Author(s) / Submitter(s) Singapore’s Agency for Science, Technology and Research (A\*STAR) Institute

of High Performance Computing (IHPC) and Maritime and Port Authority of

Singapore (MPA)

Essential Practices for Data and Management in Developing Automated Near-Miss Incident Identification

# Summary

A vessel collision incident in a given area may be preceded by several near-miss cases. The early detection, identification and analysis of these near-miss cases play a crucial role in aiding Vessel Traffic Service (VTS) operators and coastal authorities in the prompt identification of any underlying causes so as to proactively predict and prevent future vessel collision incidents. The IALA guideline G1118, published in January 2022 and titled 'MARINE CASUALTY/INCIDENT REPORTING AND RECORDING, INCLUDING NEAR-MISS SITUATIONS AS IT RELATES TO A VTS', recommends reporting near-miss cases for every VTS. This international recognition of learning from incidents and near-miss situations will eventually facilitate the proactive measures to prevent future accidents and damage towards zero incident.

The current detection and recording scheme primarily relies on manual reporting As a result, human factors such as excessive cognitive workload and inconsistent/subjective opinions lead to potential underreporting of near-miss cases. In recent years, the AI-based method for automating near-miss identification has attracted increasing attention. However, such data-driven solutions, in addressing this problem, frequently suffer from a myriad of technical challenges such as AIS data quality issues, incomplete or even missing information in AIS datasets.

This document proposes a technical recommendation on essential practices for digital data preparation in developing a near-miss case detection tool. Two aspects are addressed in this document, including recommendation on digital data quality management and structured expert knowledge incidents database.

## Purpose of the document

This document supports and contributes to the task of " Develop a discussion paper on digitalisation in the scope of IALA ", in alignment with the 2023-2027 work plan of the DTEC Committee. This document leverages the practical AI application experience of Singapore’s Agency for Science, Technology and Research (A\*STAR) Institute of High Performance Computing (IHPC) and the Maritime and Port Authority of Singapore (MPA) in the field of digitalization for maritime operations and proposes to properly measure and ensure AIS data quality and establishment of the incident knowledge database as an important part of automating near miss case detection in port waters to support maritime operation enhancement, digitalisation and transformation.

## Related documents

Related Recommendation/Guidelines on near miss reporting:

* IALA G1118, published in January 2022 and titled 'MARINE CASUALTY/INCIDENT REPORTING AND RECORDING, INCLUDING NEAR-MISS SITUATIONS AS IT RELATES TO A VTS,'

# Background

In the current landscape of maritime management, the importance of near miss cases for safety analysis and management is underestimated due to the lack of timely and accurate reporting of most ship collision near-miss cases that still rely on manual input from vessel operators or pilots. Due to varying subjective opinions and omissions, only a fraction of near-miss cases is captured through this manual method. Detection and identification methods based on AIS data can effectively address these human factors and reduce human workloads, significantly enhancing the comprehensive identification of near-miss cases. Automating the detection of near-miss cases through data-driven methods is becoming increasingly feasible with the availability and accumulation of AIS data.

However, automating the near miss case detection process is not without challenges as it requires high quality of AIS data and there are variations in the quality of AIS data from different VTS or other sources. Poor AIS data quality is a result of processes such as radio transmission, relaying, storage, and preprocessing. Proper AIS data quality recognition, assessment and enhancement is therefore important to enable the digital near miss detection technology development.

Among AIS data quality, some common quality issues include:

* Irregular/Inaccurate abnormal/erroneous vessel movement information (illogical or erratic position points)
* Insufficient spatiotemporal resolution of vessel position information
* Inconsistent/Incorrect static vessel information such as vessel length, vessel type

Low-quality data sources can easily lead to the failure of the entire automated detection solution. For instance:

* Missed cases that should have been captured
* An excessive number of false detections that obscure genuinely valuable information
* An inability to produce sufficient visual playback of cases, hindering experts from conducting accurate investigative analysis

Besides qualified AIS data for supporting near miss case detection, an initial near miss knowledge base is also important for detecting near miss cases reflecting domain experts’ risk appetite at the corresponding port water. The initial near miss knowledge base will comprise of near miss cases reported by control operators (domain experts) and candidate near miss cases detected and recognized by domain experts. Such near miss knowledge base will be a significant contributing factor, providing sufficient and detailed experiential knowledge for enhancing and refining the initial near miss case detection models to comprehensively identify the systematic characteristics of near-miss cases in the port water. This near miss knowledge base will include the following key contents:

* An insufficient number of incidents recorded to reflect expert’s recognition on near miss cases following their experience
* Corresponding expert knowledge records for usual/typical/good vessel manoeuvring cases
* Systematic/structured records of differing opinions among experts

To address these issues, it is necessary to provide information and guidance in the area above to support the development and deployment of digital near miss case detection tools which can enhance maritime operations.

# Discussion

The Singapore port has established an application practice in AIS data-driven near-miss detection. Currently, a near miss case detection tool has been successfully deployed and evaluated to assess its usefulness in a real-time operational environment. Following the practical development, we believe that the following recommendations could provide valuable information to better facilitate the digital technology development in automating near-miss case detection worldwide.

**AIS Data Quality Assurance**

* Ensure qualified AIS data is available and properly captured to support automating near-miss case detection.

In AIS data, the dynamic movement information of vessels is the most critical. Essential parameters include vessel coordinates, speed, course, and heading. It is crucial to avoid any data degradation or resolution compression during data storage and management, particularly in congested traffic areas. According to the relevant standards, position resolution should be 0.001 minutes, speed and course should be maintained at 0.1°, and heading at 1°.

* Ensure qualified AIS data is properly stored and managed to maintain data quality for retrospective analysis and comparison.

The time update frequency of vessel movement information may be compressed or downgraded during data storage. Such compression or downgrading will negatively affect the performance of automating near miss case detection, especially during encounters or other high collision-risk maneuvers. It is essential to maintain the highest recording frequency as specified by relevant standards, which is 2 to 12 seconds.

* Ensure the quality of AIS data by using proper preprocessing techniques

For the quality issue, such as irregular/inaccurate vessel movement information or deficiencies in update frequency, the preprocessing and cleansing of time-series data are essential and strongly recommended. Common and effective techniques include the slide window technique and dynamic interpolation.

**Near miss Case Knowledge Base Construction**

* Establishment of safety-critical incident dataset

In building an expert knowledge base for initiating the automation of near miss case detection, it is essential to collect sufficient safety-critical incident cases that can represent safety-critical situations in the corresponding water and to apply more appropriate and multi-level tagging. For example: accidents, near-misses, potential near-misses, and safe passages.

* Furnishment of corresponding expert inputs for safety-critical cases

Expert analysis and assessment criterion should be recorded in a digital and structured format. Additionally, it is important to encourage detailed and extensive written documentation of the reasons behind these judgments to facilitate future applications of AI language models.

* Collection of different subjective opinions for capturing comprehensive inputs for near miss detection

For each case, it is crucial to preserve as many differing opinions from multiple experts as possible. Additionally, the structured data should include their roles, experience, and other relevant personal information, as well as the important confidence level associated with their judgments.