GUIDANCE NOTES
GD06-2021

INTERNATIONAL SHIP CLASSIFICATION

Guidelines for Quality Assessment of Ship Data

2021

Effective from 1 April 2021

CONTENTS

Chapter 1 General

Section 1 General Provision

1.1.1 General requirements

1.1.1.1 The Guidelines apply to data quality assessment of shipborne systems (which may be taken as a reference by offshore installations).

1.1.1.2 The Guidelines provide the framework and methods for ship data quality assessment. Technical requirements for data quality to meet the business needs are developed based on data identification and analysis.

1.1.1.3 Any risk that might exist in data is analyzed to specify relevant security requirements, which are not limited to technical means or management measures.

1.1.1.4 The data should be identified according to certain coding rules and semantics.

Section 2 Scope of Application

1.2.1 Scope of data

1.2.1.1 The Guidelines apply to the data retained in the process of collection, storage, processing, transmission, receiving and sharing in the shipborne systems.

1.2.1.2 The Guidelines may apply to the data generated and applied throughout the life cycle of the ship, including but not limited to data generated and used in the design, construction, operation and ship recycling processes, as well as the data generated and used in the process of equipment access, platform operation, system APP application, etc. of the ship.

1.2.1.3 Data type scope of the Guidelines:

(1) Structured data (database, Key-Value storage system, etc.);

(2) Unstructured data (document, image, audio, video, etc.).

Section 3 Assessment Report and Class Notations

1.3.1 Shipborne systems/products

1.3.1.1 An assessment report is issued to shipborne system products upon application, subject to review and assessment by ISC.

1.3.2 Ships

1.3.2.1 Ships are assigned the following class notations upon application, subject to satisfactory review and assessment by ISC:

Data Quality（x）

Where: x represents the shipborne system assessed by the Guidelines

1.3.3 Application

1.3.3.1 For the systems and/or ships applying for ship data quality assessment by ISC, a written application should be filed to ISC or its designated unit or local branch, and an assessment service contract and/or agreement may be signed when necessary.

1.3.4 Assignment, maintenance, suspension, cancellation and reinstatement of class notations

1.3.4.1 The assignment, maintenance, suspension, cancellation and reinstatement of class notations for ship data quality should be according to the provisions of Section 9, Chapter 2, PART ONE of ISC *Rules for Classification of Sea-going Steel Ships*.

Section 4 Terms and Normative References

1.4.1 Terms

1.4.1.1 Architecture: (system) fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution.

1.4.1.2 Attribute: inherent property or characteristic of a target entity that can be distinguished quantitatively or qualitatively by human or automated means.

1.4.1.3 Availability: loss of availability due to accidental or unauthorized destruction of information or interruption of access to or use of the information system.

1.4.1.4 Capability area: a collection of data management related activities, processes, etc., and a collection of related data capability sub-domains.

1.4.1.5 Confidentiality: loss of confidentiality due to accidental or unauthorized disclosure of information.

1.4.1.6 Data: reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.

1.4.1.7 Data architecture: specification that defines data requirement through organization-level data model, guides distribution control and integration of data assets, deploys data sharing and application environment, as well as metadata management.

1.4.1.8 Data asset: data resources owned and controlled by the organization that can generate benefits.

1.4.1.9 Data classification: differentiation and grouping according to certain principles and methods and based on the attributes, characteristics, source, content, etc. of the data.

1.4.1.10 Data dictionary: collection of information about data such as name, description, creator, owner, provenance, translation in different languages, and usage.

1.4.1.11 Data file: set of related data records treated as a unit.

1.4.1.12 Data format: arrangement of data for storage or display.

1.4.1.13 Data governance: the procedures of processing, formatting and standardizing data.

1.4.1.14 Data grading: classification of data by level according to its importance, sensitivity, etc. to take differentiated security measures for the data.

1.4.1.15 Data item: smallest identifiable unit of data within a certain context for which the definition, identification, permissible values, and other information is specified by means of a set of properties.

1.4.1.16 Data life cycle: the evolution process of various survival forms including data acquisition, storage, integration, analysis, application, presentation, archiving and destruction.

1.4.1.17 Data management capability: the capability of an organization or institution to manage and apply data.

1.4.1.18 Data management capability maturity assessment model: model used to assess the data management capability maturity of an organization.

1.4.1.19 Data model: graphical and textual representation of analysis that identifies the data needed by an organization to achieve its mission, functions, goals, objectives, and strategies and to manage and rate the organization.

1.4.1.20 Data quality: degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions.

1.4.1.21 Data quality characteristic: category of data quality attributes that bears on data quality.
1.4.1.22 Data quality measure: variable to which a value is assigned as the result of measurement of a data quality characteristic.
1.4.1.23 Data quality model: defined set of characteristics which provides a framework for specifying data quality requirements and evaluating data quality.
1.4.1.24 Data record: set of related data items treated as a unit.
1.4.1.25 Data security: keeping and maintaining the confidentiality, integrity and availability of data, which may also include authenticity, verifiability, non-repudiation, reliability, etc.
1.4.1.26 Data standard: rules for naming, definition, structure and evaluation of data.
1.4.1.27 Data strategy: vision, purpose, goals and principles for the organization to conduct data work.
1.4.1.28 Data type: categorization of an abstract set of possible values, characteristics, and set of operations for an attribute.
1.4.1.29 Data value: content of data item.
1.4.1.30 Element: smaller part of an architecture.
1.4.1.31 Integrity: loss of integrity due to accidental or unauthorized modification of information.
1.4.1.32 Master data: basic information that meets the needs of cross-departmental business collaboration and reflects the core business entity status attributes.
1.4.1.33 Measurement: set of operations having the object of determining a value of a measure.
1.4.1.34 Measurement function: algorithm or calculation performed to combine two or more quality measure elements.
1.4.1.35 Metadata: data that describe other data.
1.4.1.36 Presentation device: device used to present data to the intended user of a system.
1.4.1.37 Quality measure: measure that is defined as a measurement function of two or more values of quality measure elements.
1.4.1.38 Quality measure element: measure defined in terms of a property and the measurement method for quantifying it, including optionally the transformation by mathematical function.
1.4.1.39 Semantics: meaning of the syntactic components of a language.
1.4.1.40 Target entity: fundamental thing of relevance to the user, about which information is kept, and need to be measured.
1.4.1.41 Vocabulary: collection of information related to a specific subset of terms related to a specific domain.

1.4.2 Normative references

1.4.2.1 The clauses in relevant documents will become the clauses of this specification after being taken as a reference, see Table 1.4.2. For dated references, only the dated edition applies to the Guidelines. For undated references, the latest edition of the referenced document (including any amendments) applies to the Guidelines.

Table 1.4.2 References

No.	Number	Name
1		ISC Rules for Classification of Sea-going Steel Ships (2018) and its Amendments
2	IACS Rec 166	Recommendation on Cyber Resilience
3	ISO/IEC 25012:2013	Software engineering—Software product Quality Requirements and Evaluation (SQuaRE)—Data quality model
4	ISO/IEC 25024:2013	System and Software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) Measurement of data quality
5	ISO 8000-8:2015	Data quality- Part 8: Information and data quality Concepts and measuring

Chapter 2 Data Classification and Grading

Section 1 General Provisions

2.1.1 General requirements

2.1.1.1 This Section applies to ship data classification and grading by the shipowner, owner, and ship management company.

2.1.1.2 The Guidelines are not applicable to the data involving state secret information, and the provisions of laws and regulations of flag States and Administrations should be followed.

2.1.1.3 The classification and grading of ship data should be designed to improve the ship data management capability, and ensure safe, efficient and reasonable application of data.

2.1.2 Principles of data classification and grading

2.1.2.1 The data may be classified and graded according to the following principles:

(1) Scientific, it refers to scientific and systematic classification according to the multi-dimensional characteristics of data and their logical relationship;

(2) Stability, it refers to develop a classification and grading scheme based on the most stable characteristics and attributes of data;

(3) Practicability, data should be available under each classification, without meaningless categories, and data categories must conform to the general understanding of data classification;

(4) Extensibility, overall, the data classification and grading scheme should be general and inclusive, be able to classify and grade various types of data, and meet the classification and grading requirements of data that may appear in the future;

(5) Applicability, it refers to not only meet the data quality requirements and data security requirements, but also meet the actual situation of the ship;

(6) Sustainability, the data classification, and grading methods should be adaptable to the development of technology.

2.1.3 Procedures of data classification and grading

2.1.3.1 It is recommended that data classification and grading requirements should be developed based on the business requirements for shipborne system, including but not limited to:

(1) Data classification methods and guidelines;

(2) Detailed list of data classification, including the initial security level of each type of data;

(3) Security requirements for data of different security levels.

2.1.3.2 It is recommended that data classification and grading should be made according to the procedures in Figure 2.1.3.2. The organization should classify the data according to the standards and requirements for data classification and grading; set the initial security level for the classified data; assess whether the initial security level meets the ship data security requirements after comprehensive analysis of business, security risks, security measures and other factors, adjust inappropriate data classification and determine the final security level of data.

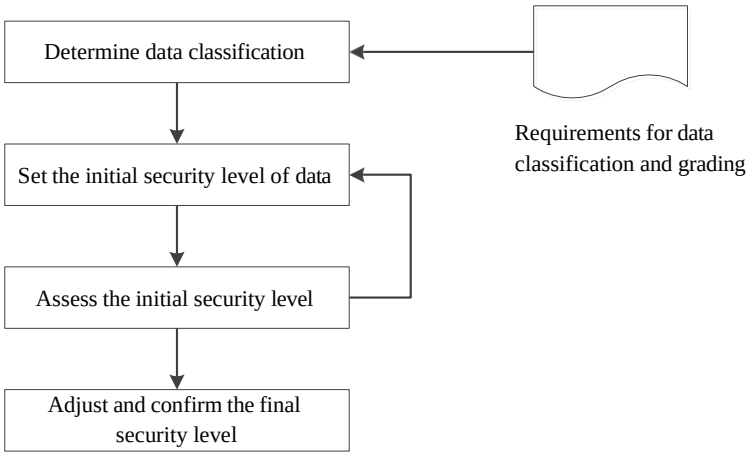

Figure 2.1.3.2 Steps of Data Grading

Section 2 Data Classification

2.2.1 General requirements

2.2.1.1 Ship data may be classified by the shipowner, owner, and ship management company according to the data asset management form, combined with the construction mode and operation mode. For data classification, the business processes and system equipment need to be analyzed and sorted out first, and data are classified and identified considering the actual situations such as industry requirements, business scale, and data complexity, in order to generate a data classification list.

2.2.2 Data classification method

2.2.2.1 Data classification may be made according to (but not limited to) the following methods:

(1) Classification related to ship life cycle, such as R&D data (R&D design data, development testing data, etc.), production and construction data (control information, working conditions, process parameters, system logs, etc.), operation and maintenance data (cargo data, service data, etc.), management data (asset information, freight information, supply chain data, business statistics, etc.), external data (data shared with other subjects, etc.);

(2) Classification related to the safety of ship, personnel, and environment, such as manipulation data, operation data, analysis data, etc.;

(3) Business-related classification, such as ship control data, data of interaction between the ship and the outside world, ship cargo data, ship maintenance data, etc.;

(4) Classification by the nature of data, such as positioning data (various coordinate data, direction data), qualitative data (data on the attributes of things), quantitative data (data on the quantitative characteristics of things, including geometric quantities including length, area, volume, and physical quantities including weight and speed), timing data (data on time characteristics, including year, month, day, hour, minute, second, etc.);

(5) Classification by representation, such as digital data (discrete values in a certain interval like symbol and text), analog data (continuous values in a certain interval, such as sound, image);

(6) Classification by recording medium and method, such as papers (voyage records, sea chart, meteorological data, etc.), tables, images, tapes;

(7) Classification by digital method, such as vector data, raster data, etc.

2.2.2.2 Ship data classification should be adapted to the construction and development of shipborne system.

Section 3 Data Grading

2.3.1 General requirements

2.3.1.1 Data grading rules should be determined based on security risks.

2.3.1.2 In the Guidelines, data grading is made from three aspects of the loss of data confidentiality, integrity or availability based on potential impacts on ship, shipowner, ship management company, and stakeholders.

2.3.2 Grading method

2.3.2.1 After falsification, destruction, disclosure or illegal use of ship data, the data can be divided into three levels from three aspects of the loss of data confidentiality, integrity, or availability—Level 1, Level 2, and Level 3 based on the potential impacts on ship, shipowner, owner, ship management company, and stakeholders, among them, Level 3 is the highest security level.

2.3.2.2 The loss of confidentiality, integrity or availability may have limited adverse effects on personnel safety, ship safety and/or threats to the environment, and the data that meet one of the following conditions are Level 1 data:

(1) The ship operating performance is reduced to a certain extent, and to a certain degree, the duration for the organization to perform its main functions lasts, but the effectiveness of the function is significantly reduced;

(2) Minor injuries to individuals.

