on data holders (e.g., subsystem manufacturers) to make data readily available to users (e.g., owners) of connected products (see also “Definitions” in section 5.1), information inferred or derived from such data, which is the outcome of additional investments into assigning values or insights from the data, in particular by means of proprietary, complex algorithms, including those that are a part of proprietary software, should not be subject to the obligation of a data holder to make it available to a user or a data recipient, unless otherwise agreed between the user and the data holder [6].



VISION

Highly integrated **ship-wide data ecosystem** as enabler to fully exploit the potential of process optimization on board of marine vessels

MAIN PILLARS

INTELLIGENT SUBSYSTEMS

- System stakeholders receive insights from intelligent subsystem suppliers
- The data providers are rewarded through the sale of these insights or through higher subsystem prices (win-win situation)

SHARING OF DATA

- Process optimization is promoted through the exchange of relevant data by the system stakeholders
- Such data delivers high value to ship owners and stakeholders while enabling data provider IP protection

SYSTEM INTEGRATION

- Usage of mature technologies for avoidance of hardware proliferation,
- standardized and secure common data exchange layer and
- structured ways for data sharing agreements between data providers and recipients

Figure 2 – CIMAC vision of a ship-wide data ecosystem for shipping process optimization, relying on three main pillars ©CIMAC

Sharing of data

The second main pillar in CIMAC’s vision concerns the sharing of data useful for exploiting the global optimisation potential of the shipping process. While all sorts of data for all sorts of purposes can be shared, the type of data that provides valuable insight useful for optimisation of ship operations is key. Insight generation is closely linked to domain knowledge. Sharing insights, rather than the data used to generate them, allows the subsystem provider to increase the value of its products while protecting its own IP and know-how, and enables the user to get data useful for process optimisation, thus generating a mutually beneficial situation.

A typical win-win solution would be for subsystem suppliers to share insights in exchange for some form of compensation, for example by charging higher prices for supplying intelligent subsystems (CAPEX-oriented business model), or through a premium service fee paid by the user to receive the insights generated by interpreting sensor data through the manufacturer's domain knowledge (OPEX-oriented business model). Another business case could be to exchange data between subsystem suppliers to enhance features at subsystem level (e.g., subsystem condition monitoring and maintenance solutions) with additional data sources. For such feature enhancement, the user (e.g., shipowner) may also become a valuable data provider. Such win-win situations would boost the development of ship-wide digital ecosystems, providing benefits for the owners to optimise their operations and incentives for the subsystem suppliers to innovate, while maintaining the IP for generating insights into their products.

To achieve efficient data sharing in a ship-wide digital ecosystem, standardised and secure data exchange with low implementation cost and operation overhead must be targeted. Furthermore, conformity with regulations such as the EU Data Act [6] needs to be achieved (e.g. the user of a connected product has the right to access any readily available product data and related service data, including metadata, see also previous section "Intelligent subsystems").

System integration

The third pillar supporting the vision of a ship-wide data ecosystem is related to system integration and, in particular, it focuses on the rationalisation of hardware resources and the definition of a common way of interfacing the various data producers and consumers. It is easy to see how, in a future where every subsystem becomes intelligent, and therefore carries with it is dedicated data processing hardware and custom user interface, the ship operator would be forced to install a multitude of computers, each one with their software and operating system to maintain, and to use multiple graphical interfaces to collect all the relevant insight.

This scenario is clearly not optimal, and presents a barrier to the adoption of advanced monitoring solutions. Hence, the recommended way forward is to create a ship-wide data ecosystem using existing technologies that avoid hardware proliferation by allowing monitoring software of different vendors to operate on the same machine and to define structured ways for managing data sharing between producers and consumers.

In this way, monitoring and optimisation applications could collect data from all producers and present it on a coherent, single user interface together with recommendations for process improvement, making life easier for the operators, and reducing the costs that each vendor nowadays needs to sustain to develop vertical applications spanning from raw data collection to insight presentation. This approach would allow every stakeholder to focus on their specific expertise, so that resources can be dedicated fully to the value creation steps and not to creating duplicates (e.g. the individual user interfaces) that carry little added value and make life more difficult for the operators.

3 Building a ship-wide data ecosystem

The vision presented in Chapter 2 can be implemented in practice in several ways. In this section, a ship-wide data processing and exchange architecture based on reliable, standard, open and proven technologies, which enables data exchange among ship subsystems and between ship and shore is proposed.

It addresses the two fundamental tasks of data processing (here intended collectively as collection, storage, analysis and transmission) and data exchange aiming to ensure minimum infrastructure and operation costs and, at the same time, to guarantee protection of data property and flexibility in the business model for all stakeholders that are involved in building and utilizing the ship-wide data ecosystem.

3.1 Recommended architecture

The recommended architecture required to build a ship-wide data ecosystem is shown in Figure 3. This arrangement implements the three vision pillars (see Figure 2), in which ***intelligent subsystems*** are ***integrated*** following a common approach to data exchange in which each subsystem monitoring software communicates towards the field to collect data and towards the other modules and the internet to ***share data***.

