

ABS ADVISORY
ON DATA QUALITY FOR MARINE AND OFFSHORE APPLICATION - METADATA

JUNE 2022

ABS

TABLE OF CONTENTS

INTRODUCTION 1

BACKGROUND 1

METADATA FOR SMART, AUTONOMOUS AND REMOTE CONTROL FUNCTIONS. 1

METADATA FOR VESSEL DESIGN 2

GOALS AND SCOPE 3

METADATA CONCEPTS. 4

METADATA FUNDAMENTALS 4

METADATA TYPES. 5

METADATA STANDARDS 7

METADATA REGISTRIES 10

CENTRALIZED METADATA ARCHITECTURE 11

DISTRIBUTED METADATA ARCHITECTURE 11

HYBRID METADATA ARCHITECTURE 11

METADATA ENCODING 12

METADATA TOOLS 13

CROSSWALKS 13

DATA LINEAGE AND IMPACT ANALYSIS. 14

METADATA QUALITY METRICS. 16

ABS ROLE 18

SUMMARY 20

REFERENCES 21

INTRODUCTION

BACKGROUND

The implementation of smart, autonomous and remote-control functions for marine and offshore assets in recent years has resulted in an increase in operational data collection onboard vessels. Those functions rely on such information to support decision-making by humans or the systems themselves. Metadata provides the necessary context for data to be processed and translated into usable information.

This document in ABS' continuing series on the utilization of vessel data, focuses on metadata and summarizes best practices on creating, using and managing metadata to support the development and implementation of smart, autonomous and remote-control functions in marine and offshore applications. Metadata plays an important role in achieving quality data due to its descriptive nature. The metadata tools, procedures and quality metrics in this document complement the data quality guidance provided in the ABS *Advisory on Data Quality for Marine and Offshore Application – Time Series Data* [3] and ABS *Advisory on Data Quality for Marine and Offshore Application – Transactional Data* [4].

METADATA FOR SMART, AUTONOMOUS AND REMOTE-CONTROL FUNCTIONS

Smart, autonomous and remote-control functions are enabled by the ability of machines to collect and process data. Metadata plays a vital role in almost all aspects of those functions. Metadata is created as the data is collected, it is involved in data exchanges between components in the system, enables the discovery of data, and can document system usage and performance.

Metadata, while data itself, is descriptive in nature. Metadata defines and describes what other data is and what the data relates to [13]. Onboard sensors and systems are a common source of time-series data for smart, autonomous and remote-control functions. Data acquired from a sensor requires metadata to describe attributes such as the type of value being recorded (temperature, pressure, etc.) and unit of measurement. Metadata is also required to relate the data to other objects such as the specific sensor that generated the data and the vessel where the sensor is located.

Health or performance monitoring reports and other transactional data generated by smart, autonomous and remote-control functions also contain metadata. The reporting period, vessel name and other descriptive metadata enable users to locate the correct records. To support interoperability, the detailed definitions of value domains ensure consistency of terms and formatting. If the functions are designed to send automated status messages, relevant metadata such as the date and time the message was sent, the recipients and message subject can be captured.

Time series and transactional data collected during vessel operation are a common source of metadata for smart, autonomous and remote-control functions. Time series data is a collection of time-stamped data obtained over a certain period at a particular frequency such as measurements obtained from sensors. Transactional data refers to the type of discrete, time-stamped business and device data generated by maritime operations, condition and performance monitoring, inspection, repair and maintenance activities. Transactional data may take on a variety of forms and is less prescriptive in structure than time series data. Noon reports, event-triggered data capture, survey reports and maintenance data are examples of transactional data.

Service deployed smart, autonomous and remote-control functions will continuously collect and transmit data to support monitoring, analysis, decisions and actions by the functions. This data will often be used beyond its primary role for purposes such as calibrating simulation models or updating digital twins to reflect the vessel's current condition. Metadata provides the necessary context, formatting and structure details to perform data exchanges among tools and systems.

A focus on metadata will increase the quality and value of smart, autonomous and remote-control functions. It will ensure data can be readily discovered and accessed. It can capture the necessary usage and revision histories to support auditing. Data quality is promoted through adherence to standards and enforced formatting restrictions. A fuller understanding of data flow will allow the impact of changes to quickly be assessed.

METADATA FOR VESSEL DESIGN

Metadata for smart, autonomous and remote-control functions, the focus of this advisory, is closely associated with the operational phase of a vessel's life cycle. With the digitization of the marine and offshore industries, the use of metadata begins at the onset of a vessel's life cycle during the design phase, continues with construction, and lasts throughout its life cycle. An abbreviated outline of current applications of metadata during vessel design and verification are presented here to provide broader context for applications related to smart, autonomous and remote-control functions.

The creation of three-dimensional (3D) computer-aided design (CAD) models of a vessel is a core activity during the design phase. Metadata embedded within the 3D CAD models documents everything from the attributes and components of a structural part to the 3D model's watermark and the date the model was created. Part names and numbers, components and descriptions, including but not limited to physical dimensions, materials, schedules and coordinates are key information used to describe CAD models. Additionally, hierarchal parent-child relationships or groupings of parts provide additional layers of metadata that allow functional intent to be captured such as compartment relationships, system associations or functional category. The various levels of metadata associated with the 3D models shape how data can be searched and accessed.

Through metadata, it is possible to create, manage, distribute, publish, access and reuse, and archive whole cycles of data, or just specific pieces. Use of metadata to manage disclosures can help protect proprietary information; rather than releasing all attributes associated with a model, a specific subset of attributes such as cost and manufacturing details can be restricted. In some cases, metadata values alone may be sufficient in eliminating the need to share entire models. For example, metadata for mass, center of gravity and moments of inertia of a small rigid component may be the only information necessary to assess compliance of its foundation to Rule requirements. In the future, metadata associated with 3D models can be used for automated rules and regulation check points, extended reality and providing higher overall process efficiency when the metadata ensures consistency.

A key proponent to the use of design metadata are product life-cycle management (PLM) applications where it could be involved in the various product stages like development, construction and operation. CAD metadata would thrive in this type of application environment where engineers, designers, manufacturers or shipyards would work with the same set of metadata but could focus primarily on those needed for their processes. For example, the product manufacturing information (PMI) of a part or a system would give the relevant set of data that a shipyard needs to assemble the structural parts together. Simulation and analysis tools may also interface with PLM applications to retrieve design metadata and 3D models stored in their repository.

Modeling and simulation constitute a powerful method for designing and evaluating complex systems and process and fall into four major application categories: proof-of-concept of a new design, modifications or retrofit to the existing design, comparison between several design options and optimization. For any type of simulation, obtaining the appropriate parameters to use in the simulation is a crucial process. The key to using the right parameters for each simulation is the ability to connect simulation models to design metadata managed by the PLM application that acts as a single source of truth. Simulation frameworks are embodied with components stored in libraries and populated with metadata for parameters and system attributes such as the number of engines, engine specifications, vessel length and displacement. Operational time series and transactional data support the calibration and improvement of simulation models and is essential for machine learning with autonomous systems.

A challenge shipyards and designers are facing is the disconnection between design data in CAD format and the data for computer-aided engineering (CAE) based evaluation. Typically, engineers are required to enter the ship data and build the ship CAE models by themselves due to the proprietary data format of different CAD and CAE software although almost all required data for CAE analysis is available in the CAD-based design. A common metadata schema for data exchange supported by various CAD and CAE software will benefit the designers by not only reducing the data and model preparation time but also avoiding potential human errors introduced during data and model preparation.

As a leader in marine and offshore classification services, ABS recognizes the efficiency that metadata can bring to the classification process. The Unified Data Model (UDM) metadata schema is developed to bridge the gap between design and evaluation. CAD metadata through UDM schema is being used throughout the design approval process to allow users to search for specific items and understand a vessel's design characteristics such as material grade and structural properties and import the design data such as tank definition, sectional configuration and properties into ABS engineering applications for design verification. Disclosure of a client's CAD model metadata to ABS makes the process of reviewing the submitted design smoother and more efficient. ABS in return will be able to add its own set of metadata through approval stamps, amendments and comments to the 3D model. CAD metadata can be extended to engineering analysis software where compatible, e.g., Finite Element Analysis (FEA), CAE and Control System Analysis all would benefit from the reuse of metadata from one true authoritative source. Where requirement management applications are deployed, users can leverage exposed CAD metadata to perform Rule checks on the designed part or structural entity. Throughout this whole process of metadata exchange between applications, the CAD metadata integrity would be maintained.