2.3.2.3 The loss of confidentiality, integrity or availability may have serious adverse effects on personnel safety, ship safety and/or threats to the environment, and data that meet one of the following conditions are level 2 data:

(1) To a certain extent, serious deterioration of ship operation is caused until the degree and duration that the organization can perform its main functions, but the effectiveness of the function is greatly reduced;

(2) Significant injury to individuals, which does not involve loss of life or serious life-threatening injuries.

2.3.2.4 The loss of confidentiality, integrity or availability may have catastrophic effects and potential impact on ships, personnel and the environment, and the data that meet one of the following conditions are level 3 data:

(1) Severe ship operation degradation or loss to the degree and duration that the organization fails to perform one or more of its main functions;

(2) Serious or catastrophic injury to individuals, involving loss of life or serious life-threatening injuries.

Chapter 3 Data Quality Assessment

Section 1 Ship Data Quality Assessment Framework

3.1.1 General requirements

3.1.1.1 The Guidelines are designed to evaluate 15 characteristics of data quality based on semantic, syntactic, and pragmatic qualities from the inherent and system dependent points of view.

3.1.1.2 Inherent data quality refers to the degree to which quality characteristics of data have the intrinsic potential to satisfy stated and implied needs when data is used under specified conditions.

3.1.1.3 From the inherent point of view, data quality refers to data itself, to:

(1) Data domain values and possible restrictions (e.g., business rules governing the quality required for the characteristic in each application);

(2) Relationship of data values (e.g., consistency);

(3) Metadata.

3.1.1.4 System dependent data quality refers to the degree to which data quality is reached and preserved within a computer system when data is used under specified conditions. From this point of view, data quality depends on the technological domain in which data are used; it is achieved by the capabilities of computer systems' components such as: hardware devices (e.g., to make data available or to obtain the required precision), computer system software (e.g. backup software to achieve recoverability), and other software (e.g. migration tools to achieve portability).

3.1.1.5 Based on ISO 8000-8 standard, the Guidelines divide the data quality into three categories:

(1) Syntactic quality: the degree to which data conforms to its specified syntax;

(2) Semantic quality: correspondence of data to the entities represented;

(3) Pragmatic quality: degree of conformance to usage-based requirements.

3.1.2 Data quality model

3.1.2.1 The data quality model defined in the Guidelines outlines the 15 quality characteristics from two points of view.

3.1.2.2 The existence of an "X" symbol in Table 3.1.2.2 indicates the relevance of the characteristics for the data quality required or evaluated from inherent and/or system dependent points of view and as a result of its measurability.

Table 3.1.2.2 Data Quality Model Characteristics

Characteristics	Data Quality	
	Inherent	System Dependent
Accuracy	×	
Completeness	×	
Consistency	×	
Credibility	×	
Currentness	×	
Accessibility	×	×
Compliance	×	×
Confidentiality	×	×
Efficiency	×	×
Precision	×	×
Traceability	×	×

Understandability	×	×
Availability		×
Portability		×
Recoverability		×
Note 1: Some characteristics are relevant from both points of view. Note 2: Data quality characteristics will be of varying importance and priority to different stakeholders.		

3.1.3 Data quality assessment framework

3.1.3.1 To meet the needs of different levels of data quality assessment, a data quality assessment framework is established, as shown in Figure 3.1.3.1.

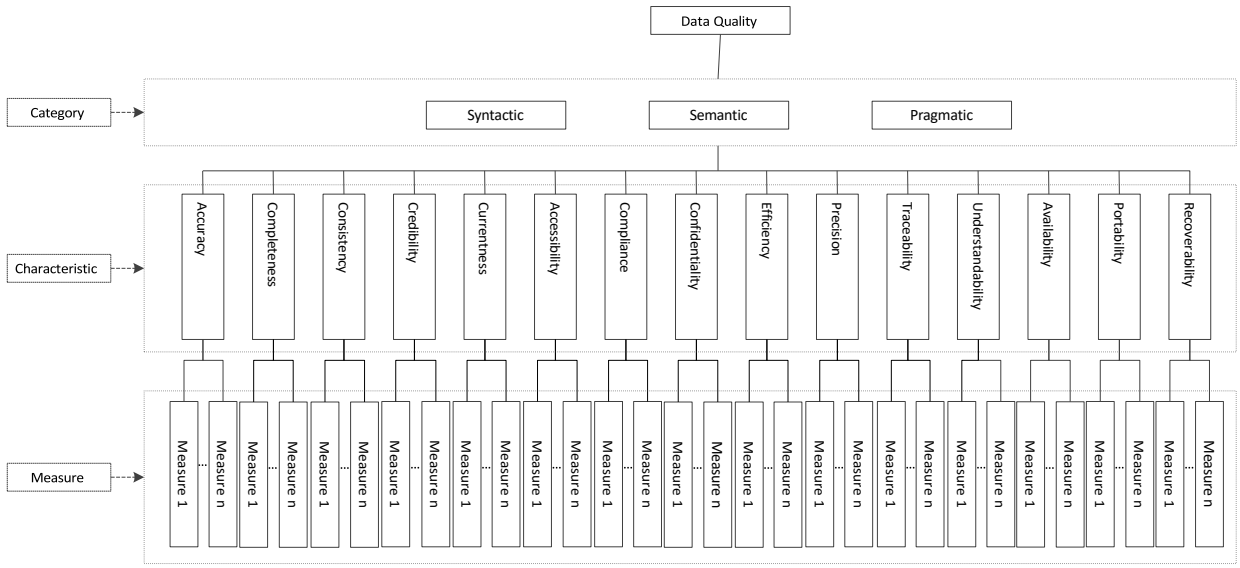

Figure 3.1.3.1 Data Quality Assessment Framework

Section 2 Data Quality Characteristics and Measures

3.2.1 QMs for accuracy

3.2.1.1 Accuracy measures provide the degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use.

Table 3.2.1.1 Accuracy Measures

ID	Name	Description	Measurement Function	Point of View
Acc-I-1	Syntactic data accuracy	Ratio of closeness of the data values to a set of values defined in a domain.	$X = {}^{A}/_{B}$ A = number of data items which have related values syntactically accurate. B = number of data items for which syntactic accuracy is required.	Inherent
Note 1: A single value is considered "syntactically accurate" when it is the same as one from an identified source of validated information: the result is "yes" or "no". Note 2: An example of a low degree of syntactic accuracy is when the word Mary is stored as Marj.				

Acc-I-2	Semantic data accuracy	Ratio of how accurate the data values in terms of semantics in a specific context are.	$X = A/B$ A = number of data values semantically accurate. B = number of data values for which semantic accuracy is required.	Inherent
Note 1: A single value is considered "semantically accurate" when the meaning (the content) corresponds to the reality. Note 2: An example of a low degree of semantic accuracy is when the name of John is recorded instead of George; both names are syntactically accurate, so only George is semantically accurate.				
Acc-I-3	Data accuracy assurance	Ratio of measurement coverage for accurate data.	$X = A/B$ A = number of data items measured for accuracy. B = number of data items for which measurement is required for accuracy.	Inherent
Note: This measure is relevant to control data quality, if applied on the raw data (especially when a software program for data error attenuation is not available). This QM does not measure the quality of the data, but it measures the thoroughness and application of the accuracy measures. It is a measure of the attention given to the accuracy matter.				
Acc-I-4	Risk of data set inaccuracy	The number of outliers in values is indicating a risk of inaccuracy for data values in a data set.	$X = A/B$ A =number of data values that are outliers. B = number of data values to be considered in a data set.	Inherent
Note 1: Outlier: a value that is numerically distant from the rest of values. An outlier is an exception. It can be calculated with different methods. Note 2: For X, lower is better. Note 3: To reduce the risk of inaccuracy, outliers should be validated by human or instrument.				
Acc-I-5	Data model accuracy	Data model describes the system with the required accuracy.	$X = A/B$ A = number of elements of the data model that accurately describe the system. B = number of elements of the data model that describe the required accuracy within the requirement specification of the system.	Inherent
Note 1: This QM derives from an assessment of the data models. Note 2: This QM is based on the subjective opinion of intended users of data models, which reflects the quality (in terms of accuracy) of these artifacts for the user's needs and goals. Note 3: This QM is related to the data models that describe the system at the same level of abstraction, related to the same context, developed with the same techniques of representation. Note 4: The accuracy of a data model A, compared for example with a previous data model B, depends on the appropriate and detailed graphical representation against the requirements within the specification. Note 5: Generally, X>0 (X=1 is better).				

Acc-I-6	Metadata accuracy	Does metadata describe data with the required accuracy?	$X = A/B$ A = number of metadata that provides the appropriate required information. B = number of metadata defined within the requirement specification of data.	Inherent
Note 1: Accuracy of metadata is relevant for critical data such as data used in GIS data models or e-health. As data moves along DLC, it is used by a number of actors who need interpretable and useful data Note 2: It can be verified by requirement design specification of data. Note 3: The criteria of accuracy of metadata is that the degree to which metadata provide a requested information based on the requirement specification of data from the stakeholders and include documentation to interpret the meaning and properties of data correctly.				
Acc-I-7	Data range accuracy	Are data values included in the required interval?	$X = A/B$ A = number of data items having a value included in a specified interval (i.e. range from maximum to minimum). B = number of data items for which can be defined a required interval of values.	Inherent
Note: Required values can be defined in the requirement specification of the system. The acceptable value of interval is decided by management decision or by statistical analysis from information of QMs or observation of natural phenomena.				

3.2.2 QMs for completeness

3.2.2.1 Completeness measures provide the degree to which data associated with a target entity has expected values for all related properties of target entity in a specific context of use.

Table 3.2.2.1 Completeness Measures

ID	Name	Description	Measurement Function	Point of View
Com-I-1	Record completeness	Completeness of data items of a record within a data file.	$X = A/B$ A = number of data items with associated values not null in a record. B = number of data items of the record for which completeness can be measured.	Inherent
Note: This QM can be used to calculate average of completeness for different set of records.				
Com-I-2	Attribute completeness	Completeness of data items within a data file.	$X = A/B$ A = number of records with associated values not null for a specific data item. B = number of records counted.	Inherent
Com-I-3	Data file completeness	Completeness of records expected within a data file.	$X = A/B$ A = number of records contained in a data file. B = number of records expected.	Inherent

Com-I-4	Data values completeness	Completeness of values of a data item in a data file.	$X = {}^{A}/_{B}$ A =number of data values for a data item in a data file connected to expected values. B = number of data values expected for a data item in a data file.	Inherent
Com-I-5	Empty records in a data file	False completeness of records within a data file.	$X = {}^{A}/_{B}$ A =number of records where all data items are empty. B = number of records in a data file.	Inherent
Note: Records exist but are empty.				
Com-I-6	Conceptual data model completeness	Completeness of entities described in conceptual data mode vs. contextual schema.	$X = {}^{A}/_{B}$ A =number of entities of the conceptual data model. B =number of entities of the conceptual data model that describe the contextual schema completely.	Inherent
Note 1: This QM is based on the subjective point of view of intended users of a conceptual data model, which reflects the quality (in terms of completeness) of this artifact for the user's needs and present or future goals. Note 2: The completeness of the contextual schema characterizes the extend to which the schema represent the corresponding real world or future evolution. The context is determined by circumstances of all influences upon a system that includes developmental, technological, business, operational, organizational, political, economic, legal, regulatory, ecological and social influences. The completeness of conceptual model characterizes the presence/absence of entities included in the model. Note 3: This QM dose not take into account attributes. Note 4: Generally, X >0(X =1 is better); a value between 0, 1 can be indicate a lack of number of entities of the conceptual data models (A) or partial consideration of complete contextual schema, on the other hand, a value ＞1 can indicate an overdescription of conceptual data mode (A) in comparison with the conceptual data model (B) that describe the contextual schema completely; this kind of situation may be an excess of specification of the conceptual data model (A).				
Com-I-7	Conceptual data model attributes completeness	Completeness of attributes defined for a conceptual data model.	$X = {}^{A}/_{B}$ A = number of attributes defined in the conceptual data model. B = number of attributes defined in the conceptual data model that describe the contextual schema completely.	Inherent
Note 1: This QM is based on the subjective opinion of intended users of a conceptual data model, which reflects the quality (in terms of completeness) of this artifact for the user's needs and present goals. Note 2: For example, "attribute" in the conceptual data model concerns the address in terms of "state, city, street and the street number", but in the context of use, it is necessary to define the geo-localization parameter. Note 3: Generally, X >0(X =1 is better); a value between 0, 1 can indicate a lack of number of attributes of the conceptual data models (A); on the other hand, a value＞1 can indicate an overdescription of attributes of				

conceptual data model (A) in comparison with the conceptual data model (B) that describe the contextual schema completely; this kind of situation may be an excess of specification of attributes of conceptual data model (A).				
Com-I-8	Metadata completeness	Completeness of attributes for metadata.	$X = {}^{A}/_{B}$ A = number of attributes with complete metadata within the data dictionary. B =number of attributes for which metadata are expected within the data dictionary.	Inherent

3.2.3 QMs for consistency

3.2.3.1 Consistency measures provide the degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use. They can be either or both among data regarding one target entity and across similar data for comparable target entities.