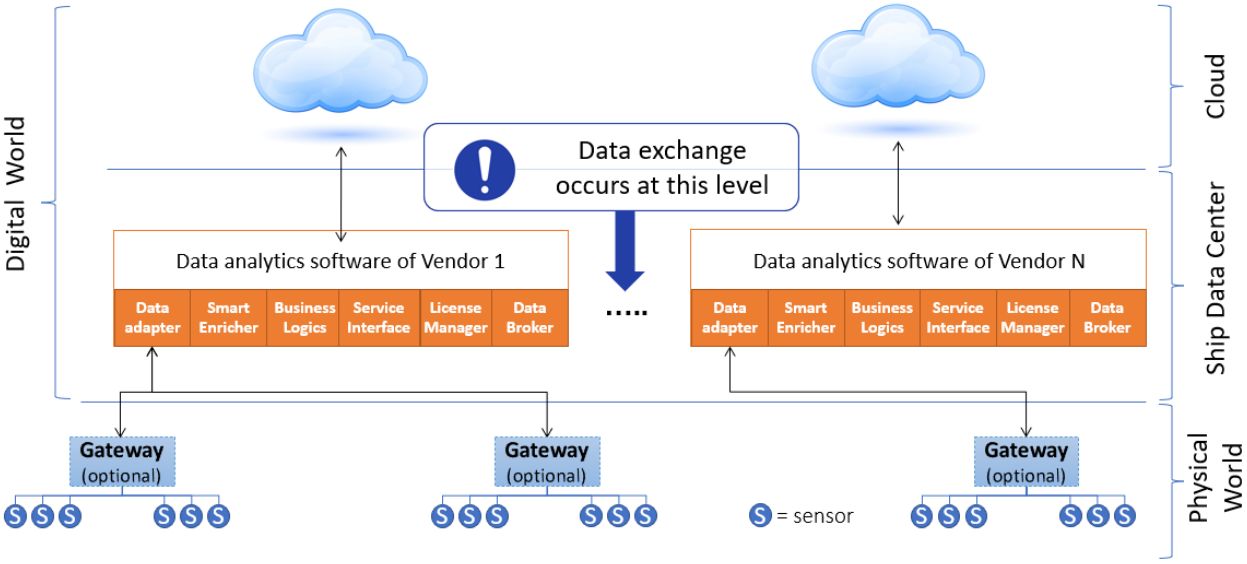


Figure 3 – Proposed architecture for the implementation of a ship-wide data ecosystem ©CIMAC

Each subsystem monitoring software is made up of various modules that perform value-adding data processing operations and manage data accessibility. Generally speaking, a software designed to operate within this architecture could implement the following logical modules:

- **Data Adapter**
  Takes care of the bidirectional communication between the digital and physical world, e.g. providing the right interfaces to collect data from the field and making available local data storage.
- **Smart Enricher**
  This is where vendor know-how is embedded in algorithms that extract precious information from raw data.
- **Business Logics**
  Applies automatic logics to the enriched data, generate notifications for people and/or other systems, and present the data via dashboards
- **Service Interface**

Synchronizes logics and data with the cloud, keeps the software and firmware updated, lets the users see the enriched data remotely.

- **License Manager**
  Enables/Disables specific features based on the business model decided by the vendor. Authorizes other systems to allow or deny access to the data on the ship
- **Data Broker**
  A module that, coupled with the “License Manager”, implements all the technology required to expose encrypted data to other systems on the ship

It is easy to see how, with this data exchange architecture, it is possible to overcome the current situation in which data is kept in “siloes” that do not easily communicate with each other. In this configuration, the IP of each vendor is kept within each monitoring software, and a structured way to exchange data in compliance with specific license agreements is provided. The data received by each monitoring software from other stakeholders is internally stored in its data adapter, eliminating the need to create a storage entity managed by a third party to collect all the data exchanged on the vessel. This approach potentially increases the need for storage space, as the same data received by multiple entities is stored in their respective space, but it avoids the need to establish three-party license agreements between data producer, consumer, and operator of the main storage.

In order to effectively and efficiently implement such architecture, a standard way must be provided for the monitoring software and the hardware it requires to run to be easily integrated with the rest of the data ecosystem. Therefore, the issues of

- Avoiding processing hardware proliferation
- Defining a framework and software development kit (SDK) for standardising the data exchange process

need to be addressed. The next sections present possible state-of-the-art solutions to implement the principles discussed above.

3.2 Data processing model

The amount of available raw data is expected to increase tremendously in the upcoming years; however, raw data in itself has limited usage. In order to provide added value, it must be collected, stored, and distilled by data processing algorithms that make use of the domain knowledge of each subsystem vendor to extract valuable insight from it. Finally, the results must be presented to the user to support its decisions.

In the absence of a ship-wide architecture for data processing and secure exchange, each vendor is forced to develop vertical solutions that span from data collection to the presentation of the results to end user. Such solutions typically require dedicated data processing hardware and user interface. It follows that while, gradually, every vendor adds intelligence to its products by providing condition monitoring systems, the ship operators face an ever-increasing number of proprietary and separate hardware and software solutions to be installed and maintained, and the need to jump from one user interface to another to gather all the needed operation data.

Such non-functional redundancy is clearly a source of additional cost and complexity with no benefit for the operator. Even if moving from the current situation will take time, it makes sense to try to envision a future in which a standardisation of the software modules of each subsystem is performed in terms of control model and data analysis, so that they can all operate as first-class citizens within the ecosystem, and can react in unison even if operating on separate hardware modules.