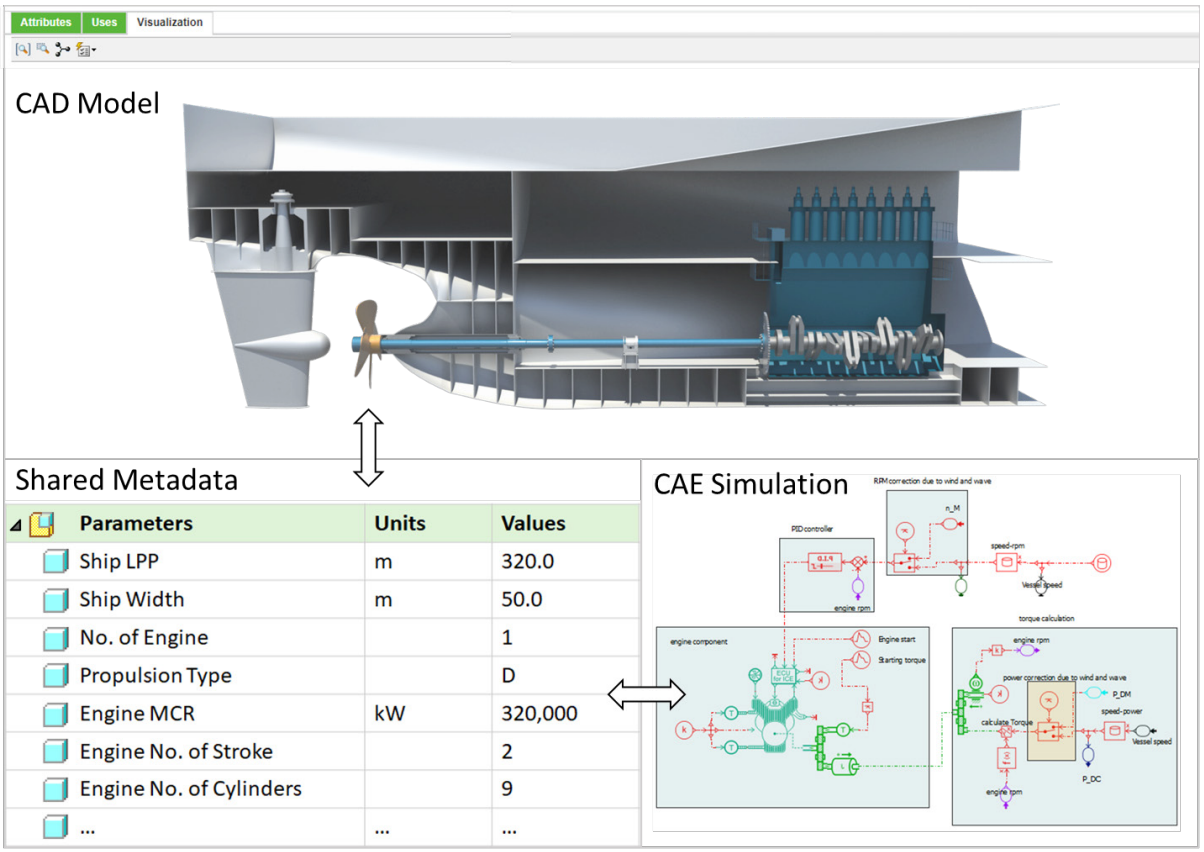


Figure 1: Metadata exchanges for modeling and simulation during vessel design

GOALS AND SCOPE

This advisory provides an overview of the types of metadata, relevant standards, metadata registries and best practices. Guidance and recommendations are offered for using metadata to perform data lineage analysis, creating metadata crosswalks to enable interoperability and metrics for quantifying metadata quality.

The scope of this advisory will address all levels of metadata implementation from conceptual data models to the value domain and finally to encoding in machine readable formats. This approach allows opportunities for metadata to be considered throughout a systematic design process for a smart or autonomous function. While the practices presented in this document apply to both new and existing systems, implementation during the design of a new system provides an opportunity for reduced effort and cost. The same techniques used during system design can be re-applied throughout the life cycle of the system to adapt the metadata to system changes.

METADATA CONCEPTS

METADATA FUNDAMENTALS

Metadata is data that defines and describes other data [13]. By definition metadata itself is data, but it plays an essential role in the process of making data reinterpretable and suitable for communication, interpretation or processing. The time-series data collected on board a vessel by a sensor is often associated with a measurement. By itself, the recorded number indicates a quantity or value but has no context. What does the number represent or refer to? Metadata is the data that provides the context necessary to allow information to be obtained from data.

The fundamental data models for metadata have three components: the data being described, the descriptive context, and the descriptive data. In the Resource Description Framework (RDF) [11] standard the terms for the three components are: subject, predicate and object. Together, the three components termed triples form a statement. An RDF statement defining the sensor type for a temperature sensor, Sensor_T1, is illustrated in Figure 2. In the graph for the RDF triple the subject and object are nodes and the predicate is an edge connecting the nodes.



Figure 2: Graph for an RDF statement

Multiple metadata statements can be made about a single subject. Additional statements describing the unit of measurement, location, compartment number, sampling rate, vessel and calibration date can also be made about the sensor. The objects in an RDF statement may also be subject of other statements. The vessel the sensor is installed on may be described with its own set of metadata statements as illustrated in Figure 3.

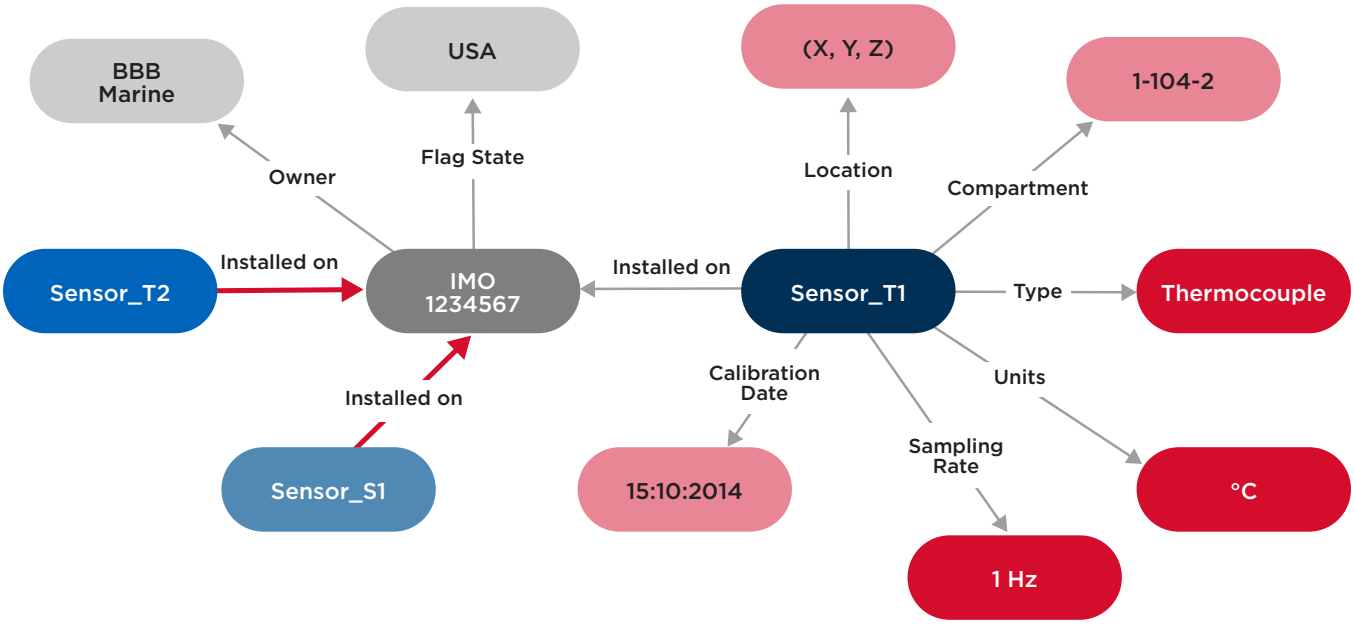


Figure 3: Graph of multiple RDF triples

The creation of metadata involves activities at three levels: a conceptual level, a representation level, and an encoding level. At the conceptual level general relationships and object classes are defined. The representation level concerns permissible values and formatting details. Encoding is required to express the metadata in a machine-readable format.

To illustrate these levels, consider the need for metadata to uniquely identify a vessel. At the conceptual level general statements can be defined: the vessel is identified by an identification number. Semantic refinements allow creation of concise metadata. For example, the statement could be revised to specify that the identification number is assigned by the International Maritime Organization (IMO). The representation level focuses on the value domain and how the metadata will be represented. For an IMO number the format is known to be a seven-digit number and it is only valid if the seventh digit is consistent with the established checksum formula. The encoding level involves structuring the metadata in the correct syntax so that it can be processed by software. Metadata is often expressed using markup languages: <SHIPID> IMO1234567 </SHIPID>.

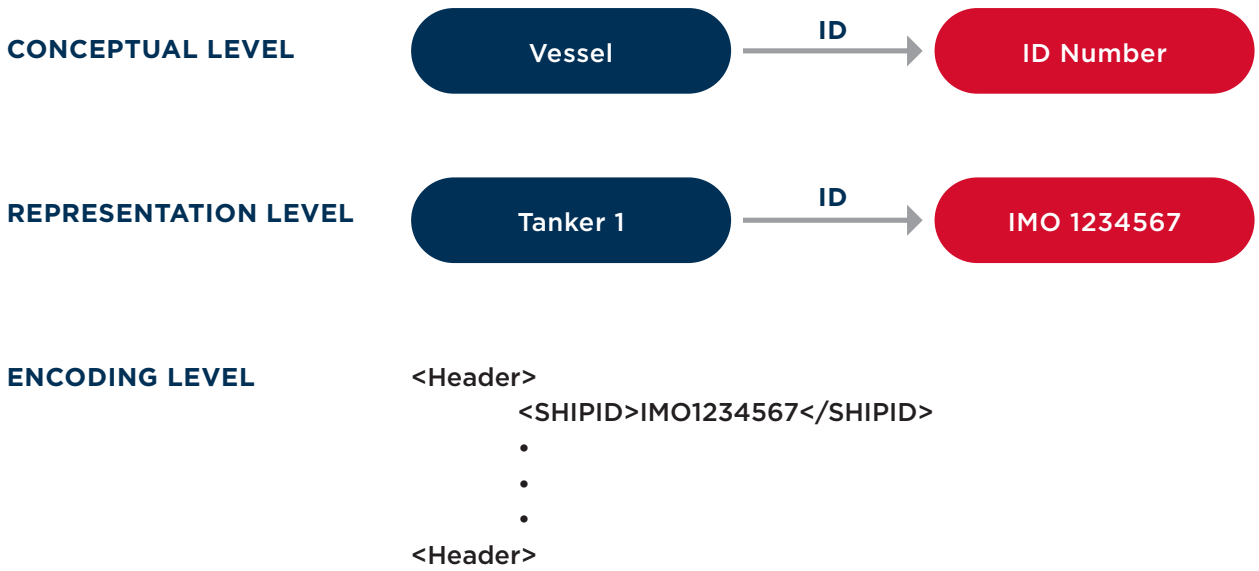


Figure 4: Metadata creation

METADATA TYPES

The breadth of metadata lends itself to categorization into various types. The four categories listed here are commonly used. Categorization is useful when establishing goals or delegating responsibility for metadata. The application of categories is generally thought of in a broader sense to avoid too much focus on metadata that may fall on the line between two categories. Other categorization methods are possible; broadly focused projects may focus on the origins of metadata such as business operations, clients or external databases.

