Department of Applied Mechanics

Data Sources for Quantitative Marine Traffic Accident Modeling

Marina Ladan, Maria Hänninen

Aalto University

SCIENCE + TECHNOLOGY

REPORT

Aalto University publication series
SCIENCE + TECHNOLOGY 11/2012

Data Sources for Quantitative Marine Traffic Accident Modeling

Marina Ladan, Maria Hänninen

Aalto University
School of Engineering
Department of Applied Mechanics
Marine Technology

Aalto University publication series
SCIENCE + **TECHNOLOGY** 11/2012

ISBN 978-952-60-4599-3 (pdf)
ISSN-L 1799-4896
ISSN 1799-490X (pdf)

Unigrafia Oy
Helsinki 2012

Finland

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Marina Ladan, Maria Hänninen

Name of the publication
Data Sources for Quantitative Marine Traffic Accident Modeling

Publisher School of Engineering

Unit Department of Applied Mechanics

Series Aalto University publication series SCIENCE + TECHNOLOGY 11/2012

Field of research Marine Technology

Abstract

Utilization of data in quantitative accident modeling is the main concern of this report. Various data sources exist in the maritime field on a global level, but the primary interest in this report are the data sources that cover the Gulf of Finland. Other databases are included for comparison purposes or when Finland does not maintain a similar database. Special attention is given to collision and grounding accidents, and to data useful in analyzing human and organizational factors. The analyzed data sources are divided into three categories: general ship traffic data, accident data, and incident data. The sources are analyzed considering following:

- What type of data is collected and stored;
- What is the quantity and the quality of the data;
- Is data available to researchers and/or public;
- Can data be utilized in quantitative accident modeling?

It is found that the data sources differ in the scope and purpose and they all have their strengths and weaknesses. The existing sources are not perfect and using any of them as the only source of input to a quantitative model seems risky. This was also acknowledged by the participants of the workshops held at IMISS conference, who agreed that marine traffic accident and incident data collection and storing has to be improved in areas such as eliminating underreporting, differences in database taxonomies, and missing and erroneous data. As the improvement of data collection systems is a long term process, an alternative approach might be to improve the models for example by combining multiple sources of data and utilizing additional prior information.

Keywords marine traffic safety, accident models, accident data, operational safety

ISBN (printed) **ISBN (pdf)** 978-952-60-4599-3

ISSN-L 1799-4896 **ISSN (printed)** 1799-4896 **ISSN (pdf)** 1799-490X

Location of publisher Espoo **Location of printing** Helsinki **Year** 2012

Pages 68

CAFE PROJECT SPONSORS

European Union
European Regional Development Fund

Leverage from
the EU
2007–2013

KOTKA
CITY BY THE SEA

Aker Arctic

PORT OF
HAMINAKOTKA

PORT OF
HELSINKI

MERIAURA

KRISTINA CRUISES

CAFE PROJECT PARTNERS

MERI
KOTKA

Aalto University

CENTRE FOR MARITIME STUDIES

Kymenlaakson
ammattikorkeakoulu
University of Applied Sciences

TURUN AMMATTIKORKEAKOULU
TURKU UNIVERSITY OF APPLIED SCIENCES

Acronyms

AIS	Automatic Identification System
ATA	Actual time of arrival
ATD	Actual time of departure
CAFE	Competitive Advantage by Safety project
CASMET	Casualty analysis methodology for maritime operations
CSR	Corporate social responsibility
DAMA	Ship accident database formerly used in Finland
DGPS	Digital Global Positioning System
DNV	Det Norske Veritas
DP	Designated person
DWT	Deadweight tonnage
EEZ	Exclusive economic zone
EM	Expectation–maximization (algorithm)
EMCIP	European Marine Casualty Information Platform
EMSA	European Maritime Safety Agency
ENC	Electronic navigational chart
ERDF	European Regional Development Fund
ESP	enhanced survey programme
ETA	Estimated time of arrival
ETD	Estimated time of departure
EU	European Union
FMI	Finnish Meteorological Institute
GIS	Geographical Information System
GOFREP	The Gulf of Finland Mandatory Ship Reporting System
GPS	Global Positioning System
GT	Gross Tonnage
H&M	Hull & Machinery insurance

HELCOM	Baltic Marine Environment Protection Commission (Helsinki Commission)
ILO	International Labour Organization
IMDatE	Intergrated Maritime Data Environment
IMISS	International Maritime Incident and Near Miss Reporting Conference
IMO	International Maritime Organization
IRIS	Incident Report Information System
FAL	Convention on Facilitation of International Maritime Traffic
IRIS	Incident Report Information System
ISM	The International Safety Management Code
KKJ	Finnish national coordinate system, 1970 - 2003/2005
LMIU	Lloyd’s Marine Intelligence Unit
LNG	Liquid natural gas
LR	Lloyds Register
LRFP	Lloyd’s Register Fairplay
LTI	Lost time incident
MAIB	Marine Accident Investigation Branch
MMSI	Maritime Mobile Service Identity
MS	(EU) Member State
P&I	Protection & Indemnity insurance
Paris MoU	Paris Memorandum of Understanding
PCS	Port Community System
PEME	Pre Engagement Medical Examination
PSC	Port state control
SAR	Search and rescue
SMHI	Swedish Meteorological and Hydrological Institute
SMS	Safety management system
SOLAS	International Convention for the Safety of Life at Sea
Trafi	The Finnish Transport Safety Agency
USCG	The United States Coast Guard
VHF	Very high frequency
VTS	Vessel Traffic Service
WP	Work Package

1. Introduction

1.1 Marine traffic and accidents

Trade by sea has increased during last century [1]. Reasons lie, amongst other things, in technological advancements [1]. More ships are sailing in the world seas, the new ones being faster and bigger than the old ones. They are more sophisticated than before, and contain much more automation. At the same time, there is a shortage of good seamen; BIMCO-ISF 2010 Manpower Update informed about 13000 officers shortage in 2010 [2]. However beneficial to society development is, it still possesses many threats. Accidents at sea still occur, and consequences to people, ship or environment, are often greater than before.

There are various ways to classify marine traffic accidents. As an example, in EMCIP database (see Chapter 3.2.3), the accident types are capsizing/listing, collision, contact, damage to ship or environment, grounding/stranding, fire/explosion, flooding/foundering, hull failure, loss of control, missing and non-accidental event. The number of accidents occurred in the Gulf of Finland ship traffic in 1997-1999 and 2001-2006 are presented in Figure 1. The most common accident types in this sea area had been groundings (48 % of all accidents) and collisions of two ships (20 %) [3].

Also the consequences of maritime accidents can be expressed in many ways. One can examine the consequences to the vessel(s) or other structures involved in the accident, to the humans either directly or indirectly involved in the accident, or the ones to the surrounding nature and environment. Further, the magnitude of the consequences can be measured on various scales. They can be probabilities or expected values of a variable describing the consequences directly, such as the number of injured persons or the amount of oil leaked to the water, or they can be described with monetary terms, such as the cost of cleaning the oil spill or the cost of human life. Although assigning monetary values for the consequences might be challenging, it is often necessary. Different types of consequences must be comparable when a cost-benefit analysis of alternative actions or decisions aiming to improve the safety or to mitigate risk is conducted.

1.2 The purpose of accident modeling

The ultimate goal of accident modeling is to learn more about the accidents in order to prevent them in the future. Especially if there is a lot of data on accidents available, it might be easier to draw conclusions based on the model than based on the raw data itself. Preventing an accident could mean minimizing the likelihood of the unwanted event, the magnitude or severity of the accident consequences, or it could include both of these elements. However, dividing the accident models into models that estimate the likelihood element and to models which address the consequences is a bit misleading, as almost any feature in the accident chain from the design or decision-making level to the final consequences can be seen as either kind, if these starting and ending points can even be defined. For example, the size of a hole in the hull after a collision can be estimated probabilistically and it has its own consequences, such as an oil leak. Further, an oil leak can be seen as the unwanted event with certain harmful consequences to nature, for example. Likewise, the unwanted event collision could be seen as a consequence of a certain organizational deficiency. Nevertheless, accident models provide support to decision making and to cost-benefit analyses when managing risks and safety.

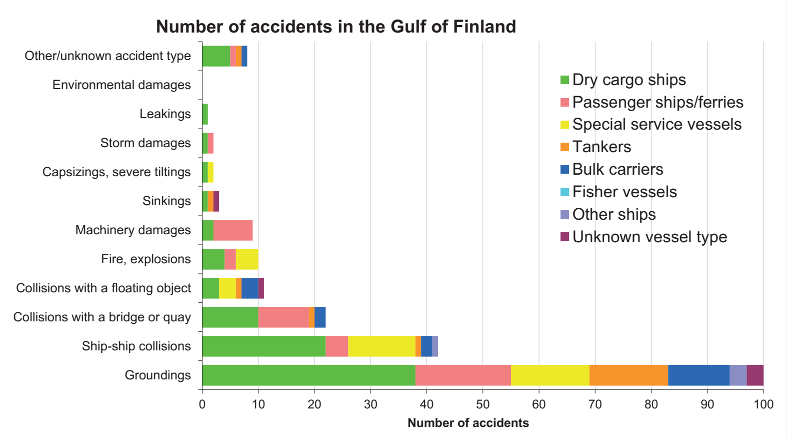

Figure 1. Marine traffic accident in the Gulf of Finland in 1997-1999 and 2001-2006 [3]

Quantitative accident models, which describe a system's quantities mathematically, statistically or computationally, provide a means to cost-benefit analysis within risk analysis or safety-related decision making. However, as accidents are low probability events, there often is a lack of data necessary for the models. To address this shortcoming, there is a growing interest in near-miss data. Reasons for this are that near-misses occur more frequently, but might be partly governed by the same underlying causes and mechanisms as accidents. Hence, a possibility to utilize near-miss data in accidents modeling appears tempting.

1.3 CAFE project

This report is a part of the research project Competitive Advantage by Safety (CAFE). CAFE aims to improve the safety, efficiency and competitiveness of maritime transportation. A special focus is on human and organizational factors and operational safety. The CAFE project tries to strengthen the knowledge obtained and shared during the "Development of Maritime Safety Culture (METKU)" project of the same consortium, where quality and environmental standards were compared with the International Safety Management Code (ISM Code) and shortcomings in the safety management of shipping companies were identified [4]. The CAFE project focuses on these shortcomings but also increases specialization to improve the sector's competitiveness. CAFE will look at how:

1. international experiences from near miss reporting systems could be used efficiently
2. maritime safety statistics could be used as an objective safety indicator
3. improving occupational safety could enhance the sector's competitiveness
4. modeling of safety management could make it possible to focus on a few key proprieties and how this focusing could lower the risk for maritime accidents
5. operational safety and human factors have high potential to improve safety
6. "weak signals", i.e. Corporate Social Responsibility, could be used improving the sector's public image
7. international networking could strengthen maritime safety on the Baltic Sea and in Europe.