Considering that the added value of a digital service is provided by the software and not by the hardware it runs onto, the data ecosystem model here proposed goes beyond the above and foresees the adoption of a single on-board data centre to perform all the data processing tasks required by each single vendor. Of course, such data centre should have all the characteristics of redundancy, safety, cybersecurity, back-up etc. typical of modern server installations to avoid creating a single point of failure, and reducing the need for IT-savvy personnel on board a vessel, but a single machine with a single operating system would be easier to maintain by the ship operators than a multitude of proprietary machines, each coming with its own hardware, operating system and application version, and the overall hardware cost would be lower, as shared hardware resources can be better utilised.

However, such arrangement could raise the question of how to safely and securely operate and maintain different vendor applications running on the same machine. Virtual machines provide a way to encapsulate operating system and application software, so that a single computer could run different vendor software designed for different target operating systems and versions. On the other hand, running multiple virtualised operating systems generates a significant overhead and multiplies the cyber-risks due to the potential presence of out-of-date components, and the need to maintain and update the operating system provided by each vendor within each virtual machine.

A step forward in this respect would be the adoption of a proven technology already widely utilised on general purpose data centres and by most cloud service providers, which is that of software containers. Much like “hardware” containers are stored on a ship using standard interfaces and transported and operated without knowing anything about their content, software containers are self-sufficient pieces of software that include the application programs and all the libraries they need to function, and that operate in their own isolated, private space. Hardware resources such as disk space, CPU or RAM are allocated to each container by a container management software (see Figure 4), and freed when they are not needed. The container management engine/orchestrator ensures that any type of policy can be applied (e.g. access, security, network) [7], and virtualises concepts like networking and service discovery, with a unique and homogeneous component for the definition and enforcement of the policies, thereby allowing the creation of an ecosystem that standardises and limits how applications are deployed and connected, enhancing cyber-security and simplifying the container lifecycle management.

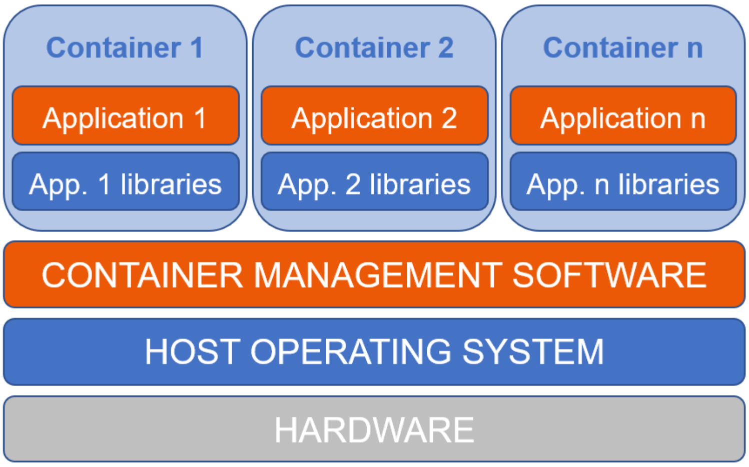


Figure 4 – Architecture of data centre operating software containers © CIMAC

Installation of a containerised vendor software only requires copying a few files in the right location, and the failure of a containerised application does not affect the rest of the system: it simply causes the container software to be rebooted. The rollback to a previous version of a vendor application is also simplified through the use of containers. Of course, an agreed process for managing the installation of container upgrades, taking into account existing standards, i.e. [8], needs to be defined to ensure that new software versions are introduced and tested when the service they provide is not required by the user or other applications, e.g., during port calls. It has been argued that small updates are more easily applicable to virtual machines over a satellite connection than containers, as the former do not require the whole image to be transferred over a limited bandwidth connection. However, it is also possible to upload only the part of container image that has changed, as described in [9].

With containers, the ship operator does not have to deal with the management and update of all the libraries needed by every vendor application whenever a new software version is released, because the correct libraries are already embedded in the container by the vendor. In this way, the vendor maintains the responsibility of providing a working containerised image of its application, reducing the maintenance burden on the ship operator. Hence, the ship operator only needs to keep the data centre up and running, and the different vendors can provide their own version of containerised software, knowing that it will work regardless of the actual configuration of the machine operating as the ship data centre, as long as sufficient hardware resources are available on. This approach provides a clear boundary between the responsibility of maintaining the operating system of the centralised machine (ship operator) and that of maintaining the monitoring applications (subsystem vendor)

It is easy to see how such technology answers the need of avoiding hardware proliferation and the associated costs and risks while ensuring that multiple vendors can share the same hardware resources while protecting their domain knowledge and intellectual property, which is safely compiled into a container that cannot be accessed by third party software. Even if the full implementation of such a model can take time due to the need to evolve from several vertical, vendor managed hardware platforms to a solution with a single data centre managed by the ship operator, developing already containerised vendor applications running on proprietary hardware will enable an easier and cheaper transition to the fully integrated model here presented.

Of course the skills of the operator need to evolve, to manage such an environment. Where a lack of IT competencies exist this can be addressed by defining a specific framework within the SDK, which can put limits on container images size, create a sort of optimization of the deployment package in order to transfer only the business logic (a small set of common target environments, such as container layers, shared among the framework users can be defined), etc. The framework should have a common application layer for observability and operations. DevOps should be capable of remotely obtaining a terminal based connection from the common platform so as to provide support to the on-board personnel.