DESCRIPTIVE METADATA

Descriptive metadata provides a description of an object. The purpose of descriptive metadata is often to enable the discovery of the object. The markings and Convention for Safe Containers (CSC) plate on shipping containers are an example of descriptive metadata. The container's serial number, owner, size and type code, manufacturer, and maximum gross weight are some of the descriptive metadata contained in the container markings and CSC plate. Descriptive metadata is important for smart functions that generate and store time-series data. In digital file formats descriptive metadata is often packaged into a header before the bulk data in much the same way as an e-mail contains metadata such as the sender, recipient(s), subject and date/time before the body of the message. The descriptive metadata allows data analysts to quickly search and locate files containing the desired data.



Figure 5: Container markings, example of descriptive metadata

ADMINISTRATIVE METADATA

Administrative metadata is concerned with the management of an object over its life cycle. The revision tables in engineering drawings and documents are examples of administrative metadata. Those tables summarize changes, specify where they occur, when they were made and who made them. Technical metadata is often a subset of administrative metadata and may include details such as hardware or software requirements. Capturing vessel or system configuration details and modifications is important for smart, autonomous and remote-control functions. Intellectual property rights information is also considered administrative metadata and may be used to control access to proprietary data.

REVISIONS				
REV	ZONE	DESCRIPTION	DATE	SIGNATURE
—	ALL	ORIGINAL ISSUE	4-26-85	D. Kortbew
A	14-D 14-G	1. REVISED SS. GEN. CONTROL WRG. DIAG. TO VENDOR DWG #3077	10-25-85	S. Pomecoy
B	16-C 5-E 16-B	1. ADDED SHORE POWER ENGERGIZED WRG DIA. 2. ADDED PIN'S 23 &24	11-20-85	S. Pomecoy
C	21-D 21-E 24-G 24-B	1. DELETED CABLE CONDUCTOR 1-PC-8 & 2-PC-8 FROM TB-1 & TB-2 RESPECTIVELY. 2. REVISED MAIN ENGINE JUCTION BOX TERMINALS PER VENDOR INFO.	10-13-86	S. Pomecoy

Figure 6: Drawing revision table, example of administrative metadata

STRUCTURAL METADATA

Structural metadata describes how objects are organized. Structural metadata is inherent to complex objects or systems such as marine vessels that are built from numerous parts. The hierarchical relationships between the parts, such as how they are arranged into assemblies or subsystems are structural metadata. Other examples of technical metadata include a document's table of contents or tags in data files that indicate the start and end of specific data types.

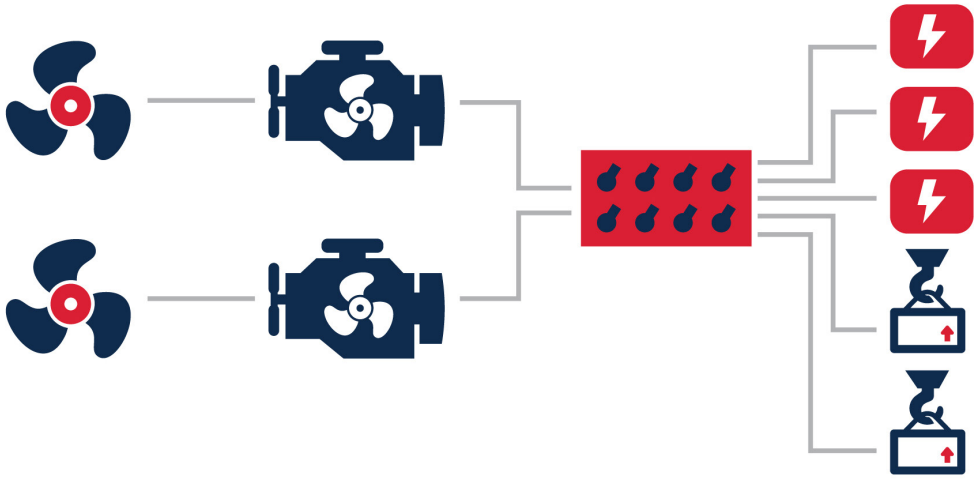
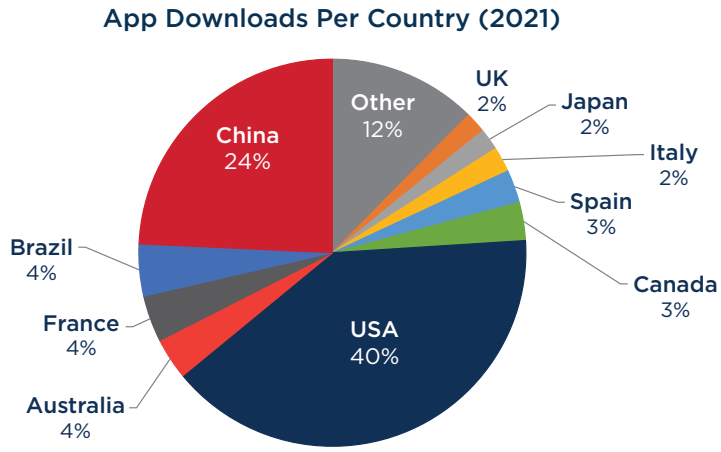


Figure 7: Propulsion system arrangement, example of structural metadata

USE METADATA

Use metadata provides information about how an object has been used. Examples of use metadata include details about how many times a file has been downloaded or the dates and times users access a system [15]. For smart functions, use metadata can be used for auditing purposes to verify scheduled inspections took place. It can document when event-triggered activities were performed and by whom. Use metadata can highlight usage patterns to guide future development or identity quality issues such as miscategorization.



Country	App Downloads
Other	1,007
UK	142
Japan	147
Italy	171
Spain	224
Canada	258
USA	3,253
Australia	282
France	315
Brazil	352
China	1,971
TOTAL	8,122

Figure 8: Performance insights, example of use metadata

METADATA STANDARDS

Metadata standards are available for all aspects of metadata creation and use ranging from standards for conceptual data models to formatting of files for computer interpretation. Four types of metadata standards have been proposed by Zeng and Qin [17].

STANDARDS FOR DATA STRUCTURE

Standards for metadata structure detail the framework and semantics for describing resources. Data structure standards formalize metadata element sets for defining the properties of resources and their purpose. The Dublin Core Metadata Element Set (DCMES) [9] is a well-known data structure standard designed to be universal to any subject. The DCMES is only comprised of 15 irreducible elements that enable it to be broadly adopted. Data structure standards may also be domain specific; ISO 19115 concerns geospatial metadata. In practice it is possible to select and apply metadata elements from one or more standards. Adobe's Extensible Metadata Platform (XMP) [10] uses a subset of the DCMES in addition to its own unique elements.

STANDARDS FOR DATA CONTENT

Data content standards guide the cataloging of resources and define required information. Best practice documentation is often available with metadata standards to provide recommendations for their uniform application. From a corporate standpoint, internal teams are often used to define data content standards for business metadata that will be used throughout the enterprise. For smart, autonomous and remote-control systems, standards for data content may be driven by transactional data requirements established by regulatory bodies, business partners or government entities.

STANDARDS FOR DATA VALUES

Data value standards are used to standardize the permissible metadata values to promote data quality and consistency. Examples included controlled lists of terms such as ISO 639-1 that is a list two-character codes for classifying languages, e.g. English = "en". Standards can also be used to control formatting. ISO 8601 is a formatting standard for dates and times.

STANDARDS FOR DATA EXCHANGE

Data exchange standards define formats for expressing metadata so that it can be exchanged or processed by machines. Metadata schemas are often implemented using markup languages that allow the structure of the data to be expressed. The Extensible Markup Language (XML) and JavaScript Object Notation (JSON) are data exchange standards.

The preceding categories for metadata standards follow the typical metadata creation process with data structure and content standards applying to the conceptual domain, data value standards corresponding to the value domain, and encoding being governed by data exchange standards. Often standards do not directly align with a single category and contain content that will fall in multiple categories. The limited scope of domain specific standards allows them to more frequently encompass all four categories. The Open Class 3D Exchange (OCX) [18] standard developed by a marine industry consortium is an example of such a domain specific standard. The goal of the OCX standard is to improve the exchange of information between shipbuilders and classification societies through an interoperability specification for data exchanges between three-dimensional (3D) computer-aided design (CAD) design tools and class confirmation tools.

Adhering to standards has the benefit of improving interoperability and content consistency. Despite the large number of standards available there is often a need to account for unique technical or business requirements. In addition to required elements, many standards have optional or conditional elements that allow for some degree of customization. When greater flexibility is required the practice of creating unique data elements to augment existing standards or combining data elements from multiple standards is generally acceptable.

Two standards, BS ISO 19847 [7] and BS ISO 19848 [8], were created for the marine industry. BS ISO 19847 defines requirements for onboard data servers that collect, store and share data from other onboard sensors, machinery and systems. The standard specifies data protocols per BS ISO 19848 that defines two concepts and their models for data exchange. These standards are directly relevant to smart, autonomous and remote-control functions and are outlined in detail. The standards are not strictly metadata standards, but their requirements concern the creation and exchange of important metadata. Both standards address data content and values while BS ISO 19848 also defines data structure and encoding examples.

BS ISO 19847 SHIPS AND MARINE TECHNOLOGY — SHIPBOARD DATA SERVERS TO SHARE FIELD DATA AT SEA

Smart and autonomous functions involve the collection of data on board vessels. Shipboard data servers are typically required to store data. Even if the smart or autonomous functions are designed to transmit data to an onshore or cloud-based data server, an onboard server is needed to store some data in the event communication with the remote server is lost. This will allow onboard data to be resent once network issues are resolved.

BS ISO 19847 specifies requirements for the onboard data server. Based on the requirements, the server generates important metadata regarding performance. Per the specification, the server shall monitor the system health for anomalies and report on systems status or failures (§ 5.1.4). The server also manages the system clock and reports loss of synchronization (§ 6.2.1). Time stamps generated by the server are important metadata and their accuracy is essential for data quality.

The data management requirements in BS ISO 19847 specify that usage metadata is recorded by the server. Additions, updates or deletions to the data channel list, the data source information and alias list must be logged (§ 6.2.2.b, § 6.2.3.b, § 6.2.4.b). The server must also indicate the current configurations and log user activity (§ 6.6).