The consortium members in the CAFE project are Kotka Maritime Research Centre, Centre for Maritime Studies at the University of Turku, Aalto

University School of Engineering, Turku University of Applied Sciences and Kymenlaakso University of Applied Sciences. CAFE is partly funded by the European Union; the financing comes from the European Regional Development Fund, the ERDF programme for Southern Finland, Priority Axis 5 (where the Regional Council for Päijät-Häme coordinates the ERDF programme and manages interregional thematic projects), the City of Kotka, Finnish Shipowners' Association and the Kotka Maritime Research Centre corporate group: Aker Arctic Technology Inc., the Port of HaminaKotka, the Port of Helsinki, Kristina Cruises Ltd, Meriaura Ltd. and the project partners. The cost estimate of the project is approximately 1.5 Million Euro and its duration from 1 October 2010 to 31 August 2013.

The CAFE project consists of several Work Packages (WPs). A work package is a subset of a project that can be assigned to a specific party for execution) that are further divided into subtasks. In CAFE, the WPs are shared with multiple consortium members, and each subtask is led by a member. The work packages, subtasks and the responsible subtask leaders are:

- WP1: Reporting of incidents and near misses in the shipping industry
  - Task A: Baltic Sea Safety Discussion Forum (Kymenlaakso University of Applied Sciences)
  - Task B: International reporting of incidents and near misses (University of Turku, Centre for Maritime Studies)
  - Task C: Applicability of statistical data for risk modeling (Aalto University School of Engineering)
  - Task D: Benchmarking occupational safety indicators (Turku University of Applied Sciences)
- WP2: Modeling safety management to increase competitive advantage in shipping
  - Task A: A conceptual model about safety management (University of Turku Centre for Maritime Studies)
  - Task B: Bayesian network model for estimating the effects of safety management on maritime accident risks (Aalto University School of Engineering)
- WP3: Corporate social responsibility in international shipping
  - Task A: Theory of corporate social responsibility (CSR) and implications to the maritime sector (University of Turku Centre for Maritime Studies)
  - Task B: Theory of corporate social responsibility (CSR) and implications to the maritime sector (Kymenlaakso University of Applied Sciences)
- WP4: International networking (all)

- WP5: Coordination and management (Kotka Maritime Research Centre)

This report belongs to the Task C of WP 1, “Applicability of statistical data for risk modeling”.

1.1 Scope of the report

The report reviews various existing information sources from the view of utilizing them as input data for quantitative marine traffic accident models. The report is based on examining the data sources, reviewing relevant literature and the results from an international expert workshop organized as a part of the CAFE project.

As the report is written within the CAFE project, a special focus is paid to utilizing data sources when modeling human, organizational and operational factors within marine traffic. However, as accidents typically involve a lot of variables with complicated dependencies, the distinction between human and technical failure is not always clear, or it should not be. Thus the report also considers utilizing data for marine traffic accident models not focusing purely on operational safety.

The report does not cover data that does not exist at the moment but could be gathered if seen feasible, such as measurements from laboratory experiments to be performed in the future. Other aspects of the information sources such as their primary purpose, the regulations concerning them and applications to other purposes than accident modeling, are not covered or they are only mentioned very briefly.

In this report, the emphasis is on data describing the marine traffic in Finland, especially in the Gulf of Finland. The reader should be aware that there are many similar data sources describing ship traffic and its safety elsewhere. However, describing all of them would be practically impossible and thus only a few additional sources are included either for comparison purposes or in cases when Finland does not maintain an appropriate database for the specific problem. The focus is on collision and grounding models, as these events have been the most common in the Gulf of Finland (Figure 1). Nevertheless, the data sources presented in this report could be utilized for other types of marine traffic accidents as well. Although the utilization of accident and incident data in modeling within other domains may have similarities with marine traffic applications, they are not discussed in this report.

The rest of the report is organized as follows. In order to get an overview to which purposes the data might be needed, some references to existing accident models and their analyses are given in Chapter 2. Chapter 3 describes the feasibility of data sources describing the ship traffic, marine traffic accidents and near misses. In Chapter4, possible ways to improve the utilization of these data sources within marine traffic accident models are discussed. Also, a brief summary from the IMISS conference workshops is provided. Finally, conclusions are drawn in Chapter 5.

2. Marine traffic accident models

2.1 General accident models

Before searching for suitable data sources, one must know what kind of model(s) the data is needed for. The models of course depend on the nature of accidents one wants to analyze. This chapter briefly introduces marine traffic accident models, especially collision and grounding models. In order to keep this report compact, the existing accident models are not presented here in detail. Instead, for a reader interested in the models, this chapter gives references to the original papers or other publications summarizing the models.

A theoretical description of accident causation can be described using an accident model. The purpose of an accident model is to serve as a compact, qualitative and/or quantitative representation of the mechanisms behind the accident. Over the years, there have been many accident models published which are either describing the accidents on a general level or within a certain domain. These models are not presented here and the reader is advised to find summaries and references to various models e.g. in [5], [6] and [7]. Probabilistic collision and grounding accident models are referred to in the following.

2.2 Collision and grounding probability and consequence models

The traditional approach to estimating the probability of collisions or groundings models the number of accidents as a product of the so-called number of geometrical accident candidates and a causation probability (see e.g. [8]). The number of geometrical collision or grounding candidates describes the theoretical number of collisions or groundings given the ship traffic properties of the area, such as the historical ship tracks, speeds and sizes, and the assumption that the ships are not performing any evasive maneuvers. For collisions, it is estimating the number of times two ships are on a collision course within the time period under examination. Similarly, for grounding it describes the number of dangerous courses towards shoals. The causation probability then describes the probability that the ships do not make any evasive maneuvers given that they are on a collision or grounding course. It is thus conditional on the approach used for geometrical collision candidate estimation. The causation probability is affected by various variables that are related to the mariners and the organization, the conditions on board and outside the ship, and to technical reliability. The existing models for accident probability estimation are not described in this report, as various techniques and models for estimating the geometrical probability of collisions or groundings have already been presented and discussed in e.g. [9], [10], [11] and [12] and for the causation probability e.g. by the author of this report in [6] and [13].

For estimating the damage resulting from impact accidents, there are many models with either analytical or numerical approach to modeling the problem. Some of these are discussed and compared in [14], [15] and [16]. Further, there are models of “the consequences of the consequences”. For example, the impacts of an oil spill have been modeled (e.g. [17] and [18]).

3. Data sources for marine traffic accident models

3.1 Ship traffic data

3.1.1 Traffic image data in marine traffic risk modeling

If the marine traffic is to be modeled, some kind of information on the ship traffic in the analyzed area is necessary. According to Larsen [19], information on cargo type and amount, ship velocity and navigational aids such as using a pilot are especially relevant to risk assessment purposes, as these variables affect the draft, collision energy, ship maneuverability, risk of pollution, and safety. Also the environment where the ships are navigating is important. Thus one might need information on ship routes, traffic intensity, and the properties of the water area. Some of these variables are static and some dynamic; whether the variable changes over time depends on the risk modeling purpose: whether risk assessment is performed on a total ship traffic certain location or on a certain ship during its voyage, for example.

Especially in the past, there might have been no information on the actual traffic image available. The traffic image could have been constructed based on various ship traffic registrations such as port registrations, pilot registrations, and coast guard or navy registrations. An example of a port registration system is described in subchapter3.1.3. Additionally, data could have been collected using visual or radar observations or VHF-radio communications. Currently, Automatic Identification System (AIS) data can be utilized in modeling the traffic. Subchapter 3.1.2 discusses the AIS data and its feasibility.

Table 1. Information included in the AIS message and the update rates [20], [21]

Type of information	Information	Update rate
Static	IMO and MMSI number Call sign and name Ship type Length and beam Location of position fixing antenna such as GPS/DGPS	Every 6 minutes and on request
Dynamic	Position and accuracy indication Time Course over ground Speed over ground Heading Navigational status Rate of turn Angle of heel (optional) Pitch and roll (optional)	2-180 seconds depending on speed and course alteration
Voyage-related	Draught Hazardous cargo type Destination and estimated time of arrival Route plan waypoints (optional) Number of persons onboard (on request)	Every 6 minutes, when data has been amended and on request
Short safety messaging	Short text messages with important navigational safety related information	As required

3.1.2 Automatic Identification System data

Automatic Identification System (AIS) operates on VHF frequency on and automatically sends and receives information between ships. The information is also received on shore by the coastal authorities such as the Vessel Traffic Service (VTS). The information transmitted via AIS can be seen in Table 1. The dynamic information is being updated automatically and the update rate depends on the speed and course alteration. Voyage-related information is entered manually. AIS is mandatory for all ships of 300 GT and upwards and to all passenger ships [22].

Although mainly used for safe navigation and VTS surveillance, historical AIS data can be used for modeling ship traffic for risk assessments. The trajectories and routes of the AIS-carrying ships that navigated within a certain water area within a certain time can be constructed, as e.g. in [23], [13] and [24]. The trajectories and routes can be further used in accident frequency estimation. AIS data has been utilized in many calculations of collision frequencies, especially when determining the so-called number of collision candidates. It has been used as input data for analytical collision candidate models, e.g. in [11] ,[13],[25] and [26]. Goerlandt & Kujala [12] used the MMSI number, time, position, speed, ship type, ship length and ship beam for a dynamic traffic simulation and collision candidate estimation for the Gulf of Finland. AIS data was also utilized in a traffic simulation –based collision risk model applied to Northern American locations [27].

In addition to reconstructing traffic patterns of detecting collision or grounding candidates, AIS data can be used for identifying vessels that have been navigating in the area. Then other data sources can be utilized in seeking information on the properties of these identified vessels, e.g. their flag state or history of detentions (see Chapter 3.3.4). The voyage-related information on hazardous cargo can be used in estimating the consequences of a possible accident.