Table 3.2.3.1 Consistency Measures

ID	Name	Description	Measurement Function	Point of View
Con-I-1	Referential integrity	For each value of one attribute of a table exists the same value of the same attribute in a different table; i.e., there is link between the same attribute represented in different tables and they contain the same values.	$X = 1 - {}^{A}/_{B}$ A = number of data items not consistent by value. B =number of data items for which referential integrity must be defined.	Inherent
Note: Data item is in the same data file or when the same data item is in different tables/data files but related to record with the same primary key values.				
Con-I-2	Data format consistency	Consistency of data format of the same data item.	$X = {}^{A}/_{B}$ A =number of data items where the format of all properties is consistent in different data files. B =number of data items for which format consistency can be defined.	Inherent
Note: It can be applicable to the same data item in the same system or different systems.				
Con-I-3	Risk of data inconsistency	Risk of having inconsistency due to duplication of data value.	$X = {}^{A}/_{B}$ A =number of data items where exist duplication in value. B = number of data items considered.	Inherent
Note 1: Risk is considered proportional to the number of duplicates. Note 2: For X, lower is better.				
Con-I-4	Architecture consistency	Degree to which the elements of the architecture have a	$X = {}^{A}/_{B}$	Inherent

		correspondence in referenced architecture elements.	A = number of elements of an architecture that have corresponding referenced elements in the installed architecture. B = number of elements of the referenced architecture.	
Note: For example, the elements of the data dictionary do not match with the elements in the conceptual data model.				
Con-I-5	Data values consistency coverage	Coverage of consistency measurement of data values.	$X = A/B$ A = number of data items considered in consistency measurement of data values. B =number of date items for which consistency is measured.	Inherent
Note: This QM is not measuring the quality of the data, but it measures the thoroughness and application of the consistency measures.				
Con-I-6	Semantic consistency	Degree to which semantic rules are respected.	$X = A/B$ A =number of data items where values are semantically correct in the data file. B =number of data items for which semantic rules are defined.	Inherent
Note: For example, an employee's birth date cannot be later than his "recruitment date".				

3.2.4 QMs for credibility

3.2.4.1 Credibility measures provide the degree to which data has attributes that are regarded as true and believable by users in a specific context of use.

Table 3.2.4.1 Credibility Measures

ID	Name	Description	Measurement Function	Point of View
Cre-I-1	Values credibility	Degree to which information items are regarded as true, real and credible.	$X = A/B$ A = number of information items where values are validated/certified by a specific process. B = number of information items to be validated/certified.	Inherent
Note 1: QM can be applicable to data items (permissible values) and could be extended to evaluate credibility of records, data files, etc. Note 2: QME A includes the verification of outliers (see Acc-I-4). Note 3: A specific process refers to QMs from inherent point of view and can include evaluation of different sources.				
Cre-I-2	Source	Degree to which	$X = A/B$	Inherent

	credibility	values are provided by a qualified organization.	A = number of data values provided or validated/certified by a qualified organization. B =number of data values for which source credibility can be defined.	
Note: Data values provided or validated by rightful data owner are considered certified data values.				
Cre-I-3	Data dictionary credibility	Degree to which data credibility provides credible information.	$X = A/B$ A =number of information items in the data dictionary for which values are validated/certified by a specific process. B =number of information items in the data dictionary.	Inherent
Cre-I-4	Data model credibility	Degree to which data model provides credible information.	$X = A/B$ A =number of elements of a data model with appropriate definition validated/certified by a specific process. B =number of elements of a data model.	Inherent

3.2.5 QMs for currentness

3.2.5.1 Currentness measures provide the degree to which data has attributes that are of the right age in a specific context of use.

Table 3.2.5.1 Currentness Measures

ID	Name	Description	Measurement Function	Point of View
Cur-I-1	Update frequency	Degree to which data items are updated with the frequency required.	$X = A/B$ A = number of data items updated with the required frequency. B =number of data items having an update frequency requirement.	Inherent
Note 1: For example, the frequency of update can be “daily”. Note 2: “A” is valid also for higher frequency.				
Cur-I-2	Timeliness of update	Degree to which data items are timely updated.	$X = A/B$ A =number of data items timely updated. B = number of data items needing updating.	Inherent
Note: The timeliness of update shall refer to requirements.				
Cur-I-3	Update item requisition	Degree to which frequency requisition of explicit update data items exists.	$X = A/B$ A =number of information items with an explicit update	Inherent

			requisition. B =number of information items for which an update requisition is necessary.	

3.2.6 QMs for accessibility

3.2.6.1 Accessibility measures provide the degree to which data can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability.

Table 3.2.6.1 Accessibility Measures

ID	Name	Description	Measurement Function	Point of View
Acs-I-1	User accessibility	Degree to which data values are considered accessible by intended users.	$X = A/B$ A = number of data items relevant to the user's task within a specific context of use having values accessible by intended users. B =number of data items that are relevant to the user's task within the context of use having values that are required to be accessible in conformance to specification.	Inherent
Note 1: Particular case of data item accessibility can refer to paper form or screen; the technology can assist users through specific software such as screen reader for the screen (assistive technology) Note 2: Example: data (or text) managed by a screen reader cannot be stored as an image.				
Acs-D-1	Device accessibility	Degree to which accessibility is allowed by a specific device (e.g., voice or sound with textual representation).	$X = A/B$ A =number of data items with values accessible through a specific device by intended users. B = number of data items for which device accessibility can be defined.	System dependent
Note: It can be verified by interviews or questionnaire.				
Acs-D-2	Data format accessibility	Degree to which data or information are not accessible by the intended users due to a specific format.	$X = A/B$ A = number of data items not accessible due to its format. B = number of data items for which format accessibility can be defined.	System dependent
Note 1: For example, number of documents in PDF not accessible by the screen reader, because of its format not managed by the device.				

Note 2: It can be verified by interviews or questionnaire.

3.2.7 QMs for compliance

3.2.7.1 Compliance measures provide the degree to which data has attributes that adhere to standards, conventions or regulations in force and similar rules relating to data quality in a specific context of use.

Table 3.2.7.1 Compliance Measures

ID	Name	Description	Measurement Function	Point of View
Cmp-I-1	Regulatory compliance of value and/or format	Degree to which data values and/or format comply with specific standards, conventions or regulations.	$X = A/B$ A = number of data items that have values and/or format that conform to standards, conventions or regulations. B = number of data items that shall conform to standards, conventions or regulations due to their value.	Inherent
Cmp-D-1	Regulatory compliance due to technology	Degree to which data items comply with specific standards, conventions or regulations.	$X = A/B$ A =number of data items that conforms to standards, conventions or regulations due to technology. B = number of data items that shall conform to standards, conventions, or regulations due to technology.	System dependent

3.2.8 QMs for confidentiality

3.2.8.1 Confidentiality measures provide the degree to which data has attributes that ensure that it is only accessible and interpretable by authorized users in a specific context of use.

Table 3.2.8.1 Confidentiality Measures

ID	Name	Description	Measurement Function	Point of View
Cnf-I-1	Encryption usage	Degree to which data values are fulfilling the requirement of encryption.	$X = A/B$ A =number of data values correctly and successfully encrypted and decrypted. B = number of data values with encryption and decryption requirement.	Inherent
Cnf-D-1	Non vulnerability	Degree to which data item defined as confidential can be accessed by authorized users only.	$X = 1 - A/B$ A = number of accesses successfully performed during formal penetration attempts by unauthorized users to reach target data item in a specific period of time. B = number of accesses attempted by unauthorized users to target data item in a	System dependent

			specific period of time.	

3.2.9 QMs for efficiency

3.2.9.1 Efficiency measures provide the degree to which data has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use.

Table 3.2.9.1 Efficiency Measures

ID	Name	Description	Measurement Function	Point of View
Eff-I-1	Efficient data item format	Ratio of using data formats that allows users to perform their operations efficiently.	$X = A/B$ A =number of data items that are stored in a format that are qualified as efficient. B = number of data items for which format is tested for efficient operation.	Inherent
Note: The qualification of format as efficient can derive from international or national regulators, results of tests, etc.				
Eff-I-2	Use efficiency	Ratio of data values that allow intended users to easily use them.	$X = A/B$ A =number of data values that intended users evaluate as "easily used". B =number of data values evaluated by users.	Inherent
Note 1: "Easily used" can be realized by interviews or questionnaires. Note 2: "Easily used" can be evaluated by comparing the time spent to use the data. Note 3: For ship navigation, nautical miles are more usable than kilometers.				
Eff-D-1	Data format efficiency	Unnecessary space occupied rate due to data format definition.	$X = 1 - A/B$ A =size in bytes of record in a data file unnecessarily occupied due to data format definition. B =size in bytes of record in a data file due to data format definition.	System dependent
Note: Unnecessary space derives from the verification that technological alternatives are possible, producing the same or better results under the same specified conditions.				
Eff-D-2	Data processing efficiency	Working time lost due to data item representation (data formatting).	$X = 1 - A/B$ A =time lost due to data item representation (data formatting) during a work. B =(total) time of processing.	System dependent
Note 1: The range of QM values varies depending on the specific context of use. Note 2: Processing can be automatic or human. For example, there will be time lost when a blind person tries to read an image of pdf format with a screen reader.				
Eff-D-3	Risk of	Wasted space in	$X = SUM(B) - A$	System

	wasted space	comparison with benchmarked average space.	A =size in bytes assumed as target for efficient data storage of the database. B =size in bytes used for data in any physical data files of the database.	dependent
Note: For X, lower is better.				
Eff-D-4	Space occupied by records duplication	Records duplication space is bigger than the space occupied by deduplicated records.	$X = A/B$ A =size in bytes of space occupied due to records duplication in a data file. B =size in bytes of space occupied due to record without duplication in the same data file.	System dependent
Note 1: The duplication problem is more relevant for data files that do not allow the definition of key constraints. Note 2: Result value can vary from 1 to infinite. Lower is better.				
Eff-D-5	Time delay of data update	Delay between the time at which values of data items change in system A and the time at which values of the same data items change in system B (when system A has to provide values to system B).	$X = t_2 - t_1$ t_1 =time when data item is updated in system A. t_2 =time when data item is updated in system B.	System dependent
Note 1: X can vary from zero to infinite. Lower is better. Note 2: The set of measured value of this QM can be applied to introduce the average update time of data item.				

3.2.10 QMs for precision

3.2.10.1 Precision measures provide the degree to which data has attributes that are exact or that provide discrimination in a specific context of use.

Table 3.2.10.1 Precision Measures

ID	Name	Description	Measurement Function	Point of View
Pre-I-1	Precision of data values	Degree of data values precision according to the specification.	$X = A/B$ A = number of data values with the requested precision. B = number of data values with the precision requirement defined.	Inherent
Note: For example, B is concerning a data file where all data values are supposed to have three decimals. For A is verified that only some data values have the requested precision. X will be less than 1.				
Pre-D-1	Precision	Degree to which data	$X = A/B$	System

	of data format	format keep precision according to the specification.	A =number of data items defined and perceived with the requested precision/data format . B = number of data items for which precision of format is required.	dependent
Note: Data format precision can be verified based on the data format specification.				

3.2.11 QMs for traceability

3.2.11.1 Traceability measures provide the degree to which data has attributes that provide an audit trail of access to the data and of any changes made to the data in a specific context of use.

Table 3.2.11.1 Traceability Measures

ID	Name	Description	Measurement Function	Point of View
Tra-I-1	Traceability of data values	Degree to which the information of user access to the data value was traced.	$X = A/B$ A =number of data values for which required access traceability of values exist. B =number of data values for which access traceability is expected.	Inherent
Note: This QM is not measuring the traceability process; traceability exists and it conforms to the requirement. The result is "yes/no".				
Tra-D-1	Users access traceability	The possibility to keep information about users access to data using system capabilities, for investigating who read/wrote data.	$X = A/B$ A =number of data items for which user access traceability is expected and realized. B =number of data items for which user access traceability is expected.	System dependent
Tra-D-2	Data items traceability	The possibility to trace the history of a data item value using system capabilities.	$X = A/B$ A =number of data items for which values are traceable using system capabilities. B =number of data items for which values are expected to be traceable using system capabilities.	System dependent

3.2.12 QMs for understandability

3.2.12.1 Understandability measures provide the degree to which data has attributes that enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use.