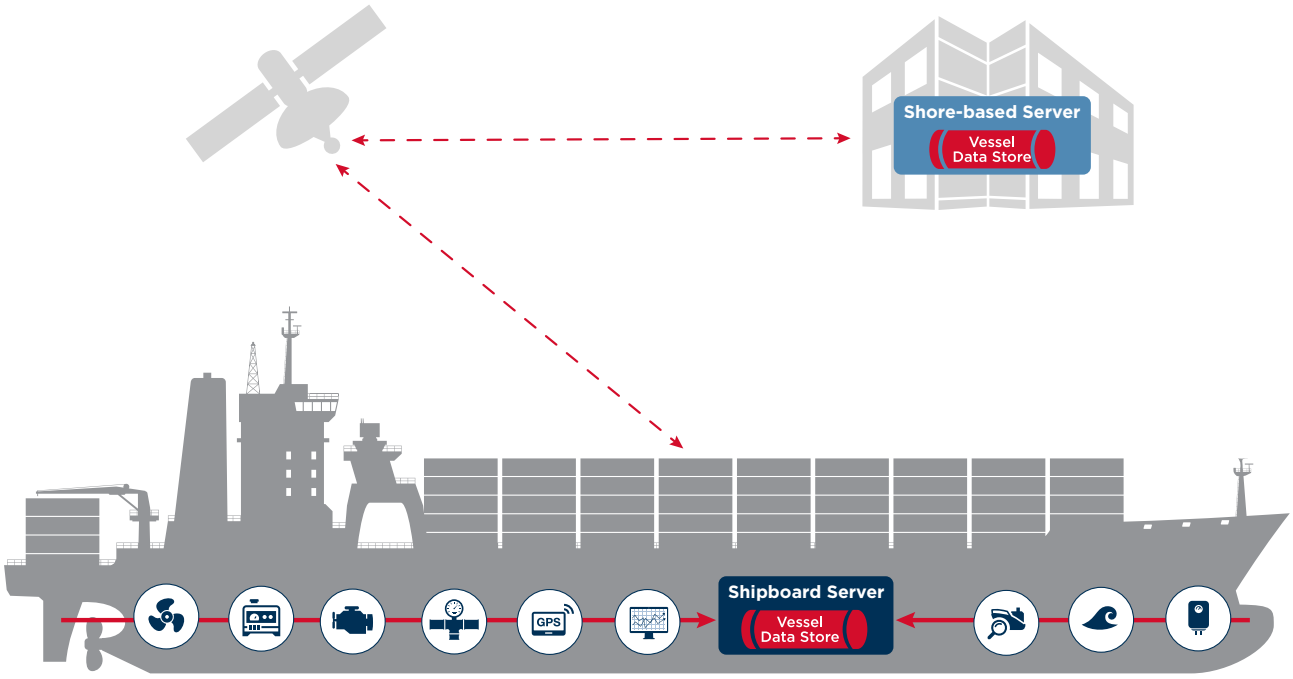


Figure 9: Metadata creation and storage on shipboard data servers

The data channel list stored on the server contains metadata pertaining to the names of channels. The data source information specifies data formatting and structural details of the data package. The alias function allows data for one or more data channel to be retrieved at one time through a data request by specifying an alias ID (§ 6.4). The alias list stores the structural metadata that details data channel(s) associated with each alias.

BS ISO 19848 SHIPS AND MARINE TECHNOLOGY — STANDARD DATA FOR SHIPBOARD MACHINERY AND EQUIPMENT

The collection of data to monitor the performance and health of ship machinery and structure has become more common. This standard defines data structures for two concepts: the Data Channel and Time Series Data. The Data Channel concept describes the time-invariant properties associated with a particular sensor. The Data Channel contains all the metadata required to identify (Data Channel ID) and describe (Data Channel Property) the sensor. The Time Series Data concept only carries minimal metadata for referencing measurement data to the channel list to reduce the size of the data package.

The Data Channel ID used to identify specific data channels is structured to allow for multiple ID definitions: Universal ID, Local ID and optional Short ID. The Universal ID is the most structured and requires metadata for uniquely identifying the vessel such as an IMO number. Local IDs allow for consistent naming with other systems that may be onboard the vessel.

The Data Channel Property contains essential metadata information to describe the data channel type, format, range, units, quality coding, name and remarks. The structure of the data channel type allows for the inclusion of statistical data for a channel such as an average value over a defined time period in addition to the raw data. Additional metadata such as the calculation period for the statistical data and the data update cycle are part of the standard. The format field provides metadata describing the type of data included: integer, real, Boolean, string or date/time. Optional format restriction metadata allows for the specification of further constraints to support data quality checks. Defined maximum and minimum values are examples of such restrictions. Data received outside of the defined boundaries could indicate poor quality data.

The Time Series Data concept can represent data in two distinct formats. The Tabular Data format supports continuously collected at the defined time interval. Each row in the Tabular Data format is a Data Set that is the collection of data channel values for the recorded time stamp. Meta-data describes the Data Channel of each column in the table. The Event Data format is designed to capture data where the number of values collected at a specific time is not fixed. Examples of event data include alarms, manual inputs and status updates. To improve storage efficiency the Event Data format only contains Data Sets and the associated Data Channel IDs when a value is available.

The data structure defined in the standard are designed to be implemented in an XML schema although Comma-Separated Values (CSV) and JSON formatting is also possible. The data structure has some provisions for some custom user defined elements. Mandatory and optional portions of the elements are detailed.

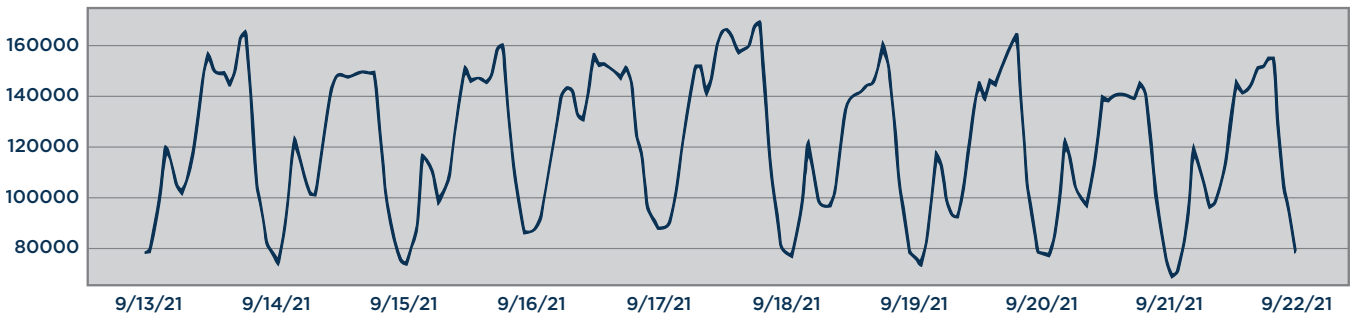


Figure 10: Time series data

METADATA REGISTRIES

Metadata registries are databases for metadata. The purpose of registering metadata in a registry is to enable identification, provide provenance and monitor quality. Metadata registries provide a central place for users to access metadata elements, vocabularies, encoding schemas or referential relationships to enable their reuse. By enabling users to browse and search existing metadata descriptions standardization is promoted. Users are also able to use metadata registries to publish completed works so they are available for others to view.

The identification role of a metadata registry is achieved by assigning a unique identification to each object loaded into the registry. The identification process should help organize the objects in the registry so that is possible to ascertain if a new object being loaded into the registry already exists or not. Due to the large number of metadata standards, it is possible for some objects to have similar or identical names due to the common use of terms such as title. The registration process can help users identify these situations and apply naming conventions to ensure uniqueness or prevent redundancy. The element "title" may be uniquely identified by specifying its namespace which is a collection of names identified by a unique reference. For example, in Extensible Markup Language (XML) encoding affixing the prefix "dc" would indicate the namespace for element "dc:title" uses the Dublin Core schema. The identification of two similar element names, "FuelOilTank1" and "FOTank1," would indicate that one of the entries is redundant due to inconsistent naming.

Metadata registries document the source of objects registered to provide provenance. Knowledge of an object's provenance can often play a role in the selection process if multiple suitable objects are found. International or industry standards, proprietary standards from equipment or software vendors, national bodies and internal corporate standards are all possible sources of metadata. Provenance is also required to document any revisions to objects. When possible, provenance metadata should document relationships between objects that influenced its history. Relationship information is applicable for objects such as regulatory standards that are often revised to comply with changes to made by higher regulation authorities.

Objects loaded into metadata registries are monitored for quality. Objects published to the metadata registry can be reviewed for conformance with the appropriate conventions for forming definitions, establishing names, and categorizing items. When similar objects are identified, the quality of the two objects can be compared to determine if one is of higher quality than the other. Decisions can then be made if there are objects that may no longer be relevant.

Standard ISO/IEC 11179 provides a framework for metadata registries. Currently the standard contains seven parts:

- ISO/IEC 11179-1:2015 Framework
- ISO/IEC 11179-2:2005 Classification
- ISO/IEC 11179-3:2013 Registry metamodel and basic attributes
- ISO/IEC 11179-4:2004 Formulation of data definitions
- ISO/IEC 11179-5:2015 Naming and identification principles
- ISO/IEC 11179-6:2015 Registration
- ISO/IEC 11179-7:2019 Metamodel for data set registration

Different architectures may be used to implement metadata registries. The architectures differ in the manner in which metadata is sourced, stored and accessed. Three architecture models are presented in the DAMA-DMBOK2 [12] are outlined below. The application specific nature of many smart, autonomous and remote-control functions on board vessels may involve the creation of user defined metadata elements. Of the three architectures, only the distributed architecture lacks the means to support user defined on manually entered metadata.

CENTRALIZED METADATA ARCHITECTURE

A central metadata architecture has a single metadata repository that contains copies of metadata from their sources. This model ensures the greatest consistency because all users will search and browse from the same central database. The searches are global in nature because centralized systems help ensure the availability and quick retrieval of metadata because the objects are independent of their source system. The extracted metadata may be transformed, customized or enhanced before being published to the registry. One disadvantage is higher maintenance costs are associated with the need to ensure changes to source data are quickly reflected in the repository and the validation and maintenance of any customized code.