Unfortunately, AIS data has had many errors and its reliability can be questioned. Errors or missing data have been found in both the information inserted manually and in the dynamic variables e.g. in [20] , [28] and [29]. The number of AIS data points collected from a sea area depends on the traffic amounts, but even for a rather short time period, it typically becomes very high. Although the large number of data points can compensate the missing or erroneous data [30], it also poses some practical challenges to storing and using the data. Nevertheless, AIS data still provides a very valuable source for modeling marine traffic and ship routes.

3.1.3 Port information systems

Over 30 ports worldwide are using either port-specific or national information systems or the so-called port community systems (PCSs), all of which are used for sharing and transferring information between different port-related actors. In Finland, the national, Internet-based information system PortNet is being used (for more information on the other systems, see e.g. [31]). The purpose of PortNet system is to provide a means to fulfill the requirements of several actors with a single notification from the ship; ship agents insert information on the ship calls into the system, which is then being used by the customs, ports, port-related companies and maritime authorities [31]. PortNet contains the following information [32]:

Basic information

- Ship register
- Agents and ship owners
- Fairway dues tables
- Ports, port areas and berths
- Customs offices

Port call information

- List of port calls
- Port service requests
- Port visit schedules (ETA, ATA, ETD, ATD)
- VTS / AIS times (ETP, ATP)

Port call notifications

- Customs ship notification
- Dangerous cargo
- Cargo info, manifest
- Cargo info, statistics
- Waste notifications
- IMO-FAL notifications
- Security notification

Ship's dues

- Annual fairway due
- Tonnage due
- Foreign traffic, single payment
- Increase or surcharge

Port enquiries

- Timetables
- Cargo information
- Dangerous cargo reports
- Waste notifications
- Port service requests
- Dangerous cargo entry permits

Reports

- Timetables
- Cargo info, statistics and manifest
- Dangerous cargo

In Finland, PortNet has been in use for almost 10 years, and there will be no changes or improvements made to the system in the near future [31]. In addition to PortNet, there are separate own information systems for the needs of various port-related parties such as port operators, customs and railway traffic. These are summarized in [31]. However, the information systems are not integrated and a PCS serving a comprehensive information exchange within the whole port community is not existing in Finland.

For marine traffic accident probability modeling, a PCS might not provide any added value. Nevertheless, based on the PortNet system data, information on cargoes within the Finnish waters can be derived. This is useful for estimating the consequences of maritime accidents, for example. Also, PortNet arrival and departure times can be used in checking and complementing AIS data. Thus a database that would combine PortNet system data with AIS data and the Mandatory Ship Reporting System GOFREP and VTS arrival reports would form a comprehensive database describing marine traffic in Finland.

3.1.4 SafeSeaNet

SafeSeaNet is a centralized European system for maritime data exchange [33]. European Maritime Safety Agency (EMSA) is responsible for the SafeSeaNet system and it involves authorities from EU member states plus Iceland and Norway. Implemented in 2004, SafeSeaNet contains information on ship movements, the nature of hazardous cargo on board and on the ships considered as risky to the shipping and environment safety [34]. The information comes primarily from AIS messages and notifications from the member states. The authorities can use the information for ship traffic monitoring and management, search-and-rescue operations and risk management. In the future, the system will also be available for other users, such as customs and port state control inspections [35].

According to EMSA's progress report from 2010 [34], SafeSeaNet has been progressing significantly, but the quality and quantity of some of its information still needs improvement. For the future, EMSA is developing Intergrated Maritime Data Environment (IMDatE), a framework that will combine SafeSeaNet with other marine traffic databases of EMSA, such as THETIS (see Subchapter 3.3.4) [36]. The purpose of IMDatE is to provide a complete, almost real time picture of marine traffic on an European level.

3.1.5 Crews, organizations, procedures

Modeling human and organizational factors could benefit from data which would describe the properties, actions and procedures of the crews and the shipping company organizations. Unfortunately, information on these factors is not very well available to a risk analyst not working within a shipping company – the data would have to be gathered from the companies, if they are willing to share it. However, an ongoing research [37] will study the manning and crews in the Baltic Sea marine traffic. This information could be utilized when developing risk modeling for the Gulf of Finland, for example.

In Finland, The Finnish Transport Safety Agency Trafi keeps a seafarer register which contains information on the tasks and work amounts of the crews working onboard Finnish vessels and of a limited number of Finnish seafarers working on foreign ships [38]. The ship-owners are obligated by law to deliver this information to Trafi. Based on the register's information, annual seafarer statistics are published. The statistics present the following information:

- seafarer man-years, the numbers of personnel onboard and the numbers of people working in seafarer professions in total and in different occupational groups
- the proportions of women seafarers in general and in different occupation groups, domestic and international traffic
- the proportions of foreign seafarers in general and in different occupation groups
- the proportions of age groups in general and in different occupation groups

Thus, the mariner age and sex distributions within the Finnish fleet could be constructed from the seafarer statistics. Unfortunately, the statistics do not present the nationalities of foreign seafarers, so only the proportion of Finns is available for utilization.

3.1.6 Environment data

Data on the environment and on location the ships are navigating and whose accidents are to be modeled can also be considered as ship traffic data. Geographic features of the water area, its currents, the weather and visibility might be included in the accident probability models. Also, the environment affects the consequences, such as the need and/or succession of evacuation, and the severity and the costs of a possible oil spill.

The Finnish Transport Agency's Hydrographic Office is responsible for paper and electronic charts of the coastal and lake areas of Finland. "The Hydrographic Office maintains a hydrographic data information service and publishes other material in connection with the charts" [39]. In 2003, the national coordinate system KKJ is replaced by WGS 84 (EUREF_FIN) standard adopted by the International Hydrographic Organisation as an international standard. Sea area charts are published as coastal charts, harbor charts, chart folios and small scale charts. Electronic navigational charts produced by Finnish Transport Agency in S-57 vector format are distributed by Primar in Norway. ENC data is encrypted in accordance with the standard IHO S-63. All charts are commercially available. Charts are updated from time to time and mariners are obliged to use the valid version.

Finnish Transport Agency also issues Notice to Mariners where different kinds of navigational notices are published, Finnish List of Lights where a complete list of lights and information about radio navigation, pilot, VTS and rescue services on the Finnish coast are available. Symbols, abbreviation and terms are published, too. As a complement to the nautical charts and publications, fairway cards are issued. They contain facts about the fairway dimensions, navigability and navigational conditions, traffic recommendations and restrictions and traffic services provided. Finland has a total of around 19,500 kilometers of public, charted fairways marked by more than 33,000 maritime aids of navigation (lighthouses, buoys, signs, leading beacons, etc.). The Finnish Transport Agency is responsible for around 25,000 of these.

In general, statistics on weather conditions have been gathered for a very long time. For example, Finnish Meteorological Institute (FMI) provides statistics for storm and wind days [40] and for waves and sea levels [41]. In Finnish waters, ice is also an environmental factor to be taken into account. For example, during a normal winter, the heavily trafficked Gulf of Finland freezes to a large extent, or even completely, and the ships need icebreaker assistance. The majority of ship collisions in the Gulf of Finland have occurred during the dark and icy winter months [3]. Safe navigation in dynamic ice conditions poses many challenges to navigators, especially for crews not familiar with winter navigation, few examples being damages to hull or propulsion, grounding or collision due to avoiding ice and loss of stability due icing [42]. During winter, ice charts describing the ice conditions in the Baltic Sea are published daily by FMI and Swedish Meteorological and Hydrological Institute (SMHI) [43]. The ice thickness, ice type, and water temperature information on the charts come from satellite image observations, observations onboard ships and from measurements. However, ice charts are too coarse for describing local ice conditions [44], which might be needed for a location-specific marine traffic risk models. For the Baltic Sea, a chart of ice probabilities has been published [45, cited in 44]. However, it also is not detailed enough for traffic risk assessment purposes.

3.2 Accident data

3.2.1 DAMA database and Finnish accident statistics

In Finland, all maritime accidents except the ones of pleasure boats must be reported according to the Maritime Act. Accident database DAMA consists of maritime casualty reports given to Finnish maritime safety authorities. The reported accidents of Finnish vessels and accidents to foreign vessels in Finnish territorial waters were stored in DAMA from the year 1990 to 2010 [46], [47]. In 2001-2005, the average number of accident cases stored per year was 50. Very minor accidents such as small dents from ice assistance situations, pleasure boat accidents or accidents that have not been reported were not included in DAMA-database. Recently, the DAMA database was judged as obsolete [47], and after 2010, Trafi began using EMCIP database (see Chapter 3.2.3) for storing the accident data.

The categories for accident types in DAMA were:

- Ship-ship collision
- Collision with an offshore platform
- Collision with a floating object
- Collision with a bridge or quay
- Grounding or stranding
- Capsizing
- Severe tilting
- Leakage
- Environmental damage
- Storm damage
- Machinery damage
- Fire/explosion in machinery area
- Fire/explosion in cargo area
- Fire/explosion in other areas
- Fire of electrical equipment
- Injury, death, poisoning
- Helicopter accident
- Missing ship
- Near accident
- Unknown

Besides the accident type, DAMA entries include the fields listed in Appendix A. However, not all fields have been filled in all accident cases. For most of the fields, the information was filled by number codes based on fixed categories. The causes in DAMA have been categorized under seven cause types:

- External factors
- Ship structure and layout
- Technical faults in ship equipment
- Factors related to equipment usage and placement
- Cargo, cargo and fuel handling and related safety equipment
- Communication, organizing, instructions and routines
- Persons, situation assessment, actions

The complete list of cause categories is presented in Appendix B.

Based on DAMA database, accident statistics analyses have been conducted. The report covering the years 1990-2000 [48] concentrated only on grounding and collision incidents. The analysis for the years 2001-2005 [46] included all the accident types. Analyses focused on accidents in Finnish territorial and inland waters.

These studies presented statistics of accident characteristics such as ship types, circumstances and causes. More in-depth analyses of the maritime accidents in Finland, such as an analysis of the correlations between the different factors, or studies for finding subgroups or clusters within the accidents, cannot be found.

The advantage of accident databases such as DAMA and the HELCOM database presented in the following subchapter is that the provided information is categorical or numerical which enables statistical analyses. However, it is also a drawback, as the categorization has been fixed and thus it may create uncertainty. When considering models describing accident causation mechanisms, DAMA provides very little information. Although it is better to be able to store four causes of the accident than only one, the database is lacking the description of the chains of events leading to the accident.