3.3 Data exchange model

Building on the assurance that no unauthorised access to proprietary vendor software and the data and know-how it contains is allowed, this section deals with the issue of how to exchange data in a controlled way between authorised parties. As each vendor data is stored in the related container, a part of the container software should be tasked with the transmission of data to the outside world. Containers can “talk” to each other using standard network protocols, just as if they were running on different machines, as well as connecting to remote machines (e.g. cloud storage servers and other physical machines on board) using the existing network infrastructure, for example for backing up the data collected on board and free local disk space, or for exposing remote user interfaces.

The data exchange model here proposed foresees the implementation, on each container, of a data broker that is tasked with exposing the data model of the physical object it monitors and implementing the relevant communication protocols. Requests can then be issued by data consumer applications to the various data brokers, and data is transmitted by the broker to the consumer after having established that a license exists between producer and consumer. This standardisation is beneficial for both data producers and consumers. The former would have to implement a single data interface on their systems, without worrying about the interface of the data consumer, and it would be easy for a ship-wide monitoring and advisory application, or any other device connected to the network, to poll all the data brokers on the network and receive information about the data available. In this way, each vendor application would be able to build and maintain its own catalogue of the data being shared in that moment by all connected devices without prior knowledge of their existence and internal data structure.

To release the full potential of this approach, it is necessary to agree on a common way to expose the model of the data that each device can exchange on board a ship and between ship and shore, and the associated licensing model. CIMAC promotes and supports the creation of a software development kit (SDK) that the industry could use for the implementation of a (i) control layer, which takes care of exposing the data model of the specific object (i.e. which data is available, in which format) and the associated licensing, and a (ii) data layer, which implements the protocols for data transmission, because such common resource would facilitate interconnectivity on board while reducing development effort. The next sections present a possible solution, laid out according to the principles of the MACH architecture [10].

3.3.1 Control layer

In a first phase of adoption of such common approach it is expected that all the relevant stakeholders would want to see quick results, maybe via the implementation of Proof of Concept (PoC) implementations, with minimum effort and maximum benefit, possibly reusing parts of existing code. Hence, it is important to ensure maximum interoperability with value generation right from the start.

In this sense, it seems ideal to implement the control layer by making use of Representational State Transfer (REST) Application Programming Interfaces (APIs) [11][12], intended as HyperText Transfer Protocol (potentially with secure implementation; HTTP/S) [13] requests to a server-type endpoint (implemented in the data broker of Figure 3) by a client device (the container requesting the data), because the protocol is already implemented in any device with internet connectivity, and many APIs can be derived from those already implemented by each vendor for proprietary configuration, setup and connectivity to user interface, IoT systems etc. Hence, by making use of most of what is already available in proprietary solutions, this approach minimises the effort needed to make each vertical application suitable to be interfaced with third party ones, and hence to become part of the ship-wide data ecosystem here presented. Furthermore, in case of the development of a new intelligent system, the vendor could use this approach both for internal intercommunication between its software modules, and for external communication to third party ones, thus reducing development cost.

This approach is a widely adopted, de-facto, standard in the internet world, used by the main cloud providers to configure the environments in which data will be exchanged through specific protocols, and to connect web-based user interfaces with back-end applications, which makes it a good choice also for the marine sector. By defining a standard way of describing APIs and data models, the maritime industry players could quickly define a common approach to implement their data brokers that securely exchange data among stakeholders. APIs can be specified according to existing standards, such as OpenAPI [14], which bring the advantage of describing their structure and syntax regardless of the programming language they are written in. Typically, this uses human readable formats such as JSON [15] or YAML [16], and fully defines the services that an application provides through its APIs. In this way, it is easy for a third-party application to poll a data broker and receive in response the data model that it exposes.

Following the same approach, it is easy for two applications to exchange license keys via specific APIs to certify that the client is authorised to receive the requested data, and therefore the server can provide it according to the service level agreement associated with that license.

Figure 5 shows an example of the API primitives and the calling sequence that could be implemented to manage discovery, licensing and data exchange. A client seeking data would first call the “DecentralisedServiceDiscovery” API, which could make use of standard ways of discovering all the subsystems present on the local network, such as mDNS, SSDP, CoAP etc.

The next step would be for the client to authenticate itself with each subsystem and to obtain permission to access its data. A very common approach is to use the Public Key Infrastructure (PKI) [17] which makes use of keys and certificates to securely identifying the actors on the network and therefore ensuring that that specific client is entitled to receive a certain dataset. In such arrangement, the vendor issues a signed certificate for the client, which is sent to the subsystem through a call to the “LicenseValidation” API, and the subsystem validates the certificate through the use of its own public key. In this way, the system can work even in the absence of cloud connection. However, connection to the cloud is required in case the certificate needs to be invalidated (e.g. the license has expired). An alternative way to manage licensing would be to implement a decentralised license registry through the use of blockchain technology [18]. This creates a secure and immutable ledger distributed on the local network where all records related to licensing (terms, permissions, and data access rules) could be stored, providing the additional advantage of redundancy (because the ledger is replicated on all the machines on the network, any single failure would keep the ledger accessible), and avoiding the need to involve PKI certification authorities.