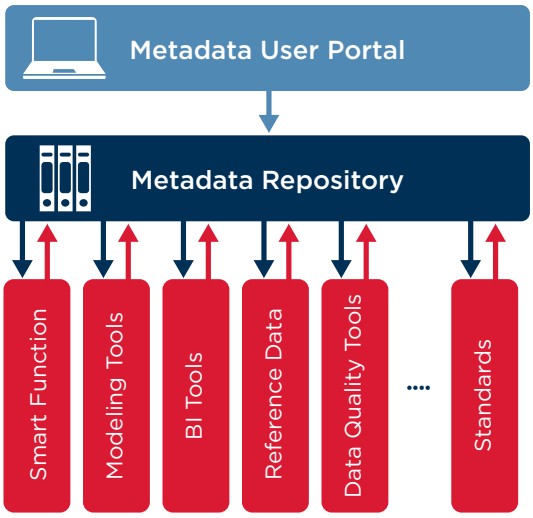


Figure 11: Centralized Metadata Architecture

DISTRIBUTED METADATA ARCHITECTURE

Distributed metadata architectures still have a single access point for users, however they directly retrieve metadata from source systems when requested in real time. No persistent repository exists, but the distributed architecture maintains the necessary source paths to access the metadata and process the user's search and retrieval requests. An advantage of a distributed architecture is that the most current metadata is obtained because it is directly retrieved from its source. An additional advantage of distributed architecture is that the cost and complexity of implementing query processing is less than that of a centralized architecture. One limitation of a distributed architecture is that the quality and availability of the metadata is dependent on the source systems. Distributed architectures also do not provide for user defined or manually entered metadata.

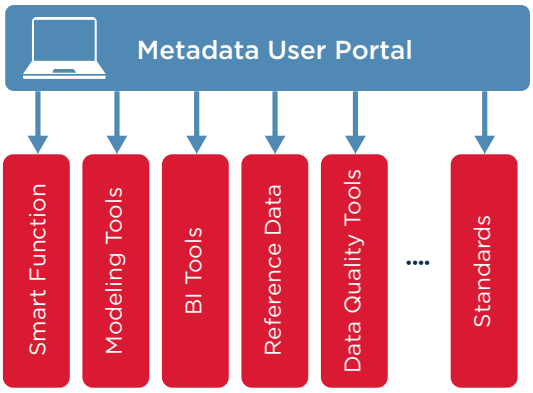


Figure 12: Distributed Metadata Architecture

HYBRID METADATA ARCHITECTURE

The hybrid metadata architecture combines the attributes of the centralized and distributed models. The hybrid model collects metadata and stores it in a central repository, however only critical standardized items, custom user added items and additions from manual sources are stored. By focusing on a narrower scope of objects in the central repository, the effort to ensure that the metadata is current and reduced. The hybrid metadata architecture has a distributed back-end systems structure and has the same availability limitations as the distributed architecture.

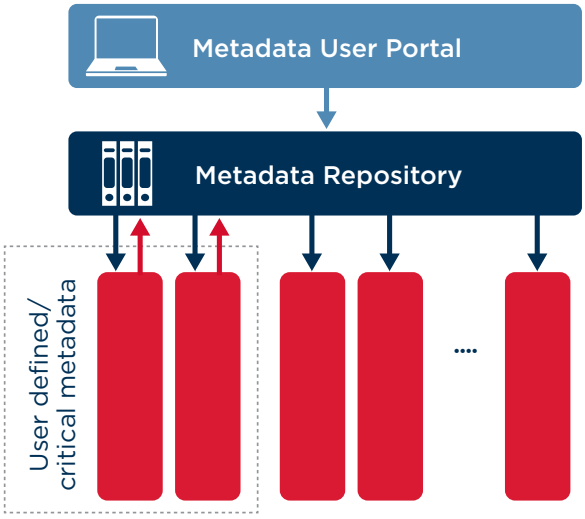


Figure 13: Hybrid Metadata Architecture

METADATA ENCODING

The digitization trend in the marine and offshore industries has led to an increase in the number of computerized systems on board vessels. To process metadata, those systems require the metadata to be encoded in a machine-readable format. While it is possible to encode metadata for the simplest of systems manually, large complex systems will greatly benefit from a set of dedicated software tools. The breadth and scope of software available to create, use and manage metadata is vast. This section highlights various options for encoding metadata.

The metadata needs of users are diverse and as a consequence users should not expect to find a single comprehensive metadata software that satisfies all their criteria. Many metadata software only perform a single or limited range of functions. The largest software developers (SAP, Oracle, IBM, etc.) offer suites of products to perform multiple tasks. Clients can select the desired software products to meet their needs. Customized products may be developed to fill any gaps. Other software companies may focus on a specific application area such as the creation of data modeling tools.

One criterion to consider in software selection is the standards that the software is designed to comply with. Software that complies with the desired standard will help ensure the metadata created is portable with other organizations that adopt the same standards. Many software enable the import or export of several different file formats, however it is advisable to understand if the import/export utilities support full or just partial coverage of those format standards. For example, a computer-aided design (CAD) software may import the three-dimensional (3D) geometry from a different program correctly but fail to translate metadata such as names of parts or their hierarchical relations. While some metadata software is designed for universal application others may be tailored to specific industries. The targeted software may be beneficial if they are optimized for the user's needs.

Depending on a business's resources it may be within their realm to develop propriety metadata tools. The development of propriety tools allows for complete customization. As an alternative to the development of standalone software tools, a company may already utilize software that support an advanced programing interface (API). The API may provide sufficient flexibility to embed and/or extract metadata using existing software. As an example, some camera systems offer simple APIs that allow user-defined tags to be added to photo or video files as they are recorded in addition to the common options for time and geospatial positioning information. Embedding vessel ID, camera orientation, a description of the subject of the images, and other details can support subsequent processing.

In addition to commercial software, free publicly available metadata software and resources exist. Those tools are often created to support research and collaboration in the areas of science, education, and the arts. Web searches may reveal countless other software tools with varying degrees of validation and testing.

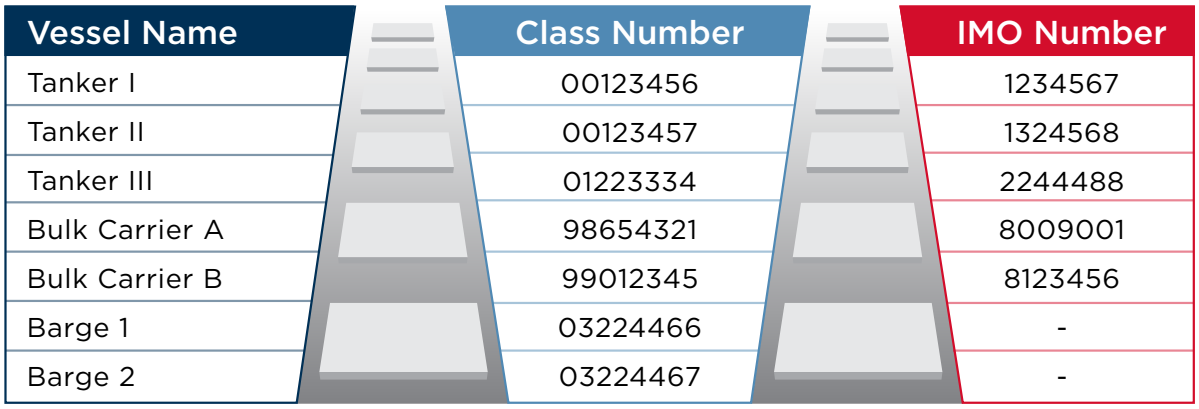


In summary software tools are available to support metadata creation, use, and management. At the highest conceptual level software is available to layout data models and visualize data lineage. At the final encoding step XML editors aid in visualization by performing syntax highlighting and support the conversion of XML to Extensible Stylesheet Language (XSL) to simplify the development of schemas.

METADATA TOOLS

CROSSWALKS

A common metadata tagging task is to perform mappings of values between different vocabularies. These mappings are termed crosswalks. The simplest crosswalks involve the mapping of vocabularies with the same number of terms where direct one-to-one mapping is possible. Consider a crosswalk between a vocabulary containing values of the vessel names for a ship operator's fleet and a vocabulary containing the class numbers for vessels in the fleet. Because name and class number for each vessel is unique the number of terms in the two vocabularies are equal and one-to-one mapping is possible.



Vessel Name	Class Number	IMO Number
Tanker I	00123456	1234567
Tanker II	00123457	1324568
Tanker III	01223334	2244488
Bulk Carrier A	98654321	8009001
Bulk Carrier B	99012345	8123456
Barge 1	03224466	-
Barge 2	03224467	-

Figure 14: One-to-one crosswalk mapping

Often the task of creating crosswalks is more challenging and one-to-one mapping is not possible. From the preceding example, assume the two barges in the fleet operate on inland waterways and therefore have no International Maritime Organization (IMO) number. The resulting vocabulary for IMO number will have two fewer terms than the other two vocabularies. Conflicts like this are particularly important when mapping tasks are automated by software. Special attention should be paid to how exceptions are handled to prevent the creation of poor-quality data. If the shorter vocabulary is the source and the longer vocabulary is the target the mapping is possible, but the reversal may result in null values for non-mappable spaces. One-to-many mappings are also possible as would be the case if a crosswalk was established for a vessel type vocabulary, (tanker, bulk carrier, barge), using the example fleet.

Crosswalks may also involve the mapping of multiple value domains to a single domain or vice versa. These type of mapping issues are illustrated below for location and date/time data that are common to time-series data. One schema may have separate fields for the latitude and longitude while a different schema may have a single field to store vessel position. In this example the value domains for the schemas also differ with one using a decimal format and the other a degree-minute-second format. These types of mappings become more challenging in the conceptual domain where differences in meaning and scope may be a factor. A conceptual domain conflict could occur between the two vessel type elements: "bulk carrier" and "ore carrier." Conceptually the two elements have different scopes as all ore carriers are bulk carriers, but the reverse relationship does not hold true.

