INTERNATIONAL MARITIME ORGANIZATION

E

MARITIME SAFETY COMMITTEE
109th session
Agenda item 5

MSC 109/5/4
27 September 2024
Original: ENGLISH
Pre-session public release: ☒

DEVELOPMENT OF A GOAL-BASED INSTRUMENT FOR MARITIME AUTONOMOUS SURFACE SHIPS (MASS)

Data quality and management for the safety of MASS

Submitted by IACS

SUMMARY

Executive summary:	A high volume of data is generated on MASS, which is critical for decision-making application. This document, while providing a brief on general data quality and management aspects, brings out aspects specific to MASS and suggests a high-level approach towards data quality, which is essential for the safety of MASS.
Strategic direction, if applicable:	2
Output:	2.23
Action to be taken:	Paragraph 20
Related document:	MSC 109/5

Background

1 Data-driven decision support for various onboard applications is on the rise. Increasing use of automation systems and development of MASS with various modes of operation is leading to increased data being processed. Consequently, the management of data has gained significance as poor data can affect human and/or system decision making. This is, to some extent, recognized by the introduction of the relevant provision on data management and quality in section 9.8, part 2, chapter 9 of the draft MASS Code, stating (MSC 109/5, annex):

> "*Efficient data management systems should be incorporated to ensure data accuracy, integrity, and quality [and design systems to leverage data for enhanced performance and decision-making].*"

Discussion

Data quality

2 Data quality refers to the degree to which dimensions of data meet quality requirements, i.e. a need or expectation that is stated, recommended or obligatory. It is a measure of the extent of data meeting the requirements of the user. High-quality data are essential for making informed decisions, ensuring operational efficiency and maintaining compliance.

3 Data management refers to the set of practices which involve deploying people, processes and technologies to control or enhance data and information throughout their lifecycle. Data management is a broader term covering entire data lifecycle, whereas data quality management is a subset of data management which specifically focuses on improving and maintaining data quality.

4 In that sense, data quality management involves the selection (e.g. translation of business expectation, data criticality, risk or priority analysis, data profiling) and definition of reference to assess the quality of data (e.g. data quality dimensions definition, metrics definition, acceptability thresholds); construction of reliable process (e.g. data transmission, storage, traceability/logging); measurement of data against those references (e.g. data quality assessment, data quality control); and reporting.

5 Data quality dimensions may be used as a vocabulary to define data quality requirement as measurable rules, assessing risk and providing a basis to translate business requirements through completeness, validity, accuracy, consistency, timeliness, integrity, reasonability, uniqueness, etc. Risk assessment plays a critical role in addressing potential threats to data quality. Threats to data quality can be assessed and addressed to ensure that reliable, secure and valuable data is made available to safety critical systems. Data in different activities can be measured by different dimensions, such as data identification, data acquisition, data storage, data integration and data processing.

6 The inherent data quality is assessed through completeness, validity, accuracy, consistency, timeliness, integrity, relevance, etc. in the initial phase. However, during the operational phase the data quality characteristics relevant to systems on which they operate gain significance.

7 System-dependent data quality refers to the degree to which data quality is influenced while the data is accessed, stored and retrieved in a computer system. Typical dimensions where data quality depends on the technology domain are availability, portability and retrievability.

8 Data quality management should be supported by proper data governance, tools or process to ensure continuous support; analysis; investigation of poor data quality (e.g. root cause investigation); reporting (e.g. data quality performance or data quality incident to responsible person); and sustainment of the programme. The roles, responsibility control and decision-making over the management of data are to be defined.

9 Data quality acceptance is an important step in data quality and is dependent on intended data application. The acceptance criterion for each data quality dimension and the corresponding metrics should be identified and recorded during the design stage. The feasibility of measuring the data quality dimension under the current technical conditions should be considered.

Data quality and data management in MASS

10 Achieving the desired performance from autonomous ship systems depends, to a large extent, on the data quality and data management. Systems which provide decision support require high-quality data to meet the desired design functionality. Poor data quality not only affects the decision system but can have a detrimental effect on safety of personnel, ship and environment. Therefore, data quality and data management are required to be addressed through a structured approach using suitable tools to meet the desired data quality requirements.

11 Data management is an iterative process that should be implemented in a sustainment loop for continuous improvement. In the design phase, the system data quality requirements should be determined according to the objective of MASS and its concept of operations, including the scope of data, the dimensions of data and the acceptance criteria to be met. During the operation of the ship, the data quality should be measured according to the design requirements.

12 Data in different activities can be measured through different dimensions. For example, in the data identification stage, consistency and compliance could be the dimensions, whereas during the data acquisition stage completeness, continuity and timeliness could be the data quality dimensional aspects. In a similar way, the data quality dimensions are to be identified for other stages of data life cycle, i.e. data storage, data integration and data processing.

13 Data management is critical for the safe and efficient operation of ships using advanced autonomy. The vast amount of data generated by sensors, systems and external sources needs to be handled effectively to ensure accurate decision-making and prevention of errors.

14 In addition to general data management methodologies like data governance and data quality assessment, the following considerations and challenges which are specific to MASS are to be addressed:

.1 real time processing;

.2 data volume;

.3 cybersecurity, and;

.4 integration with various autonomous models.

15 A MASS may have systems which may range from simple decision support systems to systems with advanced autonomy having capability to control ship systems. Such systems require high-quality data.

16 To address aspects unique to MASS, specific data management practices could be required, such as, but not limited to, edge computing, seamless data fusion, sensor calibration, real-time data visualization, etc.

17 Data quality measurement should be carried out at regular predefined intervals in a MASS. The assessment should be conducted based on quality requirements.

18 Procedures are to be identified and implemented to ensure verification of data quality, testing and validation requirements specific to the level of autonomy.

19 In the process of assessing the data quality, the content and structure of data should be learned. In most cases, metadata predefines the content and structure of data (e.g. data field type, data format, value domain precision). The metadata should include the relevant requirements from the following three dimensions:

.1 function specific requirement: describe the data characteristics associated with the function;

.2 technology requirement: data format, data source, data type, etc, and;

.3 data management requirement: data security, data standard, etc.

Action requested of the Committee

20 The Committee is invited to consider the foregoing and take action, as appropriate.