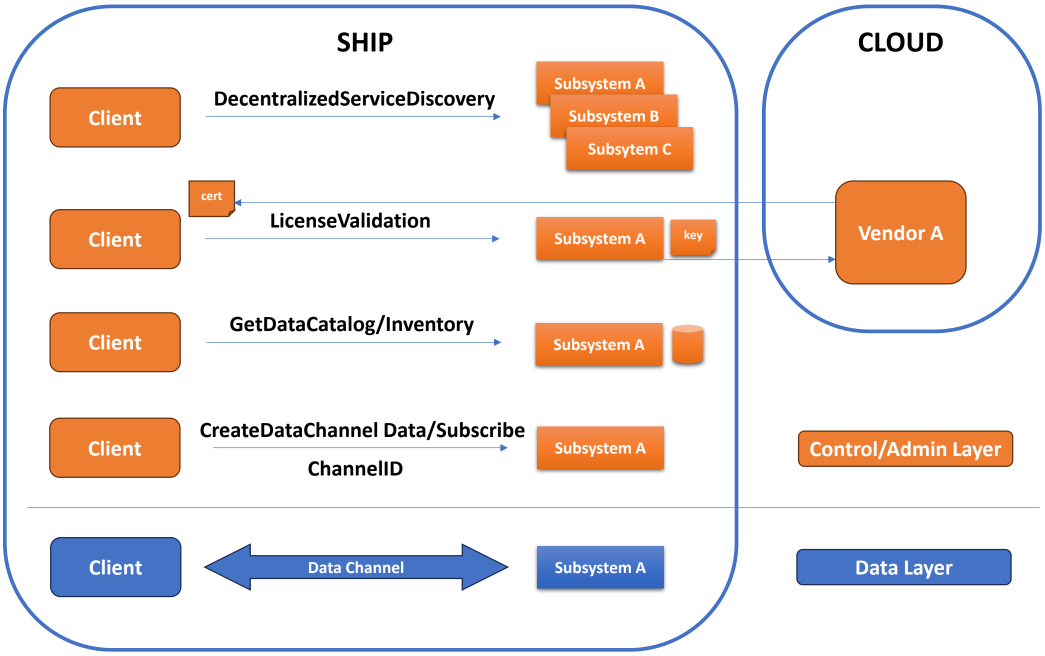


Figure 5 – Example of usage of control layer APIs for discovery, licensing, and data exchange

Once authentication and licensing have been verified, the client can request the data catalog/inventory to that particular subsystem through the use of the “GetDataCatalog” API. The data catalog [19] describes which data the subsystem exposes, and how to access it, and it is key to ensuring reliable and simple data exchange operations. This structure must be accessible even offline and remotely, to simplify the process of integrating each subsystem in the ship-wide data ecosystem already at design level.

At runtime, it is possible that not all the data specified in the catalog are actually available. Here the concept of inventory was introduced to indicate the subset of the catalog that is actually available in any given moment. Several standards exist for specifying data catalogs, i.e. Onem2m, lightweightM2M, Data Distribution Service, Web of Things etc., which simplify the process of catalog creation by allowing to focus on the content, rather than the implementation.

Once the data catalog has been received by the client, so the subsystem capabilities are known, the “CreateDataChannel” API represents the fundamental step to be able to receive the data. The client asks permission to access a specific data, and the server on the subsystem answers positively after successful verification of the license and access policy, indicating on which channel ID the data will be available, with which protocol to access it, and the channel properties (i.e. sample rate, time window, measure unit, past time histories etc.). Potential cryptographic keys, tokens and other parameters, besides IP address and port, required to create the channel must be transmitted by the server at this stage. This completes the set of administration and control operations made in the control layer to open a secure data channel between data provider and consumer. It is easy to see how this approach combines the benefits of controlled data sharing and protection of intellectual property with those of ease of integration of each subsystem in the ship-wide data ecosystem. The adoption of a specification standard such as OpenAPI, together with a data catalog description standard would allow a quick definition of an SDK for developing the data exchange interfaces between each vendor application, allowing the definition also of best practices for cybersecurity etc.

3.3.2 Data layer

Once the data channel has been opened, data can be transferred on it using the predefined protocol. Several protocols are currently used for data communication, so it is important that each application implements a set of protocols for network communication. In this way, the application will be able to select the correct protocol to use to access specific data streams as defined in the catalog.

Typical protocol models include the Client/Server model, which operates along a Request/Response pattern. Examples of such protocols are HTTP/S, already introduced when describing the operation of the control layer APIs, or OPC-UA [20]. According to such model, a client sends a request to a server and waits for its response. Communication can be synchronous, such as in the case of REST APIs, or asynchronous, i.e. after the initial request the server calls the client back on a specified address, communicated during the initial request, when the data is available. Such model is ideal for transferring data between two entities. Among data transmission protocol models, the Publish/Subscribe one is particularly suited to exchange telemetry data between one source and multiple consumers, as it allows high scalability, low latency and does not require strong coupling between data producer and consumer [21]. This model allows a single data producer to transmit the data over the network to all registered subscribers (i.e., authorised data consumers) with a single message (Figure 6).