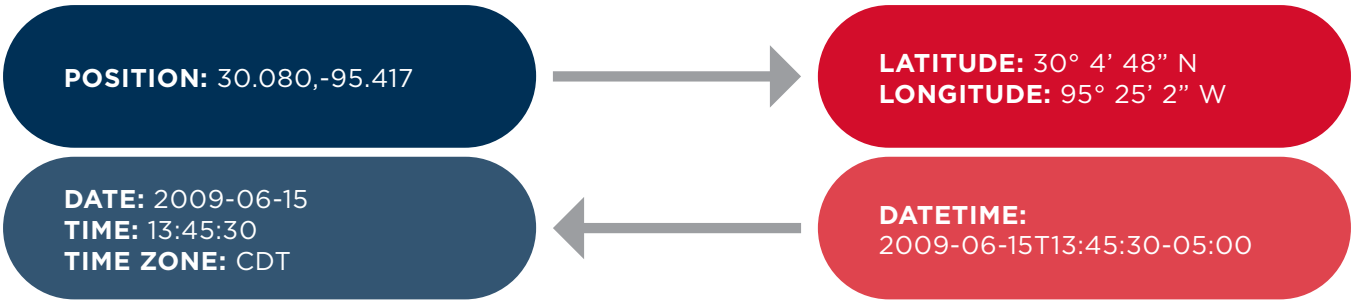


Figure 15: One-to-many crosswalk mapping

Crosswalks are also applicable to the conceptual domain as well as the value domain. A metadata mapping procedure (MMP) is established in ISO/IEC TR 20943-5 [14] for creating crosswalks for metadata elements. It provides a more detailed method for mapping metadata when direct mapping between two or more systems is not possible. The MMP is designed to be employed by metadata registries to ensure content consistency and interoperability. Data element concept mapping is divided into three processes: i) identifying metadata element sets, ii) grouping data element concepts, and iii) mapping data element concepts. The three main processes are summarized below.

IDENTIFYING METADATA ELEMENT SETS

The objective of the first process is to identify the metadata element sets to be processed and the fields to be mapped. The availability of sample data is also to be checked and the organizational authority for each metadata element set is identified.

GROUPING DATA ELEMENT CONCEPTS

The second process groups data element concepts by class and then subgroups the data element by property. It is recommended that the metadata element set with the least number of elements be selected as the primary set for the grouping process. Any data elements that cannot be grouped may be set aside.

MAPPING DATA ELEMENT CONCEPTS

The final mapping data element concept process involves creating common Data Element Concepts (DEC) for all the metadata element sets being mapped. The common DEC creation process is guided by ISO/IEC 11179-1 and ISO/IEC 11179-3. The candidate data element concepts are organized in tabular form according to common DEC. Each data element concept in the mapping can be assigned a type of heterogeneity to denote how it differentiates from the common DEC.

The conceptual domain mapping occurs after the data element conceptual mapping has been completed. The conceptual domain mapping involves conversions and recompositions. Conversions adapt the format of metadata elements. Conversions may be numeric as in the case of unit conversions. Coding conversions account for the use of different code sets for a common conceptual domain. An example of a coding conversion for flag State may be USA to United States. Recompositions can either be compositions, decompositions or order changes. An example of a composition for a date is Day: 10, Month: June, and Year: 2019 to Date: "10 June 2019." Similarly, a date order change example is DDMMYY to MMDDYY format. Following the MMP, sample data sets should be checked to test the process.

DATA LINEAGE AND IMPACT ANALYSIS

Data lineage analysis documents how data is acquired or created, how it is transformed as it moves from its source to its destination, and how it is used. The benefit of performing a data lineage analysis is that the resulting documentation enables the impact of changes to a system to be quickly assessed and the time to investigate anomalies is reduced. The documentation of metadata during the design and implementation of smart and autonomous functions supports data lineage analysis. Data lineage analyses are commonly manifested in lineage flow diagrams that allow the movement of data to be visualized. Linked metadata directly supports the creation of the flow diagrams.

Smart and autonomous functions are systems and depending on the degree of complexity may have one or more levels of subsystems. The design of these systems often involves the creation of functional block diagrams. Such diagrams lend themselves to the definition of an 'As Designed Lineage," however those models may not allow the extraction of actual implementation details. For complex systems the completion of subsystems is often delegated to specialized teams. The data lineage analysis will need to stitch together details from all the sources to create an "As Implemented Lineage."

As the size the subject system increases the complexity of the metadata grows as more and more data elements are used. Software tools may be needed to manage changes to the data lineage analysis and automate the creation of flow diagrams for large systems. Once a data lineage flow diagram has been documented it is possible to transverse the diagram in multiple directions. Two different approaches have been considered for different focus groups [12]:

BUSINESS FOCUS

For a business focus, prioritized studies are performed and are driven by business deliverables. The business focus approach is to trace backwards from targeted data elements to identify source systems. The business focus concentrates only on exchanges or locations where data elements are modified to identify points where data quality can be assessed. The business focus approach would be appropriate if user feedback indicates a problem with one function of the system. The trace back data lineage analysis will provide a means for helping to identify the root cause.

TECHNICAL FOCUS

The technical focus approach starts at the source systems and identify data consumers and subsequent consumers until the network of all systems impacted are identified. This approach allows the technical team to understand paths users will take to discover data. For smart or autonomous systems, the technical focus approach enables impact analyses to be performed. For example, the impact of switching suppliers for a component could be assessed. The new component may have a different number or type of embedded sensors that may make it incompatible with existing data elements.

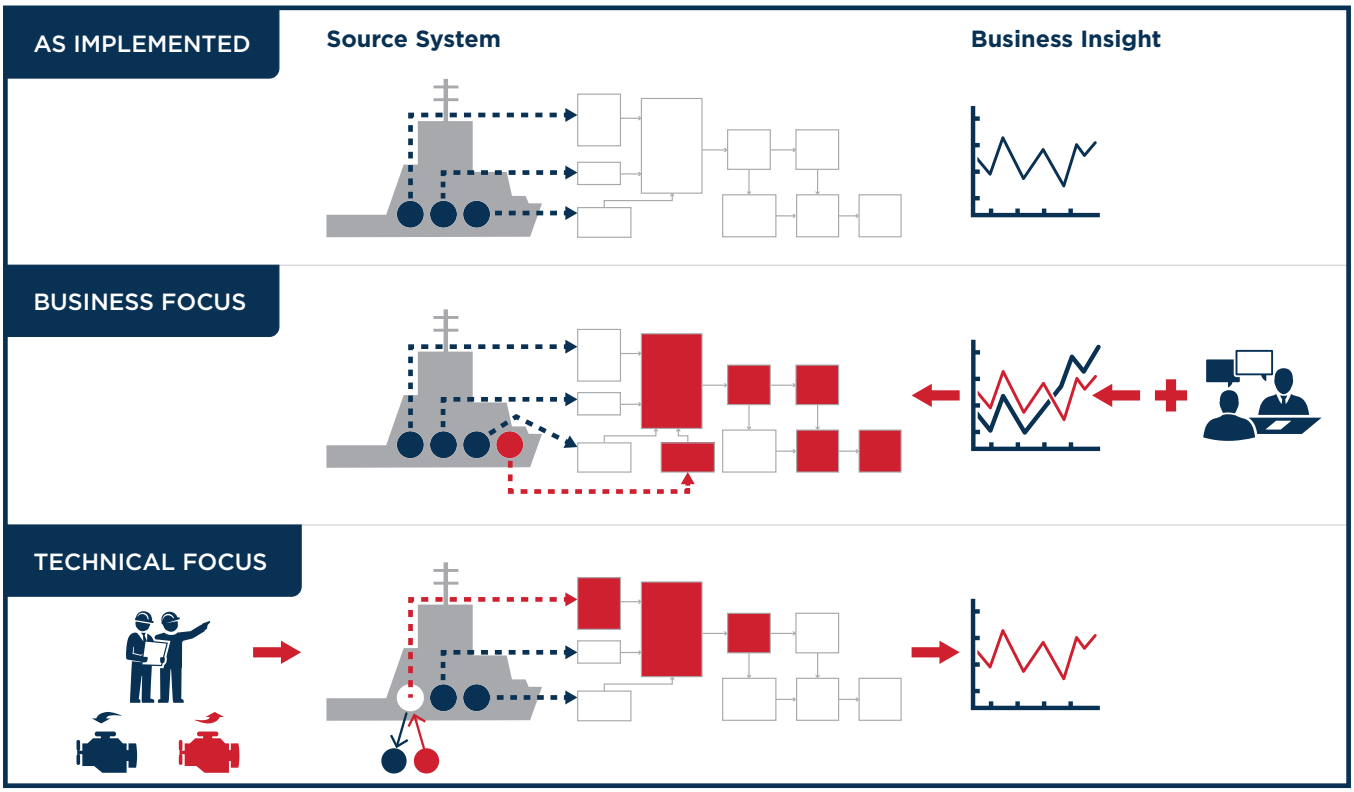


Figure 16: Data lineage analysis

METADATA QUALITY METRICS

Metadata quality can be assessed in various ways. Qualitative checks are often tasked to metadata experts to evaluate subjective details. Experts can confirm if schema are being properly adhered to and determine if items are being categorized correctly. Quantitative checks allow opportunities for automation and the quantified quality indices support statistical analysis and the monitoring of trends over time. In practice both qualitative and quantitative assessments should be performed. This section focuses on quantitative metrics for evaluating metadata quality.

A set of seven quantitative metadata metrics were proposed by Ochoa and Duval [19]. The metrics were designed to check the quality of IEEE Learning Object Metadata (LOM) stored in repositories. The following metrics were proposed:

1. Completeness
2. Accuracy
3. Provenance
4. Conformance to Expectations
5. Logical Consistency and Coherence
6. Timeliness
7. Accessibility

Of the seven metrics, the completeness and accuracy metrics are outlined in detail because they have the most universal application for transactional and time-series data and are straightforward to implement. Reiche and Höfig [16] tested five of the metrics using three public government data repositories and provide further implementation details including challenges.