Table 3.2.12.1 Understandability Measures

ID	Name	Description	Measurement Function	Point of View

Und-I-1	Symbols understandability	Degree to which comprehensible symbols are used.	$X = A/B$ A = number of data values represented by known symbols. B =number of data values for which symbols understandability is requested.	Inherent
Und-I-2	Semantic understandability	Ratio of the common recognized vocabulary which is used in terms of definitions given in the data dictionary.	$X = A/B$ A =number of data values defined in the data dictionary by using a common vocabulary. B =number of data values defined in the data dictionary.	Inherent
Note 1: The "common" vocabulary needs validation by human or instruments. The availability of synonyms can be useful to better understand the meaning. Note 2: For example, distance in "nautical miles" or "kilometers" is always necessary to be specified.				
Und-I-3	Master data understandability	Understandability of Master Data is due to metadata definition.	$X = A/B$ A =number of data items of master data files with existing metadata. B =number of data items of master data files.	Inherent
Note: Metadata are essential to understand the meaning of data items.				
Und-I-4	Data values understandability	Data values are understandable by intended users in the specific context of use.	$X = A/B$ A =number of data values easily understandable by intended users. B = number of data values that users attempt to understand during an observation period.	Inherent
Note: For example, the counting number of users' complaints during an observation period can be verified.				
Und-D-1	Data model understandability	Degree to which data models provide understandable information.	$X = A/B$ A =number of elements considered understandable in a data model. B =number of elements provided by a data model.	System dependent
Note 1: It can be verified by interviews or questionnaire. Note 2: For example, this QM can be used when a developer has to read a data model provided by an analyst to realize a software application.				
Und-D-2	Data representation understandability	Degree to which data is represented in a comprehensible way	$X = A/B$ A = number of data items considered understandable by	System dependent

		to users by system and software.	intended users. B =number of data items presented in a specific device.	
Und-D-3	Linked master data understandability	Understandability of master data is due to linked understandable metadata.	$X = A/B$ A =number of data items of master data files with automatic linked understandable metadata. B =number of data items of master data files.	System dependent
Note 1: This QM reflects the maturity of master data management that allows integration of the system and efficiency of interrelated datasets. Note 2: Metadata are linked by the system by using specified software, data formats, data dictionary.				

3.2.13 QMs for availability

3.2.13.1 Availability measures provide the degree to which data has attributes that enable it to be retrieved by authorized users and/or applications in a specific context of use.

Table 3.2.13.1 Availability Measures

ID	Name	Description	Measurement Function	Point of View
Ava-D-1	Data availability ratio	Ratio of data items available when required (e.g. during backup/restore procedures).	$X = A/B$ A =number of data items available in a specific period of time. B =number of data items requested in the same period of time.	System dependent
Note: Required specific period or duration can possibly be included during backup or restore procedures, as well as during normal processing.				
Ava-D-2	Probability of data available	The probability of successful requests trying to use data items during requested duration.	$X = A/B$ A =number of times that data items are available for the requested duration. B =number of times that data items are requested for the requested duration.	System dependent
Ava-D-3	Architecture elements availability	Degree to which architecture elements are available.	$X = A/B$ A = number of elements of the architecture available for the intended users. B = number of elements of the architecture.	System dependent
Note: The availability of architecture elements can imply the availability of related target entity.				

3.2.14 QMs for portability

3.2.14.1 Portability measures provide the degree to which data has attributes that enable it to be installed, replaced, or moved from one system to another preserving the existing quality in a specific context of use.

Table 3.2.14.1 Portability Measures

ID	Name	Description	Measurement Function	Point of View
Por-D-1	Data portability ratio	Data quality does not decrease after porting (or migration).	$X = A/B$ A =number of data items that preserve existing quality after porting. B =number of data items ported.	System dependent
Note: In this QM, the portability refers to the result of the porting.				
Por-D-2	Prospective data portability	Degree to which portability of data item conforms to requirements.	$X = A/B$ A =number of data items that can be moved to a target system. B =number of data items for which portability is expected.	System dependent
Note: This QM refers to the possibility to implement a porting activity.				
Pro-D-3	Architecture elements portability	Degree to which architecture elements are portable.	$X = A/B$ A = number of elements of the architecture that are portable for a specific user. B = number of elements of the architecture required to be ported.	System dependent
Note: For example, data models and data dictionary are portable.				

3.2.15 QMs for recoverability

3.2.15.1 Recoverability measures provide the degree to which data has attributes that enable it to maintain and preserve a specific level of operations and quality, even in the event of failure, in a specific context of use.

Table 3.2.15.1 Recoverability Measures

ID	Name	Description	Measurement Function	Point of View
Rec-D-1	Data recoverability ratio	Degree to which data stored in a device are successfully and correctly recovered.	$X = A/B$ A =number of data items successfully and correctly recovered by the system. B =number of data items that are required to be recovered.	System dependent
Rec-D-2	Periodical backup	Data is backed up periodically as stated in requirements.	$X = A/B$ A =number of data items (or data file) successfully backed up periodically. B =number of data items (or data file)	System dependent

			to be backed up periodically.	
Rec-D-3	Architecture recoverability	Degree to which architecture elements are recoverable.	$X = {}^{A}/_{B}$ A = number of elements of the architecture successfully backed up/recovered. B = number of elements of the architecture that shall be managed by back-up/restore procedures.	System dependent

Section 3 Ship Data Quality Assessment Process

3.3.1 Basic process of ship data quality assessment

3.3.1.1 The data quality assessment process is shown in Figure 3.3.1.1.

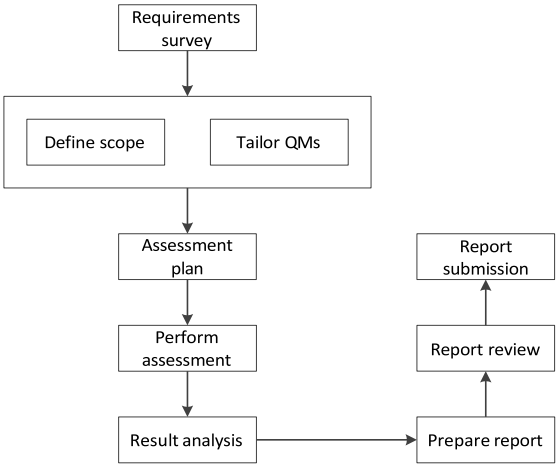

Figure 3.3.1.1 Data Quality Assessment Flowchart

3.3.2 Scope of data quality assessment

3.3.2.1 The data scope of data quality assessment can be determined according to the functional goals of the system. Please refer to Section 3.3.3-3.3.4 for the methods.

3.3.3 Demand analysis of data quality assessment

3.3.3.1 The analysis of data quality requirements generally includes identification of data sets, survey of data users and subject matter experts, initial identification of data quality issues, definition of appropriate data quality characteristics, measures and indicators, as well as setting quality goals, etc. The main steps are shown as follows:

(1) Selection/identification of data sets

①The data sets (such as the data sets to be assessed or small or comprehensive data sets similar to historical data sets generated by data applications) for initial assessment are selected/identified by using classic statistical analysis methods (such as data analysis or similar technology). The process of organizing and analyzing data sets helps to identify the data characteristics (including data type, format, precision, range, etc.), identify potential data quality problems and determine the data quality measurement rules.

(2) Preliminary analysis and assessment

①The needs of users of data are collected, and potential data quality problems/abnormalities are

identified through data analysis of the determined data sets, as well as potential problems in data collection, transmission and storage caused by hardware and software with the help of data experts;

②The technical and business opinions of data-related parties are collected, and the potential impact of identified data quality issues on potential data usage is assessed (such as the impact on data accuracy and reliability and final impact on the overall function of data applications).

(3) Definition of quality demand

①The data quality requirements are defined based on the identified data quality issues and their potential impact, as shown in Table 3.3.3.1. The requirements are defined including the determination of data quality attributes, quality measures, and measurable indicators;

② The threshold for each data quality measurement result is determined, and the acceptance criteria for data quality assessment results are determined.

Table 3.3.3.1 Data quality requirements

Characteristic	Measure	Data Quality Rules			Remarks
		Set threshold	Set weight	Applicability Description	
Accuracy	Acc-I-1				
	Acc-I-2				
	Acc-I-3				
	Acc-I-4				
	Acc-I-5				
	Acc-I-6				
	Acc-I-7				
Completeness	Com-I-1				
	Com-I-2				
	Com-I-3				
	Com-I-4				
	Com-I-5				
	Com-I-6				
	Com-I-7				
	Com-I-8				
Consistency	Con-I-1				
	Con-I-2				
	Con-I-3				
	Con-I-4				
	Con-I-5				
	Con-I-6				
Credibility	Cre-I-1				
	Cre-I-2				
	Cre-I-3				
	Cre-I-4				
Currentness	Cur-I-1				
	Cur-I-2				
	Cur-I-3				
Accessibility	Acs-I-1				
	Acs-D-1				
	Acs-D-2				
Compliance	Cmp-I-1				

	Cmp-D-1				
Confidentiality	Cnf-I-1				
	Cnf-D-1				
Efficiency	Eff-I-1				
	Eff-I-2				
	Eff-D-1				
	Eff-D-2				
	Eff-D-3		/		
	Eff-D-4				
	Eff-D-5		/		
Precision	Pre-I-1				
	Pre-D-1				
Traceability	Tra-I-1				
	Tra-D-1				
	Tra-D-2				
Understandability	Und-I-1				
	Und-I-2				
	Und-I-3				
	Und-I-4				
	Und-D-1				
	Und-D-2				
	Und-D-3				
Availability	Ava-D-1				
	Ava-D-2				
	Ava-D-3				
Portability	Pro-D-1				
	Pro-D-2				
	Pro-D-3				
Recoverability	Rec-D-1				
	Rec-D-2				
	Rec-D-3				

Note 1: The applicability of data quality characteristics is described according to the requirements specification and business scenario requirements.

Note 2: Acceptable thresholds for quality measures are set according to the requirement specification and business scenario requirements.

Note 3: The weight of each quality measure in the total data quality measures (expressed as a percentage) is set according to the requirements specification and business scenario requirements, please refer to the requirements in Section 3.4.2 of the Guidelines.

3.3.4 Rules and criteria for data quality assessment

3.3.4.1 The data quality assessment rules include the rules for validation and management of data quality:

(1) Validation rules

①The defined data quality measures should be tested and validated by means of applying the defined data quality measures to a test data set instead of a complete data set used for preliminary assessment;

②The defined data quality measures should be reviewed with the user to ensure that these measures are understood by the user.

(2) Rules and criteria for managing data quality

①Data quality rules should be recorded in a consistent format, and described in clear natural language;

②The data quality rules should be defined based on measurable data quality measures and thresholds;

③The use of data is to be considered when creating rules. The defined data quality rules should be tested based on the actual data sets. It is recommended to continuously improve the quality rules throughout the data quality improvement life cycle;

④Data users and professional experts should get involved in the formulation of data quality rules. The defined rules should be confirmed by data users (for example, the person who oversees data usage) and professional experts.

3.3.5 Data quality assessment

3.3.5.1 Data quality assessment includes:

(1) Assessing data quality

①Data quality should be assessed according to the data quality rules;

②The non-conformances and the observed data quality issues should be recorded.

(2) Identifying potential improvements and their priorities

①The importance of known data quality issues should be prioritized based on the impact on data usage, and the improvement plan should be evaluated to solve data quality issues;

②Priority should be given to repair and improvement based on comprehensive data analysis and input from related parties;

③The data quality issues that require in-depth analysis of the root cause and possible improvement plans should be identified.

(3) Developing the handling process of data issues

①The identified data quality issues are checked, and the potential root cause of the problem is identified with the assistance of data users and subject matter experts;

②Non-technical root causes are addressed: providing appropriate training for data processing personnel; improving the data processing procedures; strengthening leadership support; establishing clear accountability and ownership;

③Addressing technical root causes:

1) Defective data should be directly corrected;

2) The performance of data acquisition, transmission and storage should be improved;

3) The system and technical process should be modified to prevent the problems from occurring again;

4) Continuous monitoring is carried out without taking any immediate actions after balancing the data quality issues and the cost of correction/improvement.

(4) Addressing data quality issues

①Cost-benefit analysis should be carried out to determine feasible data problem handling procedures;

②Suggestions from data users and various professional experts are sought to choose the best options to solve the problems, including simple corrective measures and root cause corrections;

③Simple corrective measures refer to repairing and correcting data directly in the record (for example, data cleaning/analysis/formatting);

④Root cause correction refers to making a long-term improvement plan for strategic changes (such as system modifications) . It focuses on modifying the system to solve the root cause, and establishing a mechanism first to prevent problems. Prevention is usually more cost-effective than correction;

⑤A correction plan is developed and implemented, which aims to reassess the quality level of the revised data sets, and ensure that the applied changes will not introduce other errors and will be executed as expected.