3.2.2 Baltic Sea accident statistics

Baltic Marine Environment Protection Commission HELCOM (Helsinki Commission) has gathered yearly statistics of Baltic Sea accidents [49]. Accident registrations come from the coastal states of the Baltic Sea: Finland, Sweden, Denmark, Estonia, Germany, Latvia, Lithuania, Poland and Russia. All accidents of tankers over 150 GT and/or other ships over 400 GT in states' territorial waters or EEZs are reported. The accident reporting format changed in 2004 and thus the data before 2004 and the subsequent years are not fully comparable. In 2005-2009, the average annual number of accidents in HELCOM database is 125. The accident dataset, from 1989 on, can be accessed online with a map based web tool, HELCOM Map and data service [50].

The accidents in the database are divided into collisions, fire, groundings, machinery damages, physical damages, pollutions, sinkings, technical failures and other accidents. Collisions can be further classified as collisions with another vessel, with an object, or as the ones with another vessel and an object. The fields of the data entries can be seen in Table 2. Not all fields are filled for every accident case – the numbers of times the field has been filled and the following reporting percentages of the fields are also presented in Table 2. The causes of accidents are not as specifically categorized as in DAMA. The cause categories in the database are "human factor", "technical factor", "external factor" and "other factor". There is a text field for describing the cause more specifically. However, as can be seen from Table 2, it has been filled in only 21.9% of the cases.

HELCOM publishes annual accident statistics that present the number of accidents in the Baltic Sea, the spatial distributions, accident type distributions, types of vessels involved in accidents and the distributions of accident causes. The number of groundings and ship-ship collisions are presented separately for the south-western Baltic Sea, the Gulf of Finland and the whole Baltic Sea. Also the number of accidents with pollution, types of accidents and vessels involved in them and the causes of accidents resulting in pollution are presented. The accident statistics reports are available online [51].

In addition to the statistics published by HELCOM, a combination of DAMA data and HELCOM data from the years 1997-1999 and 2001-2006 has been used in evaluating accident statistics for the Gulf of Finland [3] , [52]. Salmi [53] used HELCOM accident database in comparing vessels involved in accidents and the ones recognized as accident prone ships based on VTS violation and incident reports (see Chapter 3.3.3).

HELCOM data does not contain as many accidents from Finnish waters as the DAMA does. As an example, in DAMA there are 46 accidents from Finnish waters in 2004, whereas in HELCOM database the number is 8. On the other hand, some of the accidents present in the HELCOM data are missing from DAMA. Although not complete and even containing some errors [53], HELCOM data is the largest database with uniform data format of the Baltic Sea accidents at the moment.

3.2.3 EMCIP database

The European Marine Casualty Information Platform (EMCIP) is a confidential database established on EN Directive 2009/18/EC and operated by EMSA [54]. The main purpose of the Directive is to improve maritime safety within the EU community, and the view is that the goal can be achieved better by the effect of scale. Therefore, since June 2011, all Member States (MS) are obligatory to notify EMCIP about any maritime casualty/accident occurrence and provide a report for very serious and serious accidents which they investigate. Common training for MS personnel was provided by EMSA to accomplish application of the same principles in the investigations of casualties and data analyses across the EU. EMSA also "monitors the quality of and accepts the field reports" [54].

EMCIP access is granted only to authorities entitled by MS. It is planned that this will include research institutes, but not businesses. Information about casualties involving merchant ships, recreational craft and inland waterway vessels are stored. Information about occupational accidents is also kept. All casualty events are classified according to an agreed taxonomy to the following event types [55]:

- Capsizing
- Listing
- Collision with other ship
- Collision with multiple ships
- Collision, ship not underway
- Contact with floating cargo
- Contact with ice
- Contact with other floating object
- Contact with unknown floating object
- Contact with fixed object
- Contact with flying object
- Damage to ship or equipment
- Drift grounding/stranding
- Power grounding/stranding
- Fire
- Explosion
- Foundering
- Progressive flooding
- Massive flooding
- Loss of electrical power
- Loss of propulsion power
- Loss of directional control

- Loss of containment
- Hull failure
- Missing

Collected data is divided into factual data and casualty analysis data . To describe the sequence of the events related to a casualty, results obtained in the *Casualty analysis methodology for maritime operations* (CASMET) project [56] are used. It is stated that not all accidental events necessary lead to casualty. Representation of the EMCIP approach, i.e., the casual connection between events and factors, is shown in
Figure **2**.

The database had operated only on a voluntary basis for two years until June 2011 when it became obligatory. Therefore, the number of cases it supposedly contains might be insignificant. Once the MSs transfer previous experiences collected in individual databases, EMCIP will grow.

It is still early to evaluate the quality of data in EMCIP since the database has been operated only for two years and on the voluntary basis only. We also need to wait for outcomes from the effect of scale. The results are available only to EMSA as a particular MS has access only to her own data, and not to the data of other MSs. Nevertheless, EMCIP manages to establish a common taxonomy, which can facilitate different comparison studies. We can discuss the "fact that virtually no taxonomy can represent the full spectrum of possible causes" [57], but from the research point of view having a common taxonomy is quite large improvement. Time will show if this theoretical improvement will be followed by better reports.

Some countries such as Sweden are still using, and in the near future will continue to use, their own parallel accident databases. In Finland, EMCIP has replaced DAMA. When an accident occurs in Finnish waters or to a Finnish ship, a report has to be filled, signed and sent to Trafi. The report is in a paper format and has 15 pages. First three contain a general part and are obligatory. The rest of the pages cover different categories of accidents and need to be fulfilled depending on the event occurred. There are eight categories and an additional called "other accidents/incidents". Trafi gets on average 30 reports per year, but not all of them are investigated. Finnish Accident Investigation Board has access to data stored in EMCIP and based on those data decide which accident needs to be further investigated. As already mentioned, research institutes will have access to the database, but this is still not the case.

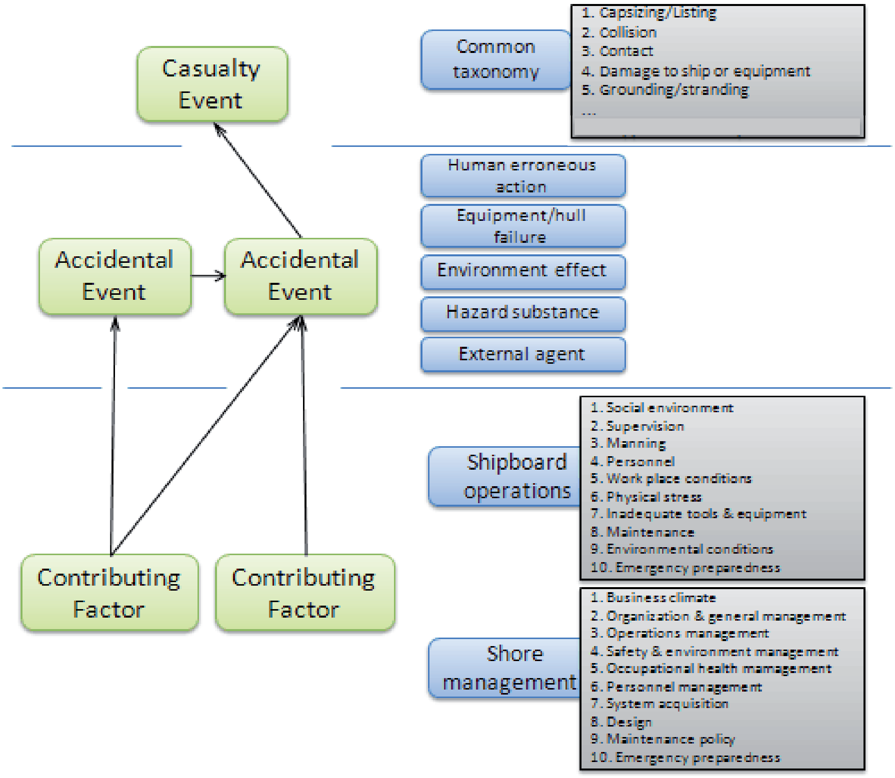

Figure 2. EMCIP approach

In a report of the marine traffic accident statistics for the year 2010 [58], merchant fleet and pleasure boat accidents within 2007-2010 were examined. Trafi provided data for 2007-2008 from the DAMA database and for the years 2009-2010 from EMCIP. Other data sources for the report were marine traffic accidents from the Finnish police forces database, search and rescue (SAR) task database of the Finnish Border Guard, and the SAR tasks of voluntary maritime rescue associations in Finland provided by the Finnish Lifeboat Institution. The accidents with casualties were checked and completed by Statistics Finland. The publication presented the numbers of marine traffic accidents and the numbers of casualties for various factors such as water areas, age, sex, month, weekday and time. Also, distributions of ship types (four merchant vessel and nine pleasure boat types) and causes (nine cause types) in the accidents were reported.

Table 2. Data fields in HELCOM accident database. The number of times reported describes the number of cases where the corresponding field has not been left blank or reported as “n.i.”, “unknown” etc. Percentages marked with * are calculated from the number of collisions with another vessel, the one with ** from the number of collisions, and the one with *** from the number of accidents with pollution. Ship2 size (dwt) values were found to be identical to the reported Ship1 size (dwt) in all but one collision with another vessel, so its correctness can be questioned and the reporting percentage is not presented in the table.