COMPLETENESS

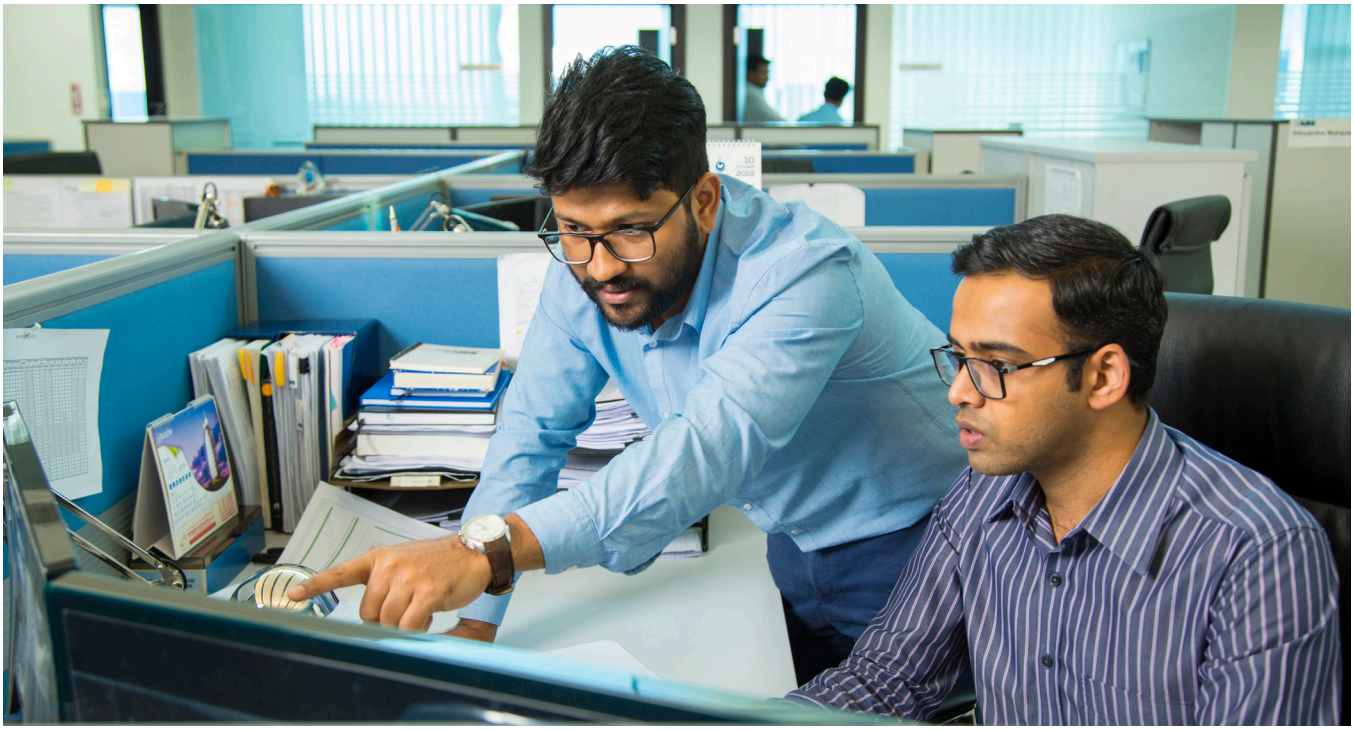
The completeness metric considers the number of completed fields in a data record. A metadata record is considered complete if all the required information needed to describe an object is present. Completeness is important for transactional data. A basic approach to assessing completeness is to count the number of non-empty fields and divide by the total number of fields to obtain a completeness percentage.

$$Q_{completeness} = \frac{\sum_{i=1}^{N} P(i)}{N}$$

The completeness metric is easy to implement although one detriment is that it considers all metadata fields equally. From a quality perspective a field for vessel ID number may have higher relevance than vessel breadth because it is more common for a user to search for a record by vessel ID than breadth. To address this concern the completeness metric can be tailored to a specific data record to improve its usefulness by introducing weighting factors for the metadata fields.

$$Q_{completeness} = \frac{\sum_{i=1}^{N} \alpha_i * P(i)}{\sum_{i=1}^{N} \alpha_i}$$

A completeness check may be more complicated if some fields are conditional based on entries in other fields. For example, a smart function for a cargo system may generate additional metadata during loading and offloading than when in transit. The additional metadata would only be expected when the field for system status indicates that cargo loading, or offloading was occurring. In such a situation additional rules or logic may need to be applied for the metric to provide a truer representation of the quality of the record.



ACCURACY

The accuracy of a metadata field concerns the degree to which the entered value corresponds to the actual value of the object it describes. The accuracy metric assesses the quality of the metadata by using the distance between the entered value and the correct value. Shorter distances imply greater accuracy. Fields with numeric or date values are well suited for obtaining the distance measurements needed to evaluate accuracy.

$$Q_{accuracy} = 1 - \frac{\sqrt{\sum_{i=1}^{N} d\,(field_i)^2}}{\sum_{i=1}^{N} d\,(field_i)} \quad \left(\text{given that } \sum_{i=1}^{N} d\,(field_i) > 0\right)$$

A difficulty with implementing an accuracy metric is identifying functions for calculating distance measurements for some fields. For time series data the correct data value may vary with time, however accuracy metrics can be evaluated for known conditions. For example, after filling a tank the data values can be compared to known tank capacity. Similarly, temperature measurements from a system during long idle periods can be compared to ambient values. Relative accuracy assessments may be performed when multiple data sources are available. Applying an accuracy metric to timestamps from two separate systems would help identify synchronization problems or offset errors such as improperly set the time zones.

Metadata may also be used to support quality checks. Smart, autonomous and remote-control functions often require prescribed periodic monitoring by humans for risk reduction. Use metadata can be used to verify the monitoring takes place as scheduled. Alarm and fault log histories can be used to compare operational performance to predictions made during the development of risk matrices.

ABS ROLE

As a classification society, ABS has published the *Guidance Notes on Smart Function Implementation* [2] (the Smart Guidance Notes) to help guide data-centric marine and offshore applications for structural and machinery health monitoring, asset efficiency monitoring, operational performance management and crew assistance and augmentation. These Smart Guidance Notes provide an actionable goal-based framework for smart function implementation along with recommended risk-informed verification and validation principles to provide confidence in the ability to conform to the functional requirements (see Figure 17).

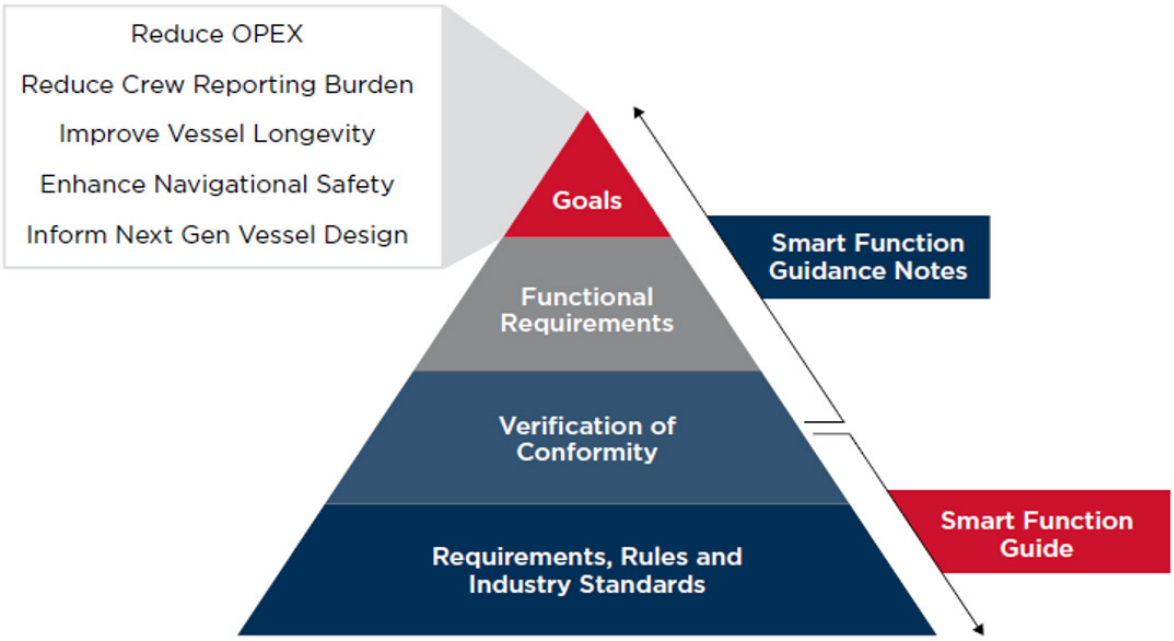


Figure 17 Goal-based smart function framework

ABS has also published the *Guide for Smart Functions for Marine Vessels and Offshore Units* [5] (the Smart Guide) to define the risk-informed technical requirements for independently assessing the Smart Function system(s) and offer optional SMART class notations and Class Record Comments (refer to Figure 17 for the coverage of the Guidance Notes and the Guide).

The ABS *Guide for Autonomous and Remote Control Functions* [6] (the Autonomous and Remote-Control Guide) provides a goal-based framework for the implementation of autonomous and remote control functions. The goal-based framework of the Autonomous and Remote-Control Guide is similar to the framework for smart functions. Optional class notations are available provided the functions satisfy the requirements of the risk-based approval process.



The metadata guidance and recommendations offered in this advisory promote the achievement of function goals for smart, autonomous and remote-control functions. The Smart Guide and Autonomous and Remote-Control Guide outline submittal requirements that require documentation of system architecture, data integrity validation and data quality assessments. The contents of this advisory are designed to support creation of those documents. Notable applications are detailed below.