(5) Monitoring the continuous compliance of data and quality rules

①The continuous compliance of data and quality rules should be monitored, and all data quality

assessment levels should be reported;

②The indicators, measures, and data quality scores related to the formulation of data quality rules should be recorded;

③The data quality over time should be recorded;

④The handling of data quality issues should be tracked based on the correction plan;

⑤The time interval should be set to periodically evaluate the quality according to the defined rules.

⑥ Threshold or acceptance criteria is set for each measurement. The data quality results usually reflect the percentage of correct data (passing the measurement rules) or the percentage of abnormal data (failing to pass the validation rules), which depends on the formula used. If the data meets the defined data quality rules, please confirm that the data is suitable for its intended application (for example, data analysis). If the data fails to meet the defined data quality rules, notification is sent in a timely manner to warn of the data quality issues, and possible actions are proposed based on the designated remedial plan.

3.3.6 Data quality assessment report

3.3.6.1 The data quality assessment report includes but is not limited to:

(1) Assessing data set;

(2) Assessment rules and criteria;

(3) Data quality assessment results;

(4) Data quality visualization;

(5) Non-conformances and data quality issues;

(6) Data quality issue analysis and handling;

(7) Settlement of data quality issues.

3.3.7 Continuous improvement of data quality

3.3.7.1 Continuous improvement of data quality can be realized by starting a new cycle. A new data quality improvement cycle may be restarted according to the following conditions:

(1) The existing data quality results are lower than the acceptance criteria;

(2) There are new data sets to be analyzed;

(3) Adoption of new data quality requirements for the existing data sets;

(4) Equipment/system changes.

Section 4 Analysis of Ship Data Quality Assessment Results

3.4.1 Data quality measurement

3.4.1.1 The calculated value of each data quality measure should be measured based on the data quality requirements determined according to Table 3.3.3, as well as attribute definitions and measurement functions in Section 3.2 of the Guidelines.

3.4.1.2 The measured value and threshold of each data quality measure should be compared according to the set threshold of the data quality measures in Table 3.3.3.

3.4.2 Data quality measure weight

3.4.2.1 Weight w_i is developed for each quality measure in the data quality requirements table for each type of data according to the requirements specification and business scenario requirements.

$$\sum_{i=1}^{n} w_i = 1$$

Where, w_i represents the weight of quality measure i.

n represents the total number of data quality measures used by a certain type of data.

3.4.2.2 For example, the weight is shown in Table 3.4.2.2 by taking the quality measures of the ship heading data retained by the system to be tested as an example. For simplicity's sake, only the two attributes —accuracy and currentness are considered in the table.

Table 3.4.2.2 Example of set weights

Characteristic	Measure	Set Threshold (%)	Set Weight (%)	Measurement Result (%)
Accuracy	Acc-I-1	90	50	100
	Acc-I-2	90	15	90
	Acc-I-3	90	0	0
	Acc-I-4	90	5	95
	Acc-I-5	90	0	0
	Acc-I-6	90	5	0
	Acc-I-7	90	15	100
Accuracy measurement		90	100	83.25
Currentness	Cur-I-1	90	80	90
	Cur-I-2	90	20	100
	Cur-I-3	90	0	0
Currentness measurement		90	100	92

The weight of data quality measures depends on the data usage requirements, and it is recommended that joint confirmation should be made by the data users and assessment experts.

3.4.3 Data quality assessment results

3.4.3.1 Overall threshold for data quality assessment should be set for data quality assessment. This indicator is used to evaluate the quality or acceptability of the data quality of the system/product. Based on this quality index, a change trend of data quality can be determined, and corresponding measures may be taken accordingly to continuously improve the data quality.

3.4.3.2 The score of data quality assessment threshold is divided into pass or fail. Based on the business requirements of the system/product being evaluated, the threshold that must be reached and the threshold that must be reached by the system/product as a whole will be provided for the data quality characteristics. The evaluation results of the system/product are obtained after weighted calculation according to 3.4.2, when the evaluation results reach the overall acceptable threshold, it should be considered passing the evaluation system. For the above thresholds, initial thresholds may be set and be gradually improved with iterative evaluation of the system/product.

3.4.3.3 The data quality assessment index allows the setting of acceptable thresholds for each data quality measure, as well as the setting of an acceptable threshold for the overall measurement result of a certain data.

Chapter 4 Data Management Maturity Assessment

Section 1 General Provisions

4.1.1. General requirements

4.1.1.1 Data management capability refers to the capability of an organization or institution to manage and apply data.

4.1.1.2 Data management capability maturity assessment model refers to the model used to assess the data management capability maturity of an organization.

4.1.1.3 The Guidelines specify the models and indicators used by ISC to implement data management maturity assessment, which divides data management maturity into 5 levels and 8 data management capability areas.

4.1.2 Data management maturity architecture

4.1.2.1 The data management maturity architecture is shown in Figure 4.1.2.1.

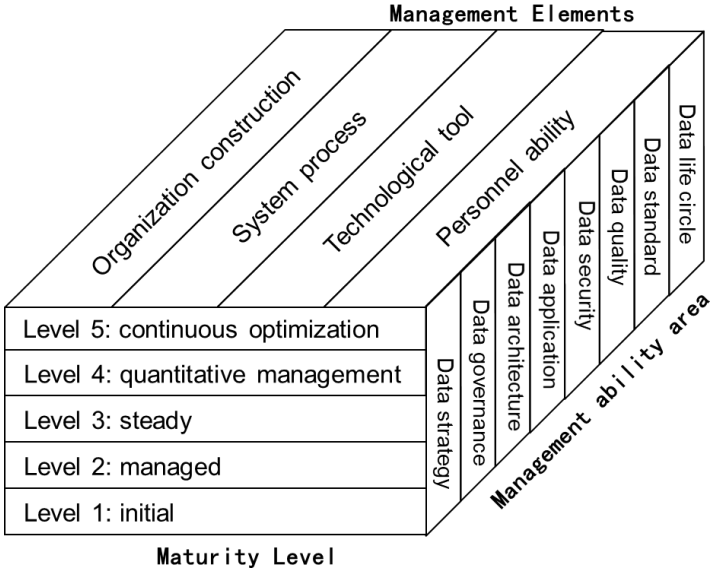

Figure 4.1.2.1 Data Management Maturity Architecture

4.1.3 Data management maturity level characteristics

4.1.3.1 An organization's data management maturity is divided into 5 levels, as shown in Table 4.1.3.1.

Table 4.1.3.1 Data Management Maturity Level

Level	Name	Description	Main Characteristics
1	Initial	The management of data requirements is mainly reflected at the project level, and it is mainly passive management without unified management process.	(1) The organization fails to receive sufficient data support when making strategic decisions; (2) There are no formal data planning, data architecture design, data management organization and process, etc.; (3) Each business system manages its own data, the data between all business systems are not consistent, and the organization is not aware of the-importance of data management or data quality;

			(4) Data management is only carried out according to the cycle of project implementation, and the cost of data maintenance and management cannot be calculated.
2	Managed	The organization has realized that data are assets, has formulated a management process according to the requirements of the management strategy, and designated relevant personnel for preliminary management.	(1) The organization is aware of the importance of data, and has formulated some data management specifications and set up relevant positions; (2) The organization have realized that data quality and isolated data island are an important management issue, but there is no solution to the problem currently; (3) The organization has carried out preliminary data integration, tried to integrate data from all business systems, and designed relevant data models and management positions; (4) The organization begins to document some important data, and designs related management measures for safety, risk, etc. of important data.
3	Steady	Data have become important assets to achieve the organization's performance goals, and a series of standardized management procedures have been developed at the organizational level to promote the standardization of data management.	(1) The organization has realized the value of data, and established rules and regulations for data management within the organization; (2) Data management and application can be combined with the business strategy, operation & management requirements and external regulatory requirements of the organization; (3) Relevant data management organization and management process have been established to ensure that all departments in the organization can work according to the process; (4) The organization can obtain data support during daily decision-making and business development, significantly improving the work efficiency; (5) The organization has participated in the training related to industry data management, and has data management personnel.
4	Quantitative management	Data are considered important resources for gaining competitive advantage, and the efficiency of data management can be subject to quantitative analysis and monitoring.	(1) The organization has recognized that data are its strategic assets, understood the important role of data in process optimization, performance improvement, etc., and can obtain the support of relevant data when formulating its business strategy; (2) A quantifiable evaluation index system is

			established at the organization level to accurately measure the efficiency of the data management process and optimize in a timely manner; (3) Participation in the formulation of relevant national and industry standards; (4) Regular training on data management and application within the organization; (5) The organization has fully borrowed the best cases of the industry and external resources such as national and industry standards during data management and application to promote its own data management and application.
5	Continuous optimization	Data are considered to be the basis for the survival and development of an organization, relevant management processes can be optimized in real time, and the best practices can be shared in the industry.	(1) The organization has taken data as its core competitiveness, and made good use of data to create more value and improve its efficiency; (2) The organization can take the lead in formulating relevant national and industry standards; (3) The organization can promote its own experience in data management capacity building as the best case in the industry.

Section 2 Data Management Maturity Assessment

4.2.1 Capability area and capacity item

4.2.1.1 The data management capability maturity assessment model contains a total of 8 data management capability areas, and each capability area covers several capability items in the data management domain, as shown in Table 4.2.1.1.

Table 4.2.1.1 Capability Area and Capacity Item

Capability Area	**Capability Item**
Data strategy	Data strategy planning
	Data strategy implementation
	Data strategy assessment
Data governance	Data governance organization
	Data system construction
	Data governance communication
Data architecture	Data model
	Data distribution
	Data integration and sharing
	Metadata management
Data application	Data analysis
	Data opening and sharing

	Data service
Data security	Data security strategy
	Data security management
	Data security audit
Data quality	Data quality requirements
	Data quality check
	Data quality analysis
	Data quality improvement
Data standard	Business term
	Reference data and master data
	Data element
	Index data
Data life cycle	Data Requirement
	Data design and development
	Data operation and maintenance
	Data retirement

4.2.2 Data strategy

4.2.2.1 Data strategy planning

The capability item of data strategy planning is shown in Table 4.2.2.1.

Table 4.2.2.1 Data Strategy Planning

Capability Item ID	Str-1
Overview	Data strategy planning refers to the result of consensus among all stakeholders. It aims to determine the motivation for data management and application from the macro and micro levels, and comprehensively reflects the needs of the data provider and the consumer.
Goals	a) It is to establish and maintain the data management strategy; b) It is to maintain the data management strategy throughout the data governance process (goals, objectives, priorities and scope) for all business areas; c) It is to identify the stakeholders and analyze the priority of various data management tasks based on the business value of data and data management goals; d) It is to develop, monitor and evaluate follow-up plans to guide the implementation of data management planning.

4.2.2.2 Data strategy implementation

The capability item of data strategy implementation is shown in Table 4.2.2.2.

Table 4.2.2.2 Data Strategy Implementation

Capability Item ID	Str-2
Overview	It refers to the process of completing data strategy planning and gradually realizing data function framework by the organization. Assess the current status of the organization's data management and data application during implementation, and determine the gap between the vision and goals; develop phased data mission goals based on the data function framework, and determine the implementation steps.
Goals	a) It is to check the implementation of data strategy, and regularly evaluate the implementation; b) It is to compare the current situation with the development goals, analyze the gaps and clarify

	the development direction; c) It is to advance the implementation of the strategy, prioritize the data function tasks according to the existing gaps, combined with the common goals and the actual business value of the organization, and provide resources and financial guarantees to advance the implementation of the strategy.

4.2.2.3 Data strategy assessment

The capability item of data strategy assessment is shown in Table 4.2.2.3.

Table 4.2.2.3 Data Strategy Assessment

Capability Item ID	Str-3
Overview	Corresponding business case and investment model should be established in the process of data strategy assessment, and the progress should be tracked during the entire data strategy implementation process, and records should be kept for audit and evaluation at the same time.
Goals	a) It is to establish the business case for data function project, which should comply with the organization's goals and business-driven requirements, help the project obtain support at the implementation level, and provide reference for the investment model at the same time; b b) It is to establish one or a group of sustainable investment model(s) to meet the culture and business case requirements of the organization; c) It is to follow the investment model, and conduct reasonable cost-benefit analysis, meanwhile, the project function support should reflect the business goals and the organization's priorities; d) It is to record, track, audit and post-evaluate the business case, financial support methods and activities.

4.2.3 Data governance

4.2.3.1 Data governance organization

The capability item of data governance organization is shown in Table 4.2.3.1.

Table 4.2.3.1 Data Governance Organization

Capability Item ID	Gov-1
Overview	The data governance organization includes organizational structure, job setting, team building, data responsibility, etc., which is the basis for performing all data functions. It aims to conduct responsibility planning and control of data management and data application of the organization, and guide the implementation of all data functions to ensure effective implementation of the data strategy goals by the organization.
Goals	a) A complete organizational structure and corresponding work flow mechanism should be established; b) Data management should be clearly subject to centralized management, and enough full-time and part-time positions should be set up to continuously promote team building; c) A performance evaluation system that supports the strategy of data management and data application should be established.