Data field	Entry format	# of times reported	Reporting %
Date	dd.mm.yyyy	1251	100,0 %
Ship1 name	text	1251	100,0 %
Ship2 name	text	145	100,0 % *
Year	numeric	1251	100,0 %
Latitude	numeric	1250	99,9 %
Longitude	numeric	1250	99,9 %
Accident type	categorical (see text)	1249	99,8 %
Ship1 category	cargo/passenger/tanker/other/n.i.	1230	98,3 %
Pollution	no/yes/n.i.	1166	93,2 %
Type of pollution	text	133	93,0 % ***
Amount of pollution (m^3)	numeric	1021	81,6 %
Collision type	object/vessel/vessel&object/n.i.	273	78,0 % **
Ship1 type	text	964	77,1 %
Ship2 category	cargo/passenger/tanker/other/n.i.	108	75,0 % *
Ship2 type	text	87	65,9 % *
Country	text	756	60,4 %
Ship1 size (gt)	numeric	725	58,0 %
Time	hh.mm	646	51,6 %
Ship2 size (gt)	numeric	68	51,5 % *
Cause, ship1	human/technical/external/other/n.i.	616	49,2 %
Ship1 draught (m)	numeric or interval	590	47,2 %
Pilot, ship1	no/yes/exemption certificate/n.i.	572	45,7 %
Cargo type	text	535	42,8 %
Ice conditions	no/yes/n.i.	507	40,5 %
Damage	text	478	38,2 %
Cause, ship2	human/technical/external/other/n.i.	46	34,8 %
Accident details	text	423	33,8 %
Ship1 size (dwt)	numeric	395	31,6 %
Offence	text	277	22,1 %
Cause details	text	274	21,9 %
Assistance need	text	209	16,7 %
Ship1 hull	single/double/n.i.	170	13,6 %
Pilot, ship2	no/yes/exemption certificate/n.i.	15	11,4 % *
Ship2 hull	single/double/n.i.	13	9,8 % *
Ship2 draught (m)	numeric or interval	55	4,4 % *
Additional info	text	38	3,0 %
Consequences/Response actions	text	36	2,9 %
Amount of pollution (tons)	numeric	15	1,2 %
Crew trained in ice navigation	no/yes/n.i.	14	1,1 %
Ship2 size (dwt)	*numeric*	*157*	-

3.2.4 Accident investigations

IMO has established a Casualty Investigation Code, which "will require a maritime safety investigation to be conducted into every 'very serious maritime casualty', defined as a maritime casualty involving the total loss of the ship or a death or severe damage to the environment. The Code will also recommend an investigation into other maritime casualties and incidents, by the flag State of a ship involved, if it is considered likely that it would provide information that could be used to prevent future accidents." [59].

In Finland, the Accident Investigation Board, located within the Ministry of Justice, investigates and reports "all major accidents regardless of their nature as well as all aviation, marine and rail accidents and their incidents" [60]. Marine accidents are investigated if they have occurred within Finnish waters, or if a Finnish vessel has been involved in the accident. The Board investigates and reports how the accident occurred, what were the circumstances, the causes, the consequences and the rescue operations. The reports based on the investigations also provide recommendations of actions for preventing similar accidents. The purpose of the marine accident investigation process is to prevent the accidents – not to blame or judge anyone. The marine traffic accident investigation reports of accidents from 1997 on and 10 older reports are available at Accident Investigation Board's web pages [61]. On 20 May 2011, 162 reports of accidents, serious incidents, incidents, damages, minor accidents and other incidents could be downloaded from the page.

For comparison purposes, Marine Accident Investigation Branch (MAIB) is examined as one of the most representative investigation bodies worldwide. The aim of MAIB is to investigate marine accidents/incidents occurred within UK territorial water or involving UK vessels worldwide. It is an independent unit within the Department for Transport which was formed in 1989. MAIB operates under a legislative framework provided in The Merchant Shipping (Accident Reporting and Investigation) Regulations 2005, amended in 2011 to bring into force EU Directive 2009/18/EC [62].

MAIB maintains a computerized database of reportable marine accidents which have occurred since 1991. The reports from 1973 to 1988 were produced by the Marine Division of the Department of Trade. The existing reporting scheme was changed in October 2001 when MAIB's new database became fully operational. An entirely new taxonomy or classification of marine accidents was devised by the MAIB inspectors. In the transition phase from the old system to the new, 21000 separate records of accidents reported to the Branch over a period of ten years were transferred electronically [63]. It can be assumed that this figure has doubled since that period. During this process, the data was cleaned in order to correct any inconsistencies. These changes have resulted in minor variations to historic data [63]. Due to the unavailability of the database, the aforementioned variations cannot be validated. However, information from the database is available through ‘Freedom of Information’ Act [64]. Service is provided on the Department for Transport’s web pages. Individuals have a right to request any recorded information held by a public authority [65]. Still, the information may not be given because it is exempt, for example if it reveals personal details about somebody else (different than the person who was asking for the information) [65]. It is unlikely that the whole database information can be asked, but an example of a request included ‘near miss and incidents related to LNG tankers’ (2011). A list of sixteen cases occurred in 1991 and after were given as an answer for the aforementioned request. In 2011, there were 17 requests for information connected with shipping. Not all of those were related to accident investigations. In 2009, one of MAIB’s objectives was to “develop a new database that will replace the existing Marine Incident Database System with a system that will reflect modern accident investigation processes and fulfill the UK’s data provision obligations to EMSA and IMO” [62]. In 2011, this is still uncompleted.

MAIB receives between 1500 and 2000 reports of accidents and incidents each year [62]. Not all of them are investigated and effort was made to select “which incidents are likely to yield the most important issues for future safety. It is only these accidents that MAIB investigates” [66]. In 2001, the maximum number of reports published as results from fully conducted investigations was 45. In 1999, a decision was made to make all reports publicly available. From June 2011 on, as to comply with the EU Directive, the MAIB (and all Member States) must conduct a full investigation into all ‘very serious accidents’, and give reasons for any ‘serious accident’ which they do not intend to fully investigate.

MAIB investigation reports are sorted by year (from 1990 to 2011, date of year published), vessel category (merchant vessel, fishing vessel, leisure craft), incident (machinery, fire/explosions, injury/fatality, grounding, collision/contact, flooding/foundering, listing/capsize, cargo handling failure, weather damage, hull defects, hazardous incidents) or alphabetically.

Shorter narrative information about accidents other than those investigated in the reports is also available online. These are given in Safety Digests, approximately 25 cases, twice per year.

Accident investigation reports were used in the study of performance shaping factors in navigation accidents in the Royal Norwegian Navy [67]. For evaluating the presence of patterns in the accidents, cluster analysis was performed to the data. Accident reports were also used when the role of human and organizational error was analyzed in Washington State Ferries risk assessment [68] and in an investigation process where marine traffic accident reports published by authorities of several countries were used for estimating occurrence probabilities and consequences of human [69]. Kiuru and Salmi [70] studied 92 marine accident investigation reports from Finland in order to find the impact of International Safety Management (ISM) Code on accident risk.

Many researchers have used and analyzed the data obtained from MAIB's investigations reports. Some studies used a specific case as an example while others provided summaries of historical accident cases. Fishing vessel accident studies can be found in [71] and [72]. Accidents occurring during transportations of packaged dangerous goods were investigated in [73]. Causal factors in the accidents of high-speed crafts are analyzed in [74] and [75] presents a human factors related study with some references to other articles with the similar topic.

According to Reason [76], the reliability of accident reports can be questioned, since they will always have a simplified presentation of the events and are mostly concerned with attributing blame. Accidents with no injuries are underreported, more severe accidents are investigated in more detail and a high risk of bias might be present when using accident investigation reports as data [67].

Accident reports are in text format and their usage typically requires human effort in extracting information from the text. The task can become tedious while humans may not always be capable of extracting the information objectively. The challenges have been attempted to tackle with text mining. Text mining is the process of automatically analyzing the contents of text documents for finding interesting features or patterns [77]. As an example, the role of lack of situation awareness in maritime accident causation was examined using a text mining software from accident reports [78]. The results from text mining were comparable to the manual analysis of the reports.

3.3 Incident and near miss data

3.3.1 The need and nature of incident and near miss data

Harrald et al. [79] stated that the collected marine accident data is not detailed enough for a human error assessment and suspected that it unlikely will ever be. Therefore they emphasized the need for data from 'near miss' situations for more advanced modeling and risk assessment.

Incident or near miss data can be collected by the shipping companies themselves or by flag state or port state authorities. Depending on the type of data, it may have been collected from voluntary or obligatory reporting. The following subchapters present a couple of types of near miss data from shipping companies and authorities.

3.3.2 Insjö/ForeSea

ForeSea is an anonymous and voluntary experience data bank initiated by Finnish and Swedish organizations and government agencies to improve maritime safety. The aim of the database is "to capture the conditions that are normally not reported to authorities" [80]. These include accidents, near misses and non-conformities. The database is a refined version of the Swedish information system Insjö which was launched in 2002. Two systems are currently running in parallel, but the plan is to replace Insjö with ForeSea.

Twelve companies are reporting to ForeSea, comparing with 76 members of Insjö [81]. The same data held in company's Safety Management System (SMS) are transferred to Insjö. This can be done automatically if the company has an IRIS (Incident Report Information System). So far 11 companies have it. An alternative method of contributing to the database is a report sent by the Designated Person (DP). Only the DP has login rights and is entitled to report to the experience bank. This is to ensure the credibility of data. After the report is received, edited and verified, information on the source is destroyed to protect anonymity. The database is administered by a third party.

DP forwards information obtained from one of the company's ships. This report is written in narrative form and should answer the questions what and where happened, and what were the causes and consequences of the event. The report is short and its quality depends on the reporter's skills. After receiving the report, Insjö/ForeSea administrator still has a possibility to contact the DP for additional information. The report is stored after being disidentified and connected with keywords to facilitate searching. In the future, ForeSea will also have a possibility of supplementing a field where more information (e.g. pictures) can be added. The database administrator is responsible for classifying the event (near miss) into 27 categories. Classification is done based on his interpretation of the event. According to [82], the task has not been perceived as difficult due to the fact that reports are short and the call-back function is on administrator’s disposal. For each report, the administrator sends feedback information to the DP. These are in form of similar cases; up to ten cases is found reasonable to send [82].

The philosophy behind the taxonomy is “what can be got into”, compared with EMCIP’s philosophy of “what a collector wants to get in” [82]. Data can be separated into five main categories: prerequisite data, the course of events, the causes, the consequences, and the measures. Each of these is further divided into subcategories, for example the causes are divided into human/manning, working environment, marine environment, technical ship and cargo and management causes. This is presented in Table 1. The taxonomy enables analysis with Ishikawa diagram [83] for all data and for a specific report. An example of such a diagram from ForeSea can be found in Figure 3. When the database was accessed (without a login), for some reason it was not possible to see the diagram for all accident reports, while for all near accidents there were no problems.

Approximately one report per year per ship is obtained in Insjö. The goal is to have ten reports per vessel per year [82]. On the 7th of December 2011, Insjö experience data bank contained 1282 accident reports, 841 near misses and 532 non conformity records, in total 2655 cases. 1268 reports are transferred into ForeSea. All reports in ForeSea are written in English.

There are less than ten active companies which contribute to Insjö [82]. Other companies are more interested to browse the cases than to contribute with their own reports. The most active companies are the ones with IRIS system (when the report is sent automatically), or the ones for which reporting is market driven such as tanker operators. To enlarge the database and secure the anonymity, the plan is to include other countries besides Sweden and Finland into the system.