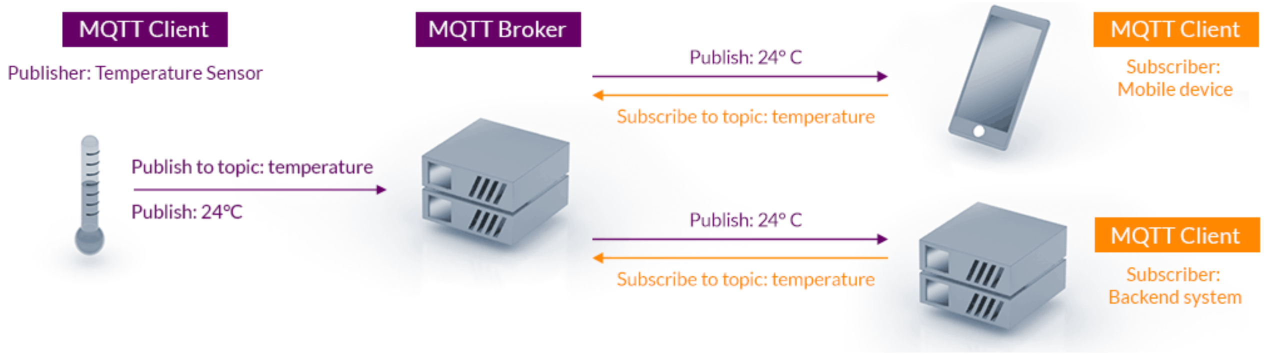

MQTT Client
Publisher: Temperature Sensor
MQTT Broker
Publish to topic: temperature
Publish: 24°C
Publish: 24° C
Subscribe to topic: temperature
MQTT Client
Subscriber: Mobile device
Publish: 24° C
Subscribe to topic: temperature
MQTT Client
Subscriber: Backend system

Figure 6 – Example of publish/subscribe data exchange with the MQTT protocol (source: mqtt.org)

The data consumers have to register to the specific “topic” (i.e., type of data needed) with the data broker operating on the data producing machine, and then all they have to do is “listen” to that topic for new data to be published. Examples of popular Publish/Subscribe protocols include MQTT [22] and OPC-UA Pub/Sub.

3.4 Implementation

As outlined in the previous sections, implementation costs can be significantly reduced and data interoperability among subsystems can be quickly achieved by making use of existing technology. ISO standards 19847 and 19848 address some aspects of data exchange on board a vessel, and thus provide a starting point and good reference to support the implementation of this proposal. However, to complete the implementation of a ship-wide data ecosystem according to CIMAC’s vision, it is necessary develop an SDK to define:

- The rules for a consistent approach to hardware virtualisation and system integration.
- The APIs required to manage the discovery, authentication, licensing and data sharing processes.
- The standard to be adopted for creating data catalogs.

- A standard data naming convention (i.e., taxonomy), so that each vendor calls the same data in the same way.

Regarding taxonomy, the lack of a standard has prompted several initiatives, such as DNV-VIS [23] or JSMEA [24], but their adoption as de-facto standards seems to pose some concerns in the industry due to the perception that one promoter could be favoured over others and could therefore acquire a competitive advantage by controlling the further development of the standard. The International Maritime Organisation also called for the adoption of a clear taxonomy standard to facilitate electronic business [25].

CIMAC encourages the creation of a consortium of marine industry players that would address these aspects and develop this SDK, thus enabling the implementation of a ship-wide data ecosystem that could leverage the power of data insight to provide concrete value for the operators and reduce the environmental impact of the marine transport sector.

3.5 Summary

State-of-the-art, mature technology is available for implementing a ship-wide data ecosystem that would enable full optimisation of the shipping process to reduce OPEX and environmental impact while adequately addressing the concerns of the marine industry stakeholders. The approach here proposed aims to provide a pathway towards seamless integration of the data subsystems on board while minimising investment costs.

In particular, virtualisation of vendor hardware, especially through the use of software **containers**:

- Ensures **protection of data property –** data inside a container is not visible to the others
- Ensures **low implementation cost** and **operation overhead** – only one data centre and one operating system need to be maintained
- Avoids the **proliferation of** data processing and connectivity **equipment** – using shared computational resources
- Avoids **vendor lock-in** – all data intelligence is in the container, it can easily be replaced.

And **the data exchange model here proposed**:

- Ensures **protection of data property –** each producer can decide which data to publish
- Ensures **business model freedom** – each producer can manage different data subscriptions according to the associated license agreement
- Ensures **secure data exchange** – only authorised subscribers can receive the (encrypted) data
- Ensures **low implementation cost** and **operation overhead** and avoids **vendor lock-in –** a common SDK is used by all players in the industry

Risk reduction associated with data sharing would be ensured by the use of proven technology.

4 Recommendations

This paper analysed the main challenges that the maritime transport industry faces when aiming to leverage digitalisation to fully optimise its processes, and presented a vision and an approach to practical implementation suitable to overcome such challenges. As explained in detail in chapter 3, and summarised in section 3.5, it is possible to leverage robust, secure, mature and scalable technology already widely used globally by many industries to minimise the implementation effort required to enable the rational integration of the intelligent subsystems that make up a ship-wide data ecosystem wherein valuable insight is shared to enable full process optimisation.

CIMAC recommends the creation of a software development kit that each vendor could use to create their digital solutions according to the proposal here presented, because this would minimise implementation cost and greatly facilitate integration with third-party systems, while ensuring that industry best practices are consistently adopted.

4.1 Technical features to be implemented in the SDK

The software development kit here promoted should implement at least the following features:

- Guidelines and methods for hardware virtualisation and integration (see section 3.2).
- Technology for asset identification and license verification, APIs for implementing system interconnection, and a standard for creating data models (see section 3.3.1)
- A set of protocols for accessing data according to the model prescriptions (see section 3.3.2)
- A taxonomy standard to name the data in a consistent way (see section 3.4)

so that scalability, ease of maintenance and integration, as well as know-how protection through secure and easy sharing of data according to specific contracts stipulated between different stakeholders can be ensured. These are key elements of any data exchange infrastructure, and they are particularly important in the case of shipbuilding, which typically involves a multitude of vendors and system integrators.