4.2.3.2 Data system construction

The capability item of data system construction is shown in Table 4.2.3.2.

Table 4.2.3.2 Data System Construction

Capability Item ID	Gov-2
Overview	It aims to guarantee standardized operation of the functions of data management and data application, and establish corresponding institutional system. The data system is usually subject to hierarchical design, and it is required to follow strict release process and regularly check and update. Data system construction is the basis for orderly development of data management and data application, and is the basis for data governance communication and implementation.
Goals	a) It is to establish the data system, and publish it after opinions are solicited extensively within the organization; b) It is to establish the management process of the system, and carry out system inspection, update, release and promotion.

4.2.3.3 Data governance communication

The capability item of data governance communication is shown in Table 4.2.3.3.

Table 4.2.3.3 Data Governance Communication

Capability Item ID	Gov-3
Overview	Data governance communication aims to ensure that all stakeholders in the organization can know the latest situation of relevant policies, standards, procedures, roles, responsibilities, and plans in a timely manner, carry out related training on data management and application, and master the knowledge and skills related to data management. Data governance communication aims to establish and improve the cross-departmental and internal data management capabilities, raise awareness of data assets and build a data culture.
Goals	a) It is to communicate to ensure that the information on data management and data application can be known and understood by relevant personnel in a timely manner; b) It is to release regulatory compliance guidance documents that affect data management and data application in a timely manner; c) It is to establish a mechanism for stakeholders to participate in data governance communication; d) It is to strengthen the understanding of data-related systems, organizations, and standards by personnel of the organization.

4.2.4 Data architecture

4.2.4.1 Data model

The capability item of data model is shown in Table 4.2.4.1.

Table 4.2.4.1 Data Model

Capability Item ID	Arc-1
Overview	Data model aims to comprehensively analyze the collected data requirements used in business operation, management and decision-making of the organization by structured language, and reorganize the requirements according to model design specifications. From the application of model, the data model is divided into organization-level data model and system application-level data model. Organization-level data model includes subject domain model, conceptual model and logic model, while system application-level data model includes logical model and physical data model.

Goals	a) It is to establish and maintain the organization-level data model and system application-level data model; b) It is to establish a set of development specifications for the organization to follow the data model design; c) It is to use the organization-level data model to guide the construction of application systems.

4.2.4.2 Data distribution

The capability item of data distribution is shown in Table 4.2.4.2.

Table 4.2.4.2 Data Distribution

Capability Item ID	Arc-2
Overview	The data distribution capability item aims to clarify the distribution relationship of data in systems, organizations and processes, define data category, clarify authoritative data sources, and provide reference and specifications for data-related work based on the definition of data in the organization-level data model. It is also designed to define the priority of data-related work, designate the person responsible for data, and further optimize the data integration relationship by means of sorting out the data distribution relationship.
Goals	a) It is to establish a classification management mechanism for data assets of the organization, and determine the authoritative source of data; b) It is to sort out the relationship between data and business process, organization, and system; c) It is to standardize the construction of data-related work.

4.2.4.3 Data integration and sharing

The capability item of data integration and sharing is shown in Table 4.2.4.3.

Table 4.2.4.3 Data Integration and Sharing

Capability Item ID	Arc-3
Overview	Data integration and sharing capability item aims to establish an integrated sharing mechanism between all application systems and departments in the organization, and promote data connectivity within the organization through the management of related systems, standards, technology, etc. for data integration and sharing within the organization.
Goals	a) It is to establish efficient, flexible, and adaptable data exchange specifications and mechanisms between organization-level application systems; b) It is to establish a data integration and sharing environment to realize structured and unstructured data processing. It is provided with such functions as complex data processing, mining analysis and easy access.

4.2.4.4 Metadata management

The capability item of metadata management is shown in Table 4.2.4.4.

Table 4.2.4.4 Metadata Management

Capability Item ID	Arc-4
Overview	Metadata management is the set of a whole set of processes about metadata creation, storage, integration and control, etc.
Goals	a) It is required to classify the metadata, establish meta-model standards, guarantee the integration and interoperability of metadata from different sources, and realize standardized

	management by means of meta-model changes according to the business requirements, data management and application requirements; b) It is to realize effective integration of metadata from different sources, form a data panorama of the organization, manage and use data from different dimensions such as business, technology, operation, management, etc., and metadata should be changed according to relevant specifications; c) It is to build metadata application and metadata service, improve the understanding of data of the stakeholders, and assist in data management and data application.

4.2.5 Data application

4.2.5.1 Data analysis

The capability item of data analysis is shown in Table 4.2.5.1.

Table 4.2.5.1 Data analysis

Capability Item ID	App-1
Overview	Data analysis aims to carry out internal and external data analysis or mining modeling of the organization to provide decision-making support of data for operation and management of the organization, as well as delivery, operation, evaluation and promotion of corresponding results. Data analysis capability will affect the way that the organization makes decisions, creates value, and provides value to users.
Goals	a) Data analysis capability should meet the business operation requirements of the organization, and adapt to the development and changes in the business and technical fields; b) Data analysis should promote data-driven decision-making and the realization of business value, and data analysis should become the core competitiveness of the organization;

4.2.5.2 Data opening and sharing

The capability item of data opening and sharing is shown in Table 4.2.5.2.

Table 4.2.5.2 Data Opening and Sharing

Capability Item ID	App-2
Overview	Data opening and sharing refers to selective opening of data within the organization in accordance with a unified management strategy, and meanwhile, introduce external data in accordance with relevant management strategies for internal application by the organization. Data opening and sharing is an important prerequisite for realizing cross-organization and cross-industry data transfer, and it is also the basis for maximizing data value.
Goals	a) Data opening and sharing should meet the safety, regulatory and legal requirements; b) Data opening and sharing should promote interoperability of internal and external data, and promote the enhancement of data value.

4.2.5.3 Data service

The capability item of data service is shown in Table 4.2.5.3.

Table 4.2.5.3 Data Service

Capability Item ID	App-3
Overview	Data service aims to provide cross-domain and cross-industry data services in the form of data analysis results through unified processing and analysis of internal and external data of the organization, combined with the needs of the public, industry and the organization. Data service is the most direct way to realize the value of data assets, and is also one of the ways to measure

	the value of data assets. Excellent data service can help improve the efficiency of the organization to the internal extent, and better serve the public and the society to the external extent.
Goals	a) The data application mode of the organization to provide services or products to the outside can be explored to meet the needs of external users through data service; b) The value of data assets can be realized through data service.

4.2.6 Data security

4.2.6.1 Data security strategy

The capability item of data security strategy is shown in Table 4.2.6.1.

Table 4.2.6.1 Data Security Strategy

Capability Item ID	Sec-1
Overview	Data security strategy is the core content of data security, which should be formulated based on the organization's management requirements, regulatory requirements and related standards.
Goals	a) It is to establish a unified data security standard; b) It is to provide an applicable data security strategy.

4.2.6.2 Data security management

The capability item of data security management is shown in Table 4.2.6.2.

Table 4.2.6.2 Data Security Management

Capability Item ID	Sec-2
Overview	Data security management aims to manage data security through authorization of data access, classification and grading control, and access to monitoring data under the guidance of data security standard and strategy to meet the business needs and regulatory requirements of data security, and realize data security management of the data life cycle within the organization.
Goals	a) Hierarchical management should be performed to data within the organization by focusing on the requirements for data management; b) It is to monitor all links of data circulation within the organization to ensure data security; c) It is to analyze potential data security risks to prevent risks.

4.2.6.3 Data security audit

The capability item of data security audit is shown in Table 4.2.6.3.

Table 4.2.6.3 Data Security Audit

Capability Item ID	Sec-3
Overview	Data security audit is a control activity that is responsible for regular analysis, validation, discussion, and improvement of the policies, standards and activities related to data security management. The audit work can be performed by an internal or external auditor of the organization, who should be independent of the data and processes involved in the audit. Data security audit aims to provide assessment and advice to the organization and external regulatory agencies.
Goals	a) It is to make sure that the security requirements and regulatory requirements of the organization are met; b) It is to discover data security risks in a timely manner, and improve the data security measures; c) It is to put forward suggestions on data security management to promote the optimization

	and improvement of data security.

4.2.7 Data quality

4.2.7.1 Data quality requirements

The capability item of data quality requirements is shown in Table 4.2.7.1.

Table 4.2.7.1 Data Quality Requirements

Capability Item ID	Qua-1
Overview	Data quality requirements aim to define the data quality goals, develop rules to measure data quality based on the business needs and data requirements, including the technical indicators, business indicators and corresponding validation rules and methods for measuring data quality. Data quality requirements are the basis for measuring and managing data quality, which should be formulated and managed in a unified manner based on the data management goals, business management needs of the organization, and industry regulatory requirements, with reference to relevant standards.
Goals	a) Clear data quality management goals should be generated; b) Various data quality management requirements should be clarified; c) A continuously updated database of data quality rules should be established.

4.2.7.2 Data quality check

The capability item of data quality check is shown in Table 4.2.7.2.

Table 4.2.7.2 Data Quality Check

Capability Item ID	Qua-2
Overview	Data quality check refers to real-time monitoring of data quality of the organization according to relevant technical indicators and business indicators, validation rules and methods in the data quality rules to find out data quality issues and provide feedback to the data management personnel.
Goals	a) It is to develop a data quality inspection plan; b) It is to comprehensively monitor data quality of the organization; c) It is to establish a mechanism for data quality problem management.

4.2.7.3 Data quality analysis

The capability item of data quality analysis is shown in Table 4.2.7.3.

Table 4.2.7.3 Data Quality Analysis

Capability Item ID	Qua-3
Overview	Data quality analysis aims to analyze the data quality problems found during data quality check and related information, identify the causes that affect data quality, and define the priority of data quality problems to serve as a reference for data quality improvement.
Goals	a) It is to establish the methods for data quality problem assessment and analysis; b) It is to regularly analyze the data quality of the organization; c) It is to establish a continuously updated data quality knowledge base.

4.2.7.4 Data quality improvement

The capability item of data quality improvement is shown in Table 4.2.7.4.

Table 4.2.7.4 Data Quality Improvement

Capability Item ID	Qua-4
Overview	Data quality improvement is the result of data quality analysis. It is required to develop and implement data quality improvement plans, including erroneous data correction, business process optimization, application system problem repair, etc., and develop a data quality problem prevention plan to ensure that the results of data quality improvement are effectively maintained.
Goals	a) It is to establish a strategy for continuous improvement of data quality; b) It is to develop a data quality improvement plan; c) It is to establish a good data quality culture.

4.2.8 Data standard

4.2.8.1 Business terms

The capability item of business terms is shown in Table 4.2.8.1.

Table 4.2.8.1 Business Terms

Capability Item ID	Sta-1
Overview	Business terms refer to the descriptions of business concepts in the organization, including Chinese name, English name, term definition, etc. Business data management is to develop a unified management system and process, and conduct unified management of the creation, maintenance and release of business terms, thereby promoting the sharing of business terms and application within the organization. Business terms are the foundation for the understanding and application of data within the organization. The consistency of understanding of specific technical terms within the organization can be guaranteed through management of the business terms.
Goals	a) Business terms should accurately describe the meaning of business concepts; b) The organization should establish a comprehensive published dictionary of business terms; c) The definition of business terms should follow relevant standards; d) The creation and change of business terms should be managed in a unified manner through the management process; e) The management and application of business terms should be improved through data governance.

4.2.8.2 Reference data and master data

The capability item of reference data and master data is shown in Table 4.2.8.2.

Table 4.2.8.2 Reference Data and Master Data

Capability Item ID	Sta-2
Overview	Reference data is the data used to classify other data. Reference data management is to manage the defined data range, including standardized terms, code value, and other unique identifiers, business definition of each value, control of the business relationship between the data range list and across different lists, as well as consistent and shared use of relevant reference data. Master data is the core business entity data that needs to be shared across systems and departments in an organization. Master data management aims to manage the master data standards and content to realize consistent and shared use of master data across systems.
Goals	a) It is to establish a recording system (SOR) for identifying the reference data and master data;

	b) It is to establish accurate records of reference data and master data; c) It is to establish the management standards for reference data and master data.

4.2.8.3 Data element

The capability item of data element is shown in Table 4.2.8.3.

Table 4.2.8.3 Data Element

Capability Item ID	Sta-3
Overview	It aims to allow the data owners and users to have a consistent understanding of the data based on the standards of the core data elements in the organization.
Goals	a) It is to establish unified data element management standards; b) It is to establish a unified directory of data elements.

4.2.8.4 Index data

The capability item of index data is shown in Table 4.2.8.4.