After ForeSea becomes fully operational in July 2013, every individual member company will be required to provide reports to the database every year. The reason behind the requirement is to “ensure that experience data bank will grow and that no shipping company uses experience of others without providing his own” [80].

Data stored in the Insjö database is available to four categories of users: public, visitors, designated persons and researchers. They all have different rights and not all features are available to all users. While public has the most limited entrée, covering only recent reports and key figures (not able to login), researchers have access to the most features, including also a right to export data to Excel format. Access to researchers has to be granted by ship owners, it is connected with a specific project and it is time limited. Only the administrator has a full access.

Studies on Insjö database can be found in [84] and [85]. More recent studies on near-miss reporting practices in Finland and Sweden are presented in [86] and [87].

Table 3. Data separation scheme of the Insjö database

PREREQUISITIES
Report ID
Regdate
Type of event
Activity of the ship
Activity on board, type of work
Location

COURSE OF EVENT
Event heading
Mechanism
Contact

CAUSES
Human/manning
Working environment
Marine environment
Technical ship and cargo
Management

CONSEQUENCES
Individual
Environment
Ship
Third party

Human/manning
Not reported
Crew composition
Criminal action
Culture/Language
Education/Training
Familiarization
Individual diminished ability
Individual motivation
Individual mental action
Qualification/Competence
Mental stress
Other

Working environment
Not reported
Living conditions
Occupational health and safety standard
Personal protective equipment
Protection device/Safe guards
Professional leadership and teamwork
Safety training standard
Workplace design/Ergonomics
Working conditions
Other

Management
Not reported
Bridge and control room procedures
Communication and information
Contigency planning
Emergency response
Familiarization
Leadership and teamwork
Reporting and corrective actions
Responsibility/Supervision
ISM instructions and manuals
Training
Work organization
Work planning
Other

Technical
Not reported
Inspection/Test/Approval
Installation
Maintenance
Passenger
Quality of materials
Reliability/Lack of equipment
Repair
Ship/Equipment design
Stewing/Packing/Lashing
Technical documentation
Other

Marine environment
Not reported
Ice conditions
Navigational conditions
Pilot assistance
Yard, port and tug assistance
SAR operations
Traffic/Navigational information
Traffic situation and other ships
Visibility
Water/Sea state
Warfare/Piracy
Wind force
Other

Taking into account that the similar causes govern accidents and near misses, it should be possible to use near-miss data for accident modeling. Insjö and ForeSea contain only a short description of the event in narrative form, with very little factual data available (the ship type, type of event, the activity of the ship, the location). Hence, traffic models cannot benefit from these two databases. Utilization might be possible in accident models, but as with accident investigation reports, one should go through all reports and extract information manually.

Even though a shipping company would not report to Insjö or Foresea databank, the ISM code still requires a near miss reporting from all SOLAS ships and thus these ships should have collected near miss data anyway. It can be assumed that the content and quality of the internal reports is the same as of the ones reported to the near miss databases. The number of reports can be larger, though. The reports are used as learning opportunities on case by case basis. However, data is not yet utilized for establishing trends [87]. Consequently, it can be assumed that any accident models have not been built either.

3.3.3 Vessel Traffic Service data

Vessel Traffic Service (VTS) provides information and navigational guidance to the vessels navigating in a VTS monitoring area. In Finland, VTS is operated by the Finnish Transport Agency and the information is given in Finnish, Swedish or English [88]. In addition, the VTS centers can organize the traffic in the area. The information the VTS provides, such as waterway conditions, icebreaker assistance and other traffic in the area, can be given when the ship reports her arrival to the VTS area, when necessary, or when requested by the ship. The navigational guidance can be given to an identified vessel by request or if the VTS center finds it necessary given the circumstances. However, the guidance is only advisory and the master of a ship remains responsible for the maneuvering. The aim of the traffic organization is to avoid dangerous encounters and traffic jams.

In Finnish territorial waters, vessels with a GT of at least 300 are obliged by law to participate in the VTS monitoring [89]. Participating means reporting their arrival to the VTS area and active listening to the VHF channel of the VTS monitoring. The vessels not obliged to participate in the VTS monitoring are also recommended to listen to the channel. In Finland, all VHF traffic and traffic image data from the VTS centers is recorded. The recordings must be stored for 30 days.

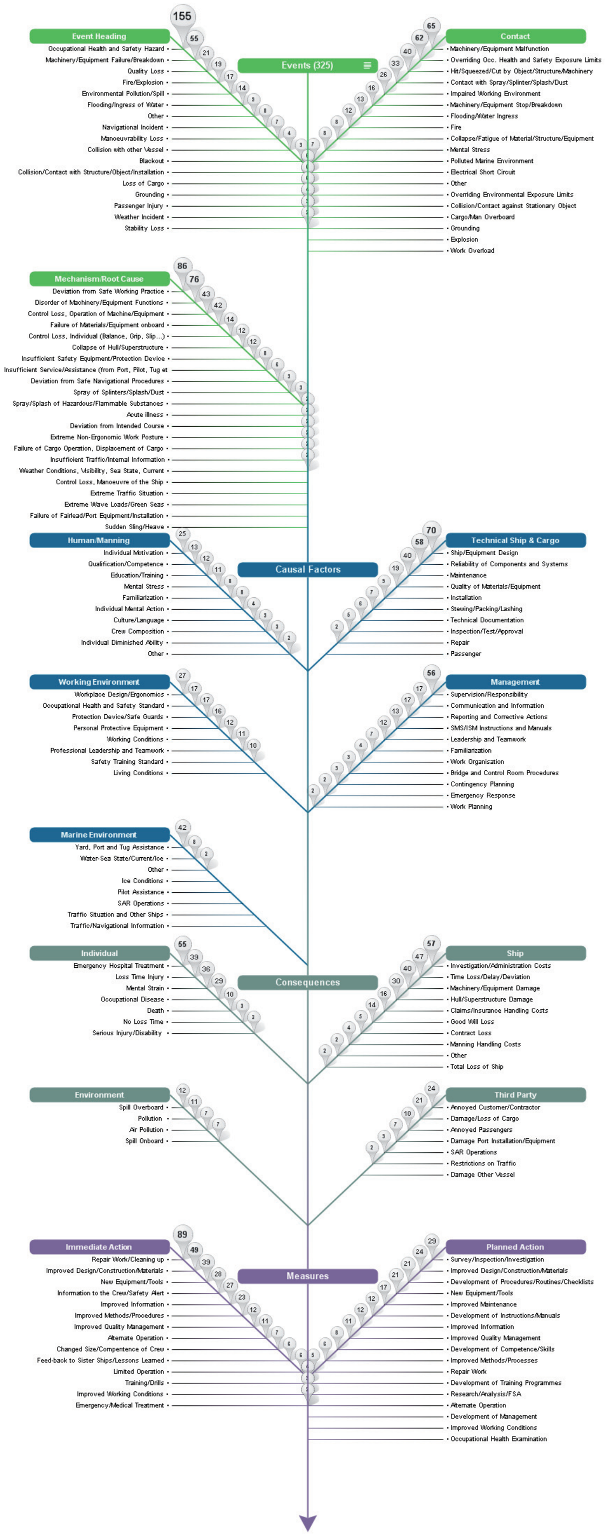

Figure 3. An example of Ichikawa diagram of the ForeSea database describing the near accidents in the database.

Table 4. Information fields of the Finnish VTS violation and incident reports from the year 2009. In addition, a capture of the situation on ECDIS is attached to the report which may include additional AIS information about the vessel's speed, course and heading. The filling percentages are calculated from 21 VTS incident forms and 37 violation forms from the first six months of 2009.

	Violation report		**Incident report**	
Type of information	**Field**	**Type of field and filling %**	**Field**	**Type of field and filling %**
Vessel identification	Name Flag Port of registry Callsign Type IMO Number MMSI GT	Text (100 %) Text (100 %) Text (65 %) Text (100 %) Text (100 %) Text (100 %) Text (100 %) Text (76 %)	Name Call sign IMO Number Pilot Master	Text (95 %) Text (90 %) Text (76 %) Text (38 %) Text (0 %)
Time	Date and time	Text (92 %)	Date and time	Text (100 %)
Position, speed and course	Latitude & longitude	Text (100 %)	Position Destination	Text (86 %) Text (81 %)
Location	Territorial waters of Finland / international waters Outside scheme / Traffic Separation Scheme / Lane / Separation zone / Other location	Check box (100 %) Check box Check box / Text (name) Check box / Text (desc.) Check box Check box / Text (desc.) (76 %)	Hanko VTS Helsinki VTS Kotka VTS GOFREP	Check box Check box Check box Check box (95 %)
Identification	Plotted by Radar / Plotted by AIS Identified by	Check box (89 %) Text (GOFREP or VTS) (0 %)	-	-
Weather	Wind direction Wind force (m/s) Sea state (douglas) Visibility (m)	Text (68 %) Text (68 %) Text (22 %) Text (8 %)	Weather	Text (visib. 67 %, wid dir. 95 %, wind force 95 %)
Type of non-conformity	-	-	Near miss Accident AIS Environment Pilot Equipment Personal injuries Emergency Other	Check box Check box Check box Check box Check box Check box Check box Check box Check box (100 %)
Contravened regulations	Rule 10 (b) i Rule 10 (b) ii Rule 10 (b) iii, joining Rule 10 (b) iii, leaving Rule 10 (c) Rule 10 (d) Rule 10 (e) Rule 10 (f) Rule 10 (g) Rule 10 (h) Rule 10 (i) Rule 10 (j) IMO Resolution MSC.139(76) Annex 1 Other rules	Check box Check box Check box Check box Check box Check box Check box Check box Check box Check box Check box Check box Check box Check box / Text (95 %)		
Addiditional information	Details of the incident	Text (97 %)	Description of incident Actions taken by VTS Operator Operator Supervisor	Text Text (Descr. and/or actions 100 %) Text (100 %) Text (95 %)

In the Gulf of Finland, Mandatory Ship Reporting System GOFREP area covers the international waters and Finnish and Estonian territorial water areas not included in their VTS areas. Helsinki VTS, Tallinn VTS and St. Petersburg VTS centers monitor the GOFREP area and provide guidance to the vessels. Ships over 300 GT must report their arrival to the area or when they are leaving a port in the Gulf. Smaller vessels must report if they have some problems with the maneuvering ability, for example [90].