- The definition of sensor configurations and mapping specifications are part of the smart function system specification required per section A3/6 of the Smart Guide. Metadata definitions such as sensor IDs and units of measure are included in the system specification. Similar criteria applies for autonomous and remote-control functions (Section 1/12.2.iii.b of the Autonomous and Remote-Control Guide). The use of standards such as BS ISO 19847 and BS ISO 19848, summarized in the metadata standards section of this advisory, to generate the required metadata definitions will support interoperability and content consistency.
- The system architecture of a smart function is to be detailed per section A3/5 of the Smart Guide. A description of the inputs and outputs at system boundaries is a typical part of a system architecture. Development of metadata crosswalks at those integration/interface points as outlined in the metadata tools section of this advisory will help ensure data exchanges at the system boundary meet the function's semantic and syntactic data requirements. Metadata crosswalks also apply internally for autonomous and remote-control functions that are systems-of-systems. Development of crosswalks at subsystem interfaces supports the creation of systems-of-systems architectures per sections 5/2.5 and 6/2.5 of the Autonomous and Remote-Control Guide.
- Management of Change (MoC) programs are required for autonomous and remote-control functions (Section 1/12.1 of the ABS Autonomous and Remote-Control Guide) and high-risk smart functions (Section A3/8 of the Smart Guide). Use of the data lineage and impact analysis techniques presented in the metadata tools section of this advisory help document the consequences and risks to system changes that MoC programs seek to capture.
- The submission of a data quality assessment (DQA) is required for medium and high-risk smart functions (Section A3/7.iii of the Smart Guide) and per section 1/12.2.iii.c of the Autonomous and Remote-Control Guide. The metadata quality metrics detailed in the metadata quality section of this advisory enable a system/equipment level DQA per section A4/5 of the Smart Guide to be performed on data considered as metadata.
- Traceability is a necessary attribute for software used by all smart, autonomous and remote-control functions. High-risk smart functions and medium/high risk autonomous and remote-control functions must also meet criteria outlined in the ABS *Guide for Integrated Software Quality Management (ISQM)*[1]. When metadata registries are used, following standards such as ISO/IEC 11179 introduced in the metadata registries section of this advisory supports metadata traceability.

SUMMARY

Metadata's role in describing and defining other data makes it an essential element of smart, autonomous and remote-control functions that rely on data to support decisions by humans or systems. The application of metadata will be further advanced as the amount of data collected on board vessels increases, and it becomes important to comprehensively describe and uniquely identify the data to enable efficient discovery. Metadata also supports data exchanges and details system usage.

An overview of the types of metadata and relevant standards was presented to provide conceptual background. Three standards provide a foundation of a framework for implementing smart, autonomous and remote-control functions:

- BS ISO 19848 provides a standard for data from equipment and machinery
- BS ISO 19847 provides requirement for onboard data servers used to collect data from onboard systems, store it and send it off the vessel
- ISO/IEC TR 20943-5 provides a metadata mapping procedure to enable interoperability

A metadata domain model for the function will identify metadata descriptions and relationships not covered by BS ISO 19847 and BS ISO 19848 that will need to be implemented. Use of ISO/IEC TR 20943-5 will ensure the interoperability of data exchanges with other systems or subsystems.

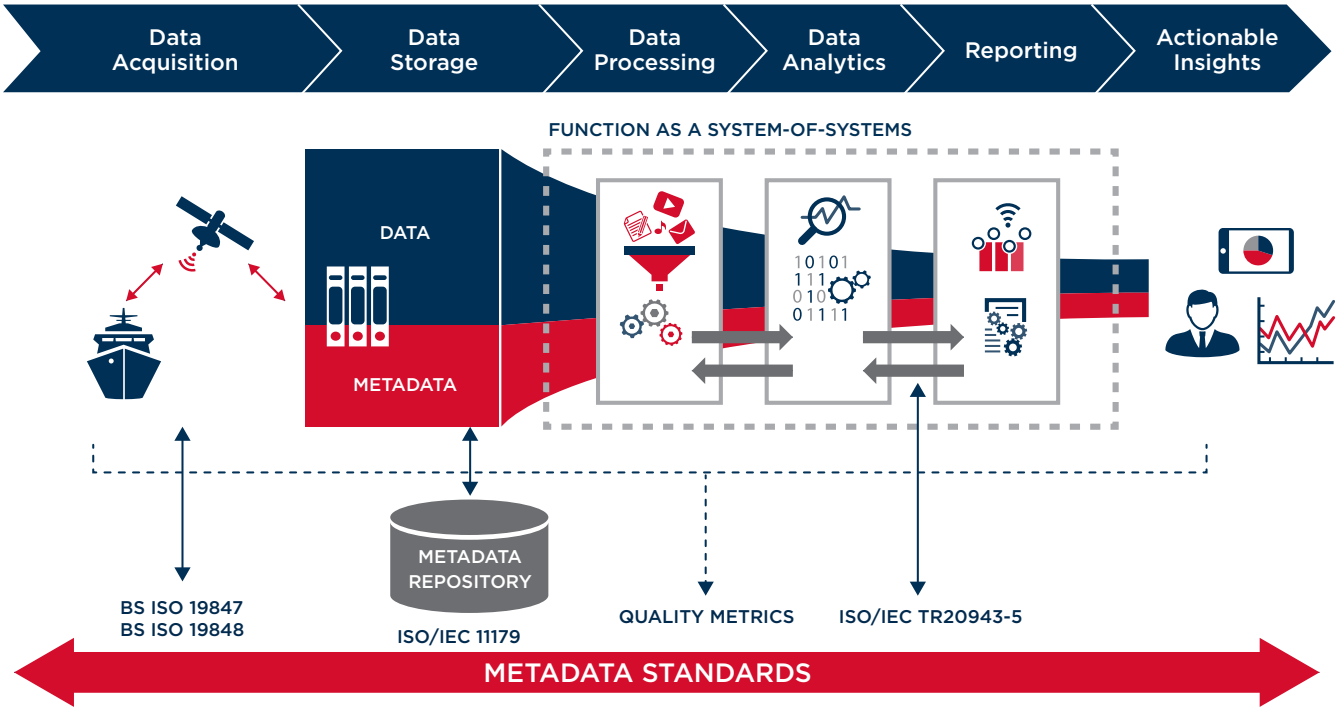


Figure 18: Application of metadata and metadata standards for smart, autonomous and remote-control functions

Metadata registries are databases for storing metadata schemas, data elements and vocabularies to promote their reuse and standardization over time. Standard ISO/IEC 11179 provides a framework for metadata registries. Three different architecture models for metadata registries were presented: centralized, distributed and hybrid.

The expansion of the Internet of Things (IoT) in the marine and offshore industry will continue to increase connectivity among machines. The adoption of linked metadata for resources on the worldwide web provides insight into the types of metadata usage scenarios that may develop in the marine and offshore industry. The number of domain specific standards for the marine and offshore industry is limited but likely to increase as new technologies are more widely adopted.

REFERENCES

1. ABS. (2012). Guide for Integrated Software Quality Management (ISQM)
2. ABS. (2018). Guidance Notes on Smart Function Implementation.
3. ABS. (2019). Advisory on Data Quality for Marine and Offshore Application – Time Series Data.
4. ABS. (2020). Advisory on Data Quality for Marine and Offshore Application – Transactional Data.
5. ABS. (2022). Guide for Smart Functions for Marine Vessels and Offshore Units.
6. ABS. (2021). Guide for Autonomous and Remote-Control Functions
7. BS ISO 19847:2018. (2018). Ships and marine technology – Shipboard data servers to share field data at sea.
8. BS ISO 19848:2018. (2018). Ships and marine technology – Standard data for shipboard machinery and equipment.
9. DCMI Usage Board. (2012). Dublin Core™ Metadata Element Set, Version 1.1: Reference Description. URL: https://dublincore.org/specifications/dublin-core/dces/
10. ISO 16684-1. (2011). Graphic Technology - Extensible Metadata Platform (XMP) Specification: Part 1, Data Model, Serialization, and Core Properties.
11. Richard Cyganiak, David Wood, Markus Lanthaler. (2014). RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation, 25 February 2014. URL: http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/. The latest edition is available at http://www.w3.org/TR/rdf11-concepts/
12. DAMA International. (2017). DAMA-DMBOK 2nd Edition. Technics Publications. Basking Ridge, NJ.
13. ISO/IEC 11179-1: 2015. (2015). Information Technology – Metadata registries (MDR) – Part 1: Framework
14. ISO/IEC TR 20943-5:2013. (2013). Information Technology – Procedures for achieving metadata registry content consistency – Part 5: Metadata mapping procedure.
15. Jeffrey Pomerantz. (2015). Metadata. The MIT Press. Cambridge, MA.
16. Konrad Johannes Reiche, Edzad Höfig. (2013). Implementation of Metadata Quality Metrics and Application on Public Government Data. IEEE. 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops [Online].
17. Marcia Lei Zeng, Jian Qin. (2015). Metadata 2nd Edition. American Library Association. Chicago, IL.
18. Open Class 3D Exchange (OCX) Consortium. (2022). URL: https://3docx.org/
19. Xavier Ochoa, Erik Duval. (2006). Quality Metrics for Learning Object Metadata. AACE. World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006 [Online].

CONTACT INFORMATION

GLOBAL SUSTAINABILITY CENTER
1701 City Plaza Dr.
Spring, Texas 77389, USA
Tel: +1-281-877-6000
Email: Sustainability@eagle.org

NORTH AMERICA REGION
1701 City Plaza Dr.
Spring, Texas 77389, USA
Tel: +1-281-877-6000
Email: ABS-Amer@eagle.org

SOUTH AMERICA REGION
Rua Acre, n° 15 - 11° floor, Centro
Rio de Janeiro 20081-000, Brazil
Tel: +55 21 2276-3535
Email: ABSRio@eagle.org

EUROPE REGION
111 Old Broad Street
London EC2N 1AP, UK
Tel: +44-20-7247-3255
Email: ABS-Eur@eagle.org

AFRICA AND MIDDLE EAST REGION
Al Joud Center, 1st floor, Suite # 111
Sheikh Zayed Road
P.O. Box 24860, Dubai, UAE
Tel: +971 4 330 6000
Email: ABSDubai@eagle.org

GREATER CHINA REGION
World Trade Tower, 29F, Room 2906
500 Guangdong Road, Huangpu District,
Shanghai, China 200000
Tel: +86 21 23270888
Email: ABSGreaterChina@eagle.org

NORTH PACIFIC REGION
11th Floor, Kyobo Life Insurance Bldg.
7, Chungjang-daero, Jung-Gu
Busan 48939, Republic of Korea
Tel: +82 51 460 4197
Email: ABSNorthPacific@eagle.org

SOUTH PACIFIC REGION
438 Alexandra Road
#08-00 Alexandra Point, Singapore 119958
Tel: +65 6276 8700
Email: ABS-Pac@eagle.org

© 2022 American Bureau of Shipping.
All rights reserved.

ABS