Table 4.2.8.4 Index Data

Capability Item ID	Sta-4
Overview	Index data is the data used by the organization to measure a certain goal or thing in the process of business analysis, which generally includes index name, time and value, etc. Index data management refers to unified and standardized definition, collection and application of index data required by the organization for internal business analysis, which is used to improve the data quality of statistical analysis.
Goals	a) It is to establish the standards for classification and format of index data; b) It is to establish a unified index data dictionary within the organization; c) The index data definition should clearly describe the index meaning, etc.;

4.2.9 Data life cycle

4.2.9.1 Data requirements

The capability item of data requirements is shown in Table 4.2.9.1.

Table 4.2.9.1 Data Requirements

Capability Item ID	Lif-1
Overview	Data requirements refer to the organization's description of classification, definition, distribution and circulation of data generated and used in the process of business operations, business analysis and strategic decision-making. In the data requirements management process, the required data are identified, the priority of data requirements is determined, and the data requirements are documented and managed by means of files.
Goals	a) It is to establish a data requirements management system to uniformly manage various data requirements; b) The stakeholders of data should have a consistent understanding of data requirements to meet the business needs; c) Various data requirements should be sorted out and defined; d) The naming, definition and representation of data should comply with relevant standards issued by the organization.

4.2.9.2 Data design and development

The capability item of data design and development is shown in Table 4.2.9.2.

Table 4.2.9.2 Data Design and Development

Capability Item ID	Lif-2
Overview	Data design and development refers to the process of designing and implementing data solutions, providing data application, and continuing to meet the data requirements of the organization. The data solutions include database structure, data collection, data integration, data exchange, data access, and data products (reports, user views), etc.
Goals	a) It is to design the data structure and solutions that meet the data requirements; b) It is to implement and maintain the solutions that meet the data requirements; c) It is to make sure that the solutions are consistent with the data architecture and data standards; d) It is to guarantee the integrity, security, availability and maintainability of data.

4.2.9.3 Data operation and maintenance

The capability item of data operation and maintenance is shown in Table 4.2.9.3.

Table 4.2.9.3 Data Operation and Maintenance

Capability Item ID	Lif-3
Overview	Data operation and maintenance refers to daily operation and maintenance of data collection, data processing, data storage, etc. after the data platform and related data service construction is completed online and put into operation, which aims to ensure normal operation of the data platform and data services and provide continuously available data content for data applications.
Goals	a) The internal and external data providers of the organization should provide the data that meet the business needs according to the agreed service level. b) Stable operation of data-related platforms and components should be guaranteed.

4.2.9.4 Data retirement

The capability item of data retirement is shown in Table 4.2.9.4.

Table 4.2.9.4 Data Retirement

Capability Item ID	Lif-4
Overview	Data retirement refers to the management of historical data, retention and destruction of historical data according to the requirements of laws and regulations, business, technology, etc., as well as archiving, migration and destruction of historical data to ensure that the organization manages the historical data based on the needs of external regulatory agencies and internal business users, rather than just meeting the requirements of information technology.
Goals	a) The scheme for use, retention, and removal of historical data should meet the internal and external business needs and regulatory requirements of the organization; b) Procedures and standards should be established to standardize the collection, scheme design and implementation of data retirement requirements.

Section 3 Assessment Measurement and Result Analysis

4.3.1 Measurement system

4.3.1.1 The applicability of the scope of data assessment should be evaluated based on the business scenarios and data risks of the organization, covering each capability area and each capability item in it insofar as practicable. For each capability item, corresponding specific assessment standards should be set according to the 5 maturity levels, and corresponding assessment results should be given based on the degree to which the actual situation of the assessed organization meets a specific standard.

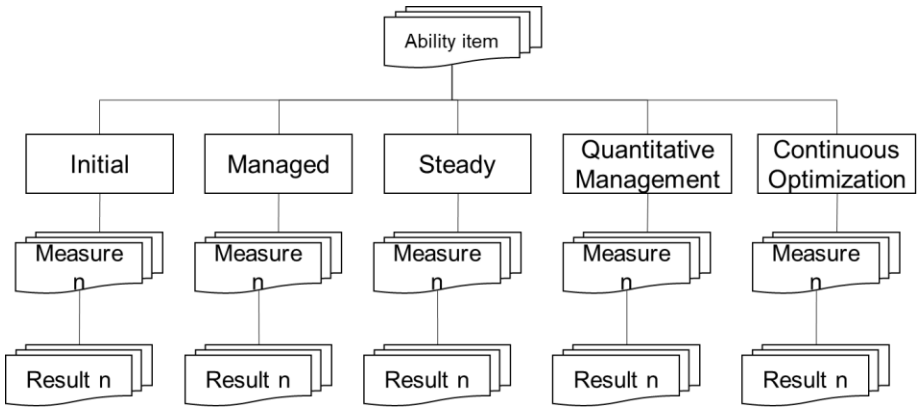

Figure 4.3.1.1 Maturity Measurement System

4.3.2 Comprehensive maturity rating

4.3.2.1 The comprehensive data management maturity level of an organization or institution can be evaluated by means of “cask principle”, “principle of maximum”, “weighted average”, etc., with overall consideration of the assessment results of capability area and capability item.

4.3.3 Evaluation result analysis

4.3.3.1 The results of data management maturity assessment can be visualized by using the maturity heat map. Figure 4.3.3.1 shows a matrix of several cells, and each cell represents the measurement results of a certain measurement standard. The assessment results present the extent to which the assessed item meets the characteristic requirements of a certain maturity level. A comprehensive and in-depth analysis of the assessment results will facilitate continuous improvement of the data management level and the maturity level.

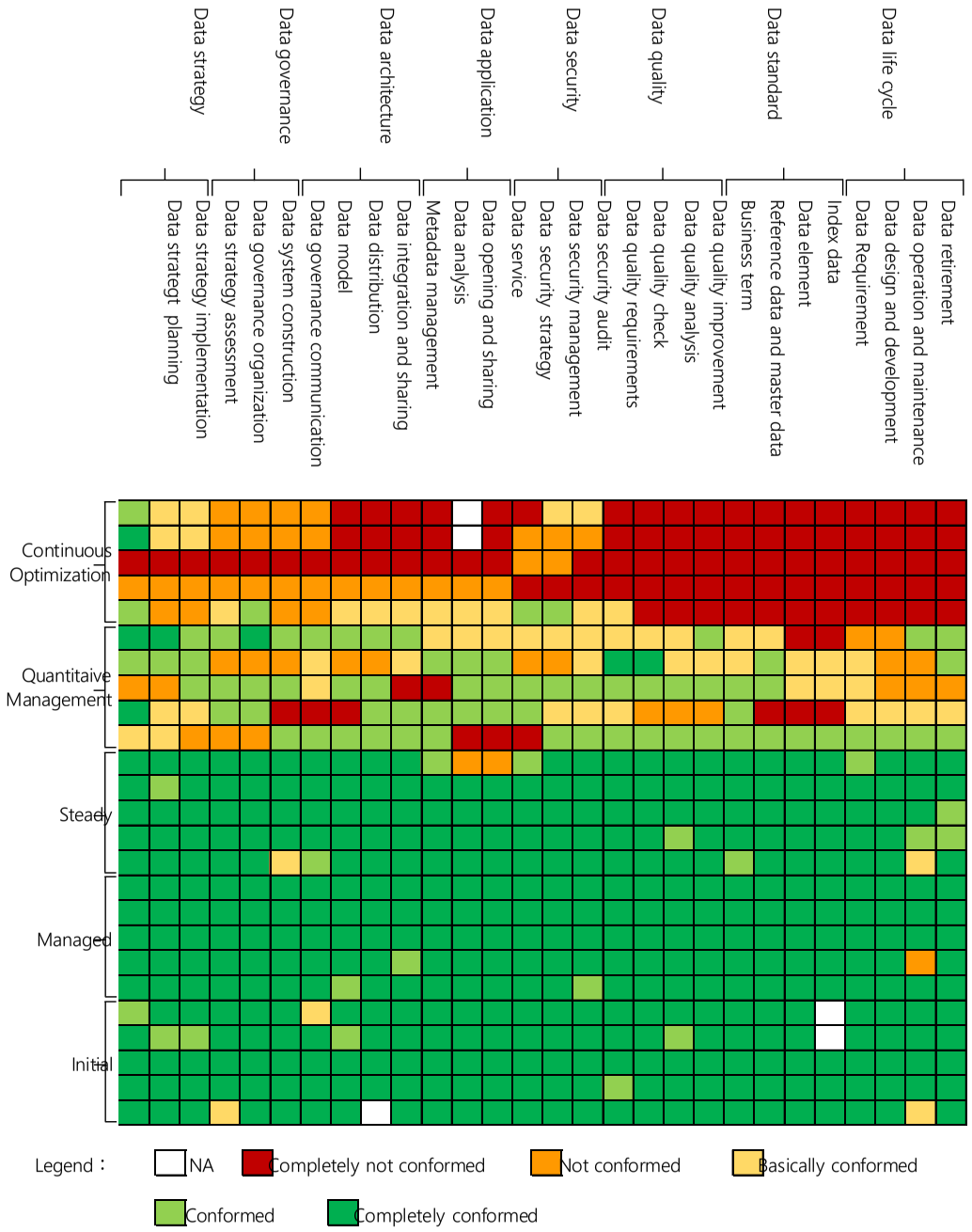

Figure 4.3.3.1 Example of Visualization of Evaluation Results

Chapter 5 Survey and Certification

Section 1 Inspections of Products

5.1.1 General requirements

5.1.1.1 The ISC Rules/guidelines specify that data quality assessment (if necessary) should be performed to the systems/products according to the requirements of this chapter.

5.1.1.2 The data scope of data quality assessment can be determined according to the functional goals of the system. Please refer to Section 3.3.3-3.3.4 for the methods.

5.1.2 Documents

5.1.2.1 In the case of data quality assessment and security assessment, the following drawings and documents should be submitted for information:

(1) System data design specifications

①Description of data input category, type, interface definition, etc.;

②Description of data storage, forwarding, access, etc.;

③Description of database identifier, status, naming conventions, etc.;

④Description of database installation, deployment, management, etc.

(2) Software manual

①Software overview;

② Software operating environment;

③Software architecture;

④Software function.

(3) Data classification instructions (when applicable)

①Methods and principles of data classification;

②Initial assessment method of data security.

(4) Data grading instructions

①Methods and principles of data grading;

②Data security assessment methods and assessment results.

5.1.3 Assessment process

5.1.3.1 Product assessment is divided into document validation and technology assessment.

5.1.3.2 Document validation: drawings and documents required in 5.1.2 are checked according to relevant requirements of Chapters 2 and 3 of the Guidelines, to determine the scope and criteria of data assessment.

5.1.3.3 Technology assessment: technology assessment is performed to the retained data of the system/product applying for assessment for 3 months of continuous operation according to the requirements of Chapter 3.

5.1.3.4 The ship data assessment report (product) should be signed by ISC after qualification through documents validation and technology assessment.

Section 2 Survey of Ships

5.2.1 General requirements

5.2.1.1 This chapter is applicable to the ships for which class notation Data Quality(x) is to be obtained.

5.2.1.2 System/product operating data for at least 3 consecutive months should be kept for the ships for

which class notation Data Quality(x) is to be obtained.

5.2.1.3 The survey requirements specified in this chapter are supplementary to all applicable ship survey requirements. This survey may be performed simultaneously with the same type of survey (i.e. annual, intermediate and special surveys) as specified in Section 2, Chapter 5, PART ONE of the *Rules for Classification of Sea-going Steel Ships*, with the same survey interval.

5.2.2 Initial survey

5.2.2.1 Document validation

(1) Relevant system/product information required by 5.1.2 of this chapter should be reviewed;

(2) The documents should be reviewed in accordance with the *Rules for Classification of Sea-going Steel Ships*.

5.2.2.2 Survey and testing

(1) The supplier should submit the description of data quality assessment requirements for the system/product applying for data quality assessment;

(2) The supplier should submit 3 months of operating data of the system/product applying for data quality assessment;

(3) ISC should evaluate the data of the system/product according to Chapter 3 of the Guidelines;

(4) If the requirements are satisfied, ISC should assign class notations to the ship of application.

5.2.2.3 Assignment of class notations

(1) After the surveys and tests are completed, ISC should assign class notations of data quality to the ships.

Section 3 Surveys After Construction

5.3.1 Annual survey

5.3.1.1 Before annual survey of the ship, ISC should submit the latest operating data of no less than 6 months to the agency that carries out the survey, and should also submit the report containing the following contents, which should include at least the following contents since the last annual survey:

(1) Overall operating condition of the system/product;

(2) Database change records (if any) of system/product;

(3) Data storage records of system/product;

(4) Data related fault/failure situation and cause analysis (if any) of system/product.