As in any professional software product, cybersecurity needs to be considered from the beginning of the development process, and to be designed in as an integral part of the product. For these aspects, it is recommended to refer to the ISA/IEC 62443 standard.

4.2 Approach to implementation

To implement the key technical features proposed in section 4.1, it is required to further develop the concept of the ship-wide data ecosystem from the strategic draft presented in this publication towards a detailed technical solution. To perform this task in a goal-oriented and efficient manner, the following recommendations are given:

- Promotion of the creation of an industry consortium that would tackle the creation of the SDK proposed in this paper, through the collaboration of ship owners/operators with equipment suppliers, with the goal of realising a first proof of concept that could become the reference for further implementations.
- Involvement in this industry consortium of the appropriate CIMAC working groups, which keep in close contact with the CIMAC Digitalization Strategy Group, to support the consortium by providing guidance and support.
- Usage of readily available technical solutions and standards wherever reasonably possible (e.g., labelling strategy for CAN systems in the automotive industry, or existing conventions and relevant ISO norms applied in the aviation industry) to avoid “re-inventing the wheel”. Thereby MTP should be generally included.

4.3 Closely coupled topics

This paper dealt with the definition and proposal of an architecture suitable to enable data exchange across the subsystems that make up the "Digital Ship Data Ecosystem" while addressing the industry concerns related to cost and IP protection.

Once this foundation is established and data exchange is enabled, further closely coupled topics, which relate to how the insight generated by data exchange and processing is utilised and corresponding implications, will arise. As two examples, safety and data attributes are further outlined below:

Safety: If a highly integrated system as proposed in this paper is established on board of a vessel and produces potentially safety-relevant advice to the crew, several safety-related items must be addressed such as *data attributes* (see below) and the avoidance of a single point of failure possibility of the system. This is in particular important to meet the IMO requirement that the risks are reduced to as low as reasonably practicable (ALARP) [26].

Data attributes: With regard to risk and safety, the data attributes are to be derived and require in-depth evaluation to ensure the correct attributes are identified [27]. In this context, a variety of items needs to be considered such as:

- The properties of data that preserve safety (e.g., accuracy, timeliness) need to be guaranteed
- Software assurance (i.e., the software lifecycle process). This will become especially important once data insights are used to recommend or initiate actions that have potential effects on safety. Different levels of software quality requirements and assurance processes will need to be put in place depending on the mission of each application and its impact on safety. (e.g. assurance of safety critical algorithms will also be required as part of the EU AI Act.)

It is the objective of the CIMAC Digitalization Strategy Group to address these and other closely coupled topics in separate, dedicated position papers to further support the concept of the ship-wide data ecosystem.

5 Appendix

5.1 Definitions

Connected product

An item that obtains, generates or collects data concerning its use or environment and that is able to communicate product data via an electronic communications service, physical connection or on-device access, and whose primary function is not the storing, processing or transmission of data on behalf of any party other than the user. [6]

Data

Any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording. [6]

Data Ecosystem

A data ecosystem is the complex environment of co-dependent networks and actors that contribute to data collection, transfer and use.

Data holder

A natural or legal person that has the right or obligation, in accordance with this Regulation, applicable Union law or national legislation adopted in accordance with Union law, to use and make available data, including, where contractually agreed, product data or related service data which it has retrieved or generated during the provision of a related service. [6]

Metadata

A structured description of the contents or the use of data facilitating the discovery or use of that data. [6]

Product data

Data generated by the use of a connected product that the manufacturer designed to be retrievable, via an electronic communications service, physical connection or on-device access, by a user, data holder or a third party, including, where relevant, the manufacturer. [6]

Related service data

Data representing the digitisation of user actions or of events related to the connected product, recorded intentionally by the user or generated as a by-product of the user’s action during the provision of a related service by the provider. [6]

Safety Related

Designated systems that both implement the required safety functions necessary to achieve or maintain a safe state for the equipment under control and is intended to achieve on its own or with other safety-related systems and other risk reduction measures the necessary risk reduction in order to meet the required tolerability risk. [28]

User

A natural or legal person that owns a connected product or to whom temporary rights to use that connected product have been contractually transferred, or that receives related services. [6]

5.2 Abbreviations

ALARP	As Low As is Reasonably Practicable refers to a level of risk for which further investment of resources for risk reduction is not justified. When risk is reduced to ALARP, it is acceptable. [26]
API	Application Programming Interface. a way for two or more computer programs to communicate with each other. It is a type of software interface, offering a service to other pieces of software. [12]
CAN	Controller Area Network
CPU	Central Processing Unit
HTTP/S	Meaning both HyperText Transfer Protocol (HTTP) and its Secure implementation (HTTPS).
IMO	International Maritime Organization
IP	Intellectual Property
ISO	International Organisation for Standardisation
MTP	Module Type Package
MQTT	Message Queuing Telemetry Transport – a lightweight, machine to machine network protocol for message queue/message queuing service.
PKI	Public Key Infrastructure
RAM	Random Access Memory
REST	Representational State Transfer. A software architectural style that was created to guide the design and development of the architecture for the World Wide Web. [11]
SDK	Software Development Kit – a set of tools for third-party developers to use in producing applications using a particular framework or platform.

5.3 Bibliography

[1] EEDI – rational, safe and effective, https://www.imo.org, 2023. Online: https://www.imo.org/fr/MediaCentre/HotTopics/Pages/EEDI.aspx (accessed 13 October 2023).