In Finland, VTS operators should report all violations they observe within the Finnish VTS areas and the GOFREP area. Also, incidents or near misses within Finnish waters are reported. The violations of regulations are reported to the maritime authorities and to the flag states. However, differences in the numbers of reported violations between VTS operators have been detected [S. Talja (Finnish Transport Agency/Gulf of Finland Vessel Traffic Centre), personal communication, 7th of October, 2011]. In 2010, a total number of 125 incident and violation reports were made at the Gulf of Finland VTS center.

The format of the violation and especially the incident reporting forms has slightly varied over the years but the basic structure, a narrative text field for describing the event and a few check box –type options for the location or circumstances, has remained unchanged. The information the reports covered in the first half of the year 2009 and the fill-up percentages is presented in Table 4. In addition to the filled form, a capture (or captures) of the situation on a sea chart is typically attached to the report. These captures may include additional information from the AIS such as the vessel's course, heading and speed. At the beginning of 2012, the reporting system will be reformed. All reporting will then be done into an electrical system. At the moment of writing this (October 2011), information about the details of the system or the contents of the reporting forms was not yet available.

Based on two two-week periods of Archipelago VTS, West Coast VTS and Gulf of Finland Vessel Traffic Centre operators reporting all the situations requiring VTS intervention, the work of the VTS was described both verbally and statistically [91]. Salmi [53] used violation reports for identifying accident-prone vessels by comparing the vessels present at the violation reports between 2004-2008 to HELCOM accident statistics. In the study it was found that for 2007 accidents, 15 % of the reported accidents had occurred to a vessel identified by the VTS reporting beforehand.

VTS violation and incident reports can be used in identifying risk-prone vessels for risk modeling purposes. The categorized data in the reports does not provide much input to the risk models. Weather, rule 10, The VTS incident and violation reports provide information on the situation itself. In order to use VTS violation or incident reports in quantitative risk modeling, the information about the situation, the vessel(s) and the circumstances must be transformed into categorical data, which, as already stated for accident databases in Chapter 3.2.1, may introduce some uncertainty. On the other hand, as with accident investigation reports, finding the truth behind the textual information may also be challenging. Nevertheless, the advantage of VTS violation and incident reports is that violations and incidents occur more frequently than accidents and thus there is more data to be utilized.

3.3.4 Port State Control inspection data

Port State Control (PSC) is the inspection of the condition, equipment, manning and operation of foreign state vessels conducted by the port state authority when the foreign ships are visiting a port in the port state [92]. The purpose of the PSC inspections is to verify that the aforementioned aspects on board comply with the international regulations. Finland is a member of Paris Memorandum of Understanding (Paris MoU),which is an agreement on a harmonized system on Port State Control covering European coastal states and the west coast of Canada [93]. Similar MoUs cover all oceans in the World.

Since the beginning of 2011, the priority, frequency and scope of the Paris MoU inspections are determined with Ship Risk Profile. Ship Risk Profile classifies ships into High risk ships, Standard risk ships and Low risk ships. It is determined based on various factors such as ship type, age, flag, company performance and the number of deficiencies recorded in the previous inspections. The details of determining the Ship Risk Profile can be found in the Paris Mou text [94]. As some of the factors behind the Ship risk profile are dynamic, it is updated daily. Ships are inspected periodically with an inspection interval depending on the ship risk profile: 5-6 months after the last inspection in the Paris MoU region for a high risk ship, 10-12 months for a standard risk ship and 24-36 months for a low risk ship. In case of presence of overriding or unexpected factors listed in the Paris MoU text, and additional inspection must (overriding factor) or may (unexpected factor) be carried out before reaching the end of the inspection interval. Before the Ship Risk Profile was established, the inspected ships were chosen very similarly (e.g. as in [95]).

Paris MoU inspections can be divided into four categories [96]. An initial inspection visit consists of checking certificates and documents listed in Paris MoU text [94], performing an overall condition and hygiene check of the ship and verifying that any possible deficiencies found in the previous inspections have been corrected as were required. If during an initial inspection there are clear grounds to believe the ship may have some deficiencies, a more detailed inspection is carried out. These clear grounds are mentioned in the Paris MoU text. A more detailed inspection will cover the area where the clear grounds were established or that are relevant to overriding or unexpected factors and areas chosen randomly from the following list [94]:

1. Documentation
2. Structural condition
3. Water/weathertight condition
4. Emergency systems
5. Radio communication
6. Cargo operations
7. Fire safety
8. Alarms
9. Living and working condition
10. Navigation equipment
11. Lifesaving appliances
12. Dangerous goods
13. Propulsion and auxiliary machinery
14. Pollution prevention

An expanded inspection will cover all the categories mentioned above. The fourth inspection category is a concentrated inspection campaign. It has a certain focus area and runs for a limited time, during which all PSC inspections will additionally address the details of this area. As an example, a three-month campaign on structural safety and load lines was launched in September 2011 [96].

The results from the PSC inspections are gathered to a database that is accessible by public on Paris MoU web site [97]. ParisMoU inspections are also available through a European PSC database THETIS [98].

The web interface provides a possibility to search the inspections of a certain vessel identified by IMO Number and/or Name. It can also be used when searching multiple vessels based on their flag, ship type, size, age, classification society, the date period of inspection, port state, type of inspection, inspection port, the number of deficiencies and/or duration of detention. The search results in information on the factors listed in Table 5. As an example, when writing this report, a search of all ships under Finnish flag resulted in a list of 792 inspections, and a search of inspections conducted by Finland as the port state resulted in 1596 inspections, inspected between 29 October 2007 and 24 October 2011.

Table 5. The information the Paris MoU inspection database search provides on an PSC inspection of a vessel

Information type	**Details**
Ship details	IMO Number Type Name GT Flag Keel Laying Date
ISM Company	IMO Number Name Address City Country
List of charterers (if any)	Type Name Address City Country
List of Class Certificates	issuing authority issue date expiry date
List of the Statutory Certificates	Certificate Issuing authority issue date expiry date Surveying authority Date of last survey Place of last survey
A list of the ports in route	
Inspection Details	Type of Inspection Place of Inspection Date of first visit Data of final visit Nb. of Deficiencies Nb. of Deficiencies ground for detention
List of the inspected areas	
List of the operational Controls Carried Out	
A list of the deficiencies	Area Defective item Nature of defect Ground for detention RO Related

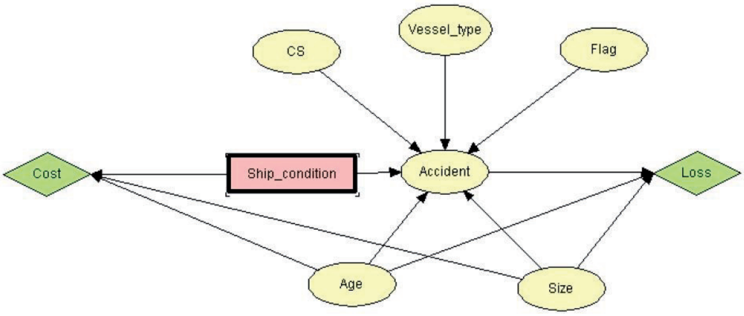

Figure 4. A Bayesian network model of ship accidents proposed by Li et al. [101]. Port State Control data, accident data and ship fleet data were used for the model parameters

Port State Control data has been used in many previous studies. In a data analysis of 42 000 Indian MoU inspections [99], types of deficiencies found during inspections and the changes in these deficiencies over time and between successive inspections were examined. Knapp and Frances [100] studied the effect of PSC inspections on the probability of accidents and incidents using a binary logistic regression model. Their data consisted of more than 180 000 PSC inspections for approximately 26 000 ships, over 11 000 casualty records from Lloyd's Register Fairplay, Lloyd's Maritime Intelligence Unit and the IMO, and, from Lloyd's Register Fairplay, information on almost 44 000 ships for inspections. Li et al. [101] also used PSC inspections, accident data and static ship data in their model. From the data, they constructed the probability parameters of a Bayesian network model for shipping accidents. The structure of their model, which was based on expert assumptions and accident data, can be seen in Figure 4.

The information the PSC inspection database contains is easily usable in quantitative risk modeling as the information is in categorical format. However, PSC data alone only provides information on the deficiencies and inspection history of a vessel. Also, it should be kept in mind that PSC inspection data is not describing the ship fleet on average but the vessels chosen to be inspected.

3.3.5 Occupational safety data

Within the maritime domain, insurance companies and authorities collect occupational safety data [102]. According to interviews within Finnish shipping companies [102], the levels of occupational safety and general maritime safety are not independent. Thus occupational safety data might bring some additional information for marine traffic accident models as well. With the hypothesis that an indicator for occupational safety, such as the number of occupational safety incidents or the lost time incident (LTI) frequency, and the number of marine traffic accidents, such as the number of collisions, are dependent, one could use the occupational safety data for identifying the accident-prone ships. In a model describing the human and organizational factors, an indicator of occupational safety such as LTI frequency could be seen dependent on a company's level of safety culture. It would thus provide indirect information on the hidden safety culture variable, which can be hard to measure and model as such. Models with hidden variables are discussed more in Chapter 4.3.

3.4 Insurance company data

Hull&Machinery (H&M) and Protection&Indemnity (P&I) are two of the best known ship insurances. The former is more related to technical damages to the ship, while the P&I insurance is connected with the operation of the vessel and covers claims related to crew, cargo and liabilities for pollution and wreck removal. There are two categories which can be covered either by H&M or P&I insurance and these are [103]:

1. Collisions – damage sustained to the ship and sometimes also liability towards the other ship
2. Striking other objects – damage inflicted on the own ship and sometimes also liability towards the owners of the other object

As the interest is mainly for data which can help in modeling human and organizational factors, P&I insurance is found more suitable to study.

In Finland, Alandia Marine provides insurance services. Company operates in the Baltic Sea area and had 1847 insured vessels on 31 December 2010, of which more than 100 are Finnish flagged ships [104]. Alandia offers the service of H&M, but P&I insurance is "developed only for smaller tonnage (not exceeding 2,000 GT) in limited trade". On their web pages [ref to the web page], links are given to two P&I Clubs, namely Gard (Norway) and UK (United Kingdom). We are focusing on the latter. UK P&I Club handles 7000 claims a year [105].