5.3.1.2 During annual survey, ISC should carry out technology assessment of the operating data for the ship, and should also checked the following items on board:

(1) Operating status of equipment/system;

(2) Database running log of equipment/system;

(3) Checking the operation and maintenance management records of equipment/system.

5.3.1.3 If the system/product is unqualified after data quality assessment, ISC should give suggestions for rectification within the time limit or cancel the ship data quality class notations.

5.3.1.4 If data quality rectification of the ship has not been completed beyond the time limit, ISC will cancel the ship data quality class notations.

5.3.2 Occasional survey

5.3.2.1 When there are significant changes in ship system/product and/or the data retention conditions are not met, generally, the equipment/system supplier should apply to ISC for temporary ship data quality assessment and make sure that the ship data quality meets the requirements of the Guidelines.

5.3.2.2 At occasional survey, ISC should check the following items:

(1) System/product change manual;

(2) System/product database architecture change specification;

(3) Determining whether to carry out technology assessment of data quality according to the situation.

5.2.3.3 The ship data should meet the conditions and passed the assessment, a formal report should be issued, and class notations should be assigned.

5.3.2.4 If the equipment system data quality assessment fails to meet the requirements, ISC should give suggestions for rectification within the time limit or cancel the class notations of ship data quality.

5.3.2.5 If data quality rectification of the ship has not been completed beyond the time limit, ISC will cancel the ship data quality class notations.

Appendix 1 Data Security Requirements

Section 1 General Provisions

1.1.1 General requirements

1.1.1.1 The overall goal of data security is to guarantee the confidentiality, integrity and availability of data.

1.1.1.2 Data export should meet relevant laws and regulations of the flag State and the Administration.

1.1.1.3 For ship data exchange, please refer to relevant technical clauses of ISO19645, ISO19847 and ISO19848.

1.1.1.4 Ship data security should at least meet relevant requirements of Section 5, Chapter 3 of the *Guidelines for Requirement and Security Assessment of Ship Cyber System*.

Section 2 Data Security Framework

1.2.1 Data security framework

1.2.1.1 Data security provides the data security requirements from data collection, storage, transmission, processing, sharing, destruction and other data life cycle dimensions. The data security framework is shown in Figure 1.2.1.

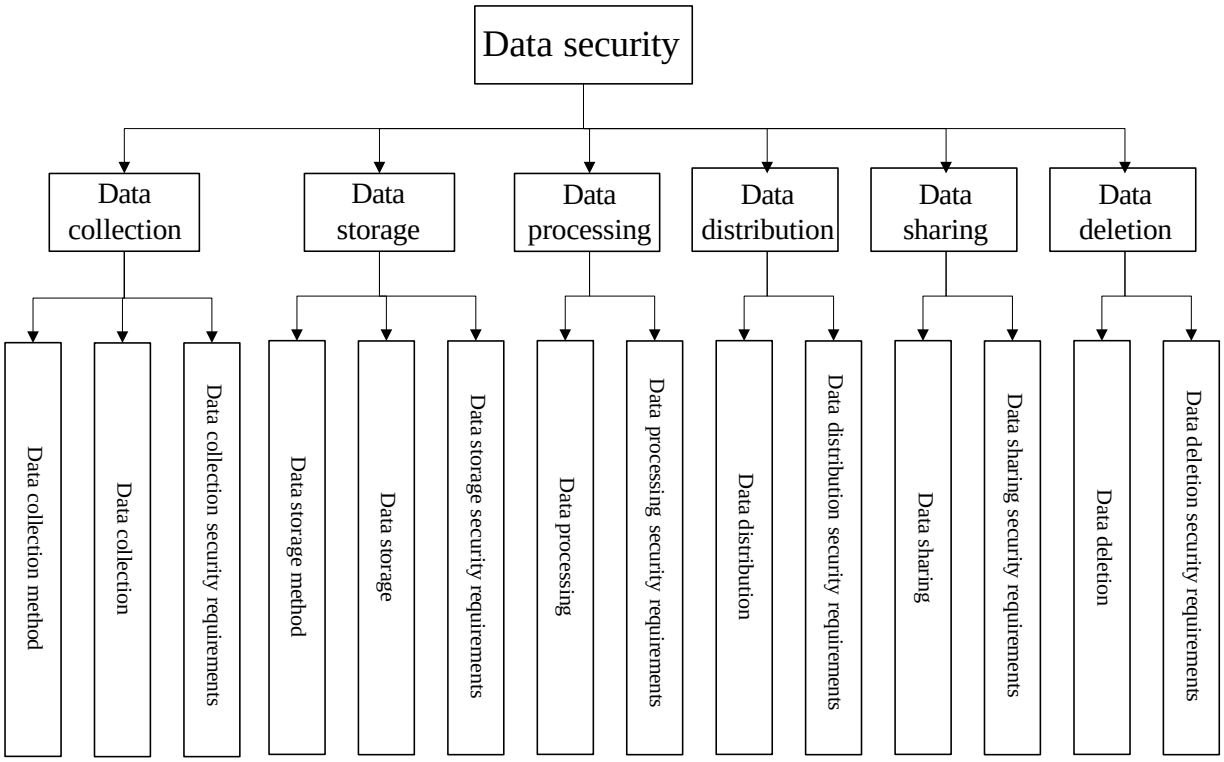

Figure 1.2.1 Data Security Framework

Section 3 Data Collection

1.3.1 Concept of data collection

1.3.1.1 The purpose of data collection is to obtain data, and the data collection methods include but are not

limited to:
(1) Data collection by sensors. It includes the data collected and generated during the operation of various ship systems;
(2) Communication network data. Data obtained through ship-to-ship, ship-to-shore communication;
(3) System data. It includes business data collected and generated during the operation of the system within the organization, as well as a large amount of operation and maintenance and log data generated during operation of various systems, programs and services.
1.3.1.1 Data collection mainly includes but is not limited to: discovery of data sources, transferring of data, generation of data, cache data, creation of metadata, data conversion, data integrity validation, etc.

1.3.2 Security requirements

1.3.2.1 During data collection, the security requirements should:
(1) Define the purpose and usage of the collected data, and clarify the data collection source, type of data collection and range of collection data;
(2) Follow the “principles of compliance”, and ensure the legality, legitimacy, and necessity of data collection;
(3) Follow the “principle of minimization”, and only collect the minimum data required by the business;
(4) Follow the “quality assurance principle”, and develop the strategies, procedures and requirements for data quality assurance;
(5) Follow the “principle of safety assurance”, classify and mark the collected data, and implement corresponding security management strategies and supporting measure for different types and levels of data. Take necessary safety control measures for the data collection environment, facilities and technology.

Section 4 Data Storage

1.4.1 Concept of data storage

1.4.1.1 Data storage refers to static storage of data on the data platform, and the stored data includes collected data, analyzed and processed data, etc. The storage system can be a relational database, a non-relational database, etc., and should support the storage of different types and formats of data, and should provide multiple data access interfaces, such as file system interface, database interface, etc. The stored data should be properly protected by the organization until the data is completely deleted.
1.4.1.2 Full consideration should be given to the security risks of saving data by a third-party data storage platform. Due to intellectual property rights, laws and regulations, etc., even if the organization can effectively control the data in the storage system, including personal information or health data, etc., but it may not be the owner of the data, and the organization still needs to bear the responsibility of data storage management.
1.4.1.3 The data storage mainly includes but is not limited to data encoding, data encryption and decryption, hierarchical storage of hot and cold data, data archiving, data backup, data update, data access, etc.

1.4.2 Security requirements

1.4.2.1 When carrying out data storage, the organization should:
(1) Store the data of different types and levels separately, and take physical or logical isolation mechanism;
(2) Follow the “principle of safety assurance”, and give major consideration to the following aspects:
①Storage architecture security;
②Logical storage security;
③Storage access control;
④Data copy security;

⑤Data archiving security;

⑥Data timeliness management.

(3) Establish a data storage redundancy strategy and management system, as well as operating specifications for data backup and recovery;

(4) Relevant requirements of Section 5, Chapter 3 of the *Guidelines for Requirement and Security Assessment of Ship Cyber System*;

Section 5 Data Processing

1.5.1 Concept of data processing

1.5.1.1 Data processing refers to a series of operations that extract information from data through such technologies as data analysis and data visualization and extract useful knowledge and value.

1.5.1.2 Data processing mainly includes but is not limited to data query, data reading, data indexing, batch processing, interactive processing, stream processing, data statistical analysis, data predictive analysis, data association analysis, data visualization, generating analysis reports, etc.

1.5.2 Security requirements

1.5.2.1 When carrying out data processing, the organization should:

(1) Define the purpose and scope of data processing according to the requirements of laws and regulations on protecting personal information and important data;

(2) Establish an internal responsibility system for data processing, and make sure that the stated purpose and scope of data use are not exceed during analysis, processing and use of data;

(3) Follow the “principle of minimum authorization” and provide a fine-grained access control mechanism of data;

(4) Follow the “principle of safety assurance”, and major consideration should be given to the following aspects:

①Distributed processing security;

②Data analysis security;

③Data encryption;

④Data Masking;

⑤Data traceability.

(5) Follow the “principle of auditability”, and record and manage the operations in the data processing;

(6) Conduct risk assessment on the data processing results, and avoid inclusion of recoverable sensitive data in the processing results.

Section 6 Data Distribution

1.6.1 Concept of data distribution

1.6.1.1 Data distribution refers to the delivery of different forms of data including raw data and processed data to other roles within the organization, external entities or the public, etc. Data distribution is subject to online or offline distribution, etc.

1.6.1.2 The reasons for data distribution include but are not limited to:

(1) Data exchange between departments within the organization;

(2) Generating reports for external parties, such as statistics from government departments;

(3) Data exchange between enterprises, such as providing customers with usage reports, etc.;

(4) Selling data to other organizations;

(5) Business realization requirements.

1.6.1.3 Data distribution mainly includes but is not limited to: data transmission, data export, data exchange,

data transaction, data sharing, etc.

1.6.2 Security requirements

1.6.2.1 When carrying out data distribution, the organization should:

(1) Follow the “principle of responsibility not to be transferred with data”;

(2) Before data distribution, conduct risk assessment of the data to ensure that the risks after data distribution are tolerable, and clarify the data protection responsibility of the data recipient through the contract;

(3) Before data distribution, assess the sensitivity of data, and desensitize sensitive information that needs to be distributed based on the assessment results.

(4) Follow the “principle of auditability”, and record relevant information including time, distributed data, data recipient, etc.;

(5) Evaluate the transmission security risks in data distribution to ensure the security of data transmission;

(6) Provide an effective data security sharing mechanism;

(7) Establish a review system for data release, and define the content and scope of data release to carry out regular review of the data released.

Section 7 Data Sharing

1.7.1 Concept of data sharing

1.7.1.1 The process of data sharing involves different business scenarios such as providing data to third parties and disclosing data to the outside world.

1.7.1.2 The ships, integrators, suppliers, shipowners and ship management companies should share data with overseas organizations and disclose data to the outside world according to the requirements of relevant laws and regulations of the flag State.

1.7.2 Security requirements

1.7.2.1 A unified data distribution platform can be established while implementing relevant management system regulations on data sharing security to serve as the only exit for data to leave the data security domain combined with the data security domain technology, thereby effectively managing data sharing behaviors and preventing security risks such as data theft and leakage while meeting the business needs. For a unified data distribution platform, it needs to integrate all data sharing business scenarios, such as sharing with third parties, downloading data analysis reports to the office terminal equipment, etc. based on the processing and analysis results of raw data, and realizing a differentiated online approval process for each type of data sharing scenario.

Section 8 Data Deletion

1.8.1 Concept of data deletion

1.8.1.1 Data deletion refers to the deletion of data and its copies by the organization. If the data comes from an external real-time data stream, the link to the real-time data stream should also be disconnected.

1.8.1.2 The reasons for data deletion include but are not limited to:

(1) Reducing the risk of data leakage and preventing data from being inappropriately distributed or processed;

(2) Deleting irrelevant or incorrect data, the data is no longer associated with the original purpose of use, or the data is incorrect;

(3) Data deletion after the business is completed, and the data business fulfills the service goals with no need to save relevant data anymore;

(4) Meeting the customer's data deletion requirements, except for the data required to be retained by laws and regulations.

1.8.1.3 Data deletion mainly includes but is not limited to: deleting metadata, deleting raw data and its copies, disconnecting from external real-time data streams, deleting the data access interfaces, and destroying unrecoverable data, etc.

1.8.2 Security requirements

1.8.2.1 When carrying out data deletion, the organization should:

(1) Delete relevant data that has exceeded the data retention period, and implement according to relevant regulations if there are clear regulations on the retention period;

(2) Establish a corresponding data deletion mechanism according to data classification and grading requirements, clarify the data, methods and requirements for data destruction, and define the scope and process of data destruction;

(3) Follow the "principle of auditability", establish a data deletion strategy and management system, and record the operation time, operator, operation method, data content and other related information of data deletion.