[2] EEXI ad CII – ship carbon intensity and rating system, https://www.imo.org, 2023. Online: https://www.imo.org/en/MediaCentre/HotTopics/Pages/EEXI-CII-FAQ.aspx (accessed 13 October 2023).

[3] International Convention for the Safety of Life at Sea (SOLAS), https://www.imo.org/, 2023. Online: https://www.imo.org/en/About/Conventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-(SOLAS),-1974.aspx (accessed on 24 October, 2023).

[4] RTCA DO-178C, Software Considerations in Airborne Systems and Equipment Certification, https://www.rapitasystems.com/, 2023. Online: https://www.rapitasystems.com/do178 (accessed on 24 October, 2023).

[5] ISO 26262-1:2018: Road Vehicles – Functional Safety.

[6] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), https://eur-lex.europa.eu/, 2023. Online: https://eur-lex.europa.eu/eli/reg/2023/2854/oj (accessed 29 March, 2024).

[7] The 4 Kubernetes policy types, Cloud Native Computing Foundation, 2023. Online: https://www.cncf.io/blog/2023/03/23/the-4-kubernetes-policy-types/ (accessed 13 October 2023).

[8] ISO 10007 – Quality management — Guidelines for configuration management, International Standardisation Organisation, 2017.

[9] Efficient Docker image deployment for low-bandwidth connectivity. 2023. Online: https://learn.microsoft.com/en-us/azure/architecture/example-scenario/iot/efficient-docker-image-deployment (accessed 03 October, 2023).

[10] MACH Architecture, https://macharchitecture.com/, 2023. Online https://macharchitecture.com/ (accessed 11 October 2023).

[11] REST, https://en.wikipedia.org, 2023. Online: https://en.wikipedia.org/wiki/REST (accessed 7 October 2023).

[12] API, https://en.wikipedia.org, 2023. Online: https://en.wikipedia.org/wiki/API (accessed 7 October 2023).

[13] HTTPS, https://en.wikipedia.org, 2023. Online: https://en.wikipedia.org/wiki/HTTPS (accessed 7 October 2023).

[14] Open API, https://www.openapis.org/, 2023. Online: https://www.openapis.org/ (accessed 7 October 2023).

[15] JSON, https://json.org, 2023. Online: https://www.json.org (accessed on 14 May 2023).

[16] YAML, https://yaml.org, 2023. Online: https://yaml.org/ (accessed 7 October 2023).

[17] Public Key Infrastructure, https://en.wikipedia.org, 2023. Online: https://en.wikipedia.org/wiki/Public_key_infrastructure (accessed 7 October 2023).

[18] Blockchain, https://en.wikipedia.org, 2023. Online: https://en.wikipedia.org/wiki/Blockchain (accessed 7 October 2023).

[19] What Is a Data Catalog? Data Catalog Features and Benefits, https://www.alation.com, 2023. Online: https://www.alation.com/blog/what-is-a-data-catalog

[20] OPC Unified Architecture, https://opcfoundation.org, 2023. Online: https://opcfoundation.org/about/opc-technologies/opc-ua/ (accessed 7 October 2023).

[21] PubNub Inc.: Guide, Everything You Need to Know About Pub/Sub, https://www.pubnub.com, 2023. Online: https://www.pubnub.com/guides/everything-you-need-to-know-about-pub-sub/ (accessed on 19 April, 2023).

[22] ISO/IEC 20922:2016: Information Technology – Message Queuing Telemetry Transport (MQTT) V3.11.

[23] DNV-VIS, https://vista.dnv.com, 2023. Online: https://vista.dnv.com/docs (accessed on 14 May, 2023).

[24] JSMEA, https://www.jsmea.or.jp, 2023. Online: https://www.jsmea.or.jp/ssap/topics/jsmea_iso19848.html (accessed on 14 May 2023).

[25] The IMO Compendium on Facilitation and Electronic Business, https://www.imo.org, 2023. Online: https://www.imo.org/en/ourwork/facilitation/pages/imocompendium.aspx (accessed 11 October 2023).K

[26] MSC.1/Circ 1455: 24 June 2013 Guidelines for the Approval of Alternatives and Equivalents as provided for in various IMO Instruments.

[27] SCSC-127 Data Safety Guidance: Version 3.4 Feb 2022: ISBN 9798401357663.

[28] IEC 61508-4: Functional Safety of electrical/electronic/programmable electronic safety related systems – Part 4 Definitions and Abbreviations.

Imprint

CIMAC e. V.
Lyoner Strasse 18
60528 Frankfurt
Germany

President: Rick Boom
Secretary General: Peter Müller-Baum

Phone +49 69 6603-1567
E-mail: info@cimac.com

Copyright

CIMAC is the leading global non-profit association promoting the development of ship propulsion, train drive and power generation. The association consists of National Member Associations and Corporate Members in America, Asia and Europe. CIMAC provides a forum for technical interchange with all parties interested in piston engines, gas turbine systems, non-shaftline propulsion systems, automation and controls, system integration and digitalization solutions.

CIMAC acts as a global platform for discussion through a range of events, namely the CIMAC Congress (once every three years), CIMAC Circles, CASCADES and web seminars. The content-related work evolves around CIMAC's Strategy and Workings Groups which produce publications on various topics.

For further information about CIMAC please visit http://www.cimac.com.