Even though the UK P&I Club is aware that "human error dominates the underlying causes of major claims", the Club "does not record the root causes of the human error which played a part in the incident" [106]. Statistical data on issues such as fatigue, lack of training, inadequate management or the myriad of mental, motivational or emotional causes of human error are not collected, as they have been found to be unnecessary in settling liability claims. They have promised to study and publish report about the root causes of human errors. This was stated in 1997. Data will be classified using taxonomy from the US Coast Guard. Despite the fact that the Club "has for some years sought a methodology for both defining and analyzing human error in the maritime context" [105], no report on root causes has been published so far.

P&I Clubs conduct regular inspections and surveys on ships owned by their clients [107]. Vessels are chosen randomly and, by the end of 1994, 2000 ships had been visited. At present, there are more than 600 ship visits every year. Inspection visits "should supplement the owner's own management system" [107]. A visit lasts approximately four hours and it is done by P&I Club's own inspectors. The purpose is to see whether international and classification society's requirements are met regarding cargoworthiness, manning, general maintenance, safety including safety working practices, operational status and pollution. The inspectors have a guide notebook containing relevant questions [107]. No information is available on how the answers to these questions are stored or if they are stored at all. Interestingly, two forms contain information relevant to human factors modeling. Officer qualifications are recorded in those, as well as information about manning and management, language and pilotage. If the inspector makes a comment, it will be recorded. Inspectors are also armed with a digital camera and any information obtained in this manner is attached to the report.

P&I Clubs also carry conditional surveys whose main purpose is to agree on the damage cause, nature and extent [108] . These are done by independent consultants. The process applied at Swedish Club is explained here. According to Swedish Club web pages [108], “Before attending a casualty, the surveyor will search the Club’s own computerized records for any claims associated with the vessel that have not been finally settled. The surveyor will include details on the status of such casualties in the report, if relevant”. The report forms are available online [109]. Vessel particulars, crew matrix information, circumstances of the survey and some answers to the survey questionnaire could be used in modeling human and organizational factors. The following documentation is available to a surveyor [108]:

- Vessel log books, covering relevant period
- A signed statement from the Master, Chief Engineer and/or ship’s personnel directly involved
- Vessel’s ISM damage/non conformity report
- Maintenance records
- Classification records
- A repair specification, if available
- A statement outlining the cause of the damage, in the owner’s opinion, and any documentation supporting the owner’s statement
- Drawings

In addition to inspections and surveys, P&I Clubs also provide their members with Pre Engagement Medical Examination (PEME) program, where “accredited clinics are held accountable to both the Club and Members for their performance” [110]. It is unclear whether this also means that Clubs possess medical data obtained during examination.

It can be concluded that insurance companies have a large amount of data potentially usable for risk modeling, but these are not available to public or researchers. Hence the structure of the database, the quantity or the quality of the data cannot be checked.

3.5 Classification societies

Classification societies have a large role when it comes to the safety of a ship as a system. If suspected that ship safety might be endangered, Class can intervene during design and construction phases, as well as during surveys. Insight into different data is on their disposal, starting from drawings, material deficiencies, maintenance actions, etc. The collected data is mainly related to technology and might be used in a reliability assessment or for maintenance studies. Classification societies use accident data to revise the Rules. Information about hull characteristics can be used for modeling the consequences in case of a grounding or a collision.

Except for accident reports (see for example DNV web pages [111]), the majority of the data which the classification societies collect is confidential. Nevertheless, some societies offer other services, too. For example, Lloyd's Register Group's principal business in the maritime domain is the classification of ships. Nevertheless, they have offered different commercial services as well. An example is Lloyd's Register Fairplay (LRFP), which after ownership changes is nowadays entitled IHS Fairplay [112]. In the IHS Fairplay, data on vessels characteristics and accidents of the world fleet is available. LRFP was investigated for example in [113].

Lloyd's List Intelligence, formerly known as Lloyd's Marine Intelligence Unit (LMIU) is a primary provider of global commercial maritime data and the only provider of global shipping movements [114]. In LMIU, AIS data is integrated in GIS (Geographical Information System). Besides positions reported by AIS, GIS also displays static data sets, including maritime charts, showing clear harbor, coastal and waterway areas [114]. Historical vessel tracks along with dynamic real time reports are shown. Data is archived for 7 days online [114]. According to [115,] the classification of the accidents in LMIU applied to the first event that has occurred and hence did not include other consequences that may have happened in the same accident. Similar is stated in [116], where the authors were able to populate only the top event in their fault tree analysis. [117] found that LMIU is poor at picking small spills and criticized its false causality and design. Still, the Lloyd's List Intelligence database remains as one of the most popular as it provides data on the global level.

Other services originally provided by Lloyd's Register are IHS Fairplay World Fleet Statistics, World Shipbuilding Statistics, and World Casualty Statistics, nowadays published in electronic format [118]. The latter is published annually and contains a summary of reported losses and disposals of merchant ships. The casualty incident categories include foundered, fire/explosion, collision, wrecked/stranded, contact, and hull/machinery

[119]. Regarding individual ships, reported information include ship name, her flag, Gross Tonnage, year of build, casualty incident location and a summary of any casualty incident suffered, including the fate of the vessel and crew [120].

To support their clients, Lloyd's Register Group also offers a service called ClassDirect LIVE. This online tool is available for operators of Lloyd's Register classed ships. The information is confidential and the operators can access data on ships in their fleet with a given password. Provided information is held on the Lloyd's Register Group's databases. The following main items are offered through ClassDirect LIVE [121]:

- Fleet particulars irrespective of Classification Society
- Up-to-date Survey Status for all LR-classed vessels
- Hull and Machinery Master Lists
- Up-to-date status of Condition of Class and Memoranda items
- Detailed Survey Histories with complete Survey Reports for at least 12 years
- Incident summaries which link to details of hull and machinery defects
- Details of Hull and Machinery "as built" configuration
- Survey Checklist as used by LR Surveyors
- ISM Code certification status for all ships, irrespective of Class
- Access to Rules, Regulations, Classification News, Approved Suppliers Lists and technical services
- Owners/Operators of ships with the ESP (enhanced survey programme) notation, can view "hull related" ESP survey reports

To conclude, classification societies have information only on the ships under their class. The most comprehensive is data regarding vessels particulars. Accident data is only collected when a surveyor is called or whena failure is observed during a survey [122]. Further, the focus is on technical failures. This data is confidential [117] and as such not available to the public. If advancements in technology are rapid, failure data collected from previous accidents will not be usable for modeling future systems. For a detailed risk and reliability analysis, the data from class societies might be insufficient.

3.6 Equasis

Equasis database combines multiple existing maritime safety related data sources of the word's merchant ships. For registered users, the database is freely searchable online [123]. The data providers for Equasis are listed in

Table 6. It is stated on the Equasis web page [124] that a special attention is paid to the accuracy and validity of the data and that the database is being constantly improved.

In Equasis, one can search information on a specific ship (based on the IMO number, name or the call sign) or on a specific company. The following ship information is returned:

- Ship information
    - IMO number
    - Name of ship
    - Call Sign
    - MMSI
    - Gross tonnage
    - DWT
    - Type of ship
    - Year of build
    - Flag
    - Status of ship
    - Last update
- Key indicators
    - The ship is classed by (at least) one of the IACS member societies (Y/N)
    - The ship's flag is not on the black list of the Paris MoU (Y/N)
    - The ship's flag is on the white list of the Paris MoU (Y/N)
    - Percentage of inspections having led to a detention in last 36 months
    - The ship's flag is not on the targeted list of the USCG (Y/N)
- List of management details
    - IMO number
    - Role
    - Name of company
    - Address
    - Date of effect
    - Details (a link to company info)
- List of classification status
    - Classification society
    - Date change status
    - Status
    - Reason
- List of classification surveys
    - Classification society

  - Date survey
  - Date next survey
  - Details (link to ClassDirect LIVE)
- P&I Information
  - Name of P&I insurer
  - Date of inception

For inspections and manning, the search returns a list of previous port state controls including the PSC Organization, authority, the port of inspection, the date of the report, detention (Y/N), detention duration, the number of deficiencies per category and details about the statutory and classification surveys at the time of the inspection. For passenger ships, information on Ferry directive is provided. Also, the ILO convention by flag state and the working conditions and collective agreement is shown. The search also provides historical information about the ship such as her former name(s) , flag(s), classification(s) and companies.

Table 6. Data providers for Equasis [125]

Category	**Provider**
Port State Control Regimes	Paris MOU on Port State Control (PMOU) The US Coast Guard (USCG) Tokyo MOU on Port State Control (TMOU) Indian Ocean MOU on Port State Control (IOMOU)
Private inspections	Chemical Distribution Institute (CDI) Oil Companies International Marine Forum (OCIMF)
IACS Classification Societies	American Bureau of Shipping (ABS) Bureau Veritas (BV) China Classification Society (CCS) Det Norske Veritas (DNV) Germanischer Lloyd (GL) Korean Register of Shipping (KRS) Lloyds Register (LR) Nippon Kaiji Kyokai (NKK) Registro Italiano Navale (RINA) Russian Maritime Register of Shipping (RS)
Associate Members of IACS	Indian Register of Shipping (IRS)
Other Classification Societies	Türk Loydu (TL)
International Group of P&I Clubs	American Steamship Owners Mutual P&I Association Inc.(American Club) Assuranceforeningen Gard - Norway Assuranceforeningen Skuld – Norway Britannia Steamship insurance Association Ltd Japan Shipowners P&I Association London Steam-Ship Owners Mutual Insurance Assoc. Ltd (The London Club) North of England P&I Association Steamship Mutual Underwriting Assoc. (Bermuda) Ltd The Shipowners' Mutual P&I Association (Luxembourg) The Standard P&I Club The Swedish Club The West of England Shipowners UK P&I Club
Other	Green Award Foundation Intertanko Intercargo Intermanager International Maritime Organization (IMO) International Labour Office (ILO) International Transport Workers' Federation (ITF) IHS Fairplay (IHSF) (previously Lloyd's Register Fairplay (LRF)) European Maritime Safety Agency (EMSA) Q88

4. Challenges of using data in marine traffic accident models

4.1 Challenges identified in the IMISS conference

The reporting procedures and data quality of near-misses and accidents were discussed in an International Maritime Incident and Near Miss Reporting Conference (IMISS) held in Espoo, Finland on 1-2 September 2011. In addition to presentations, expert workshops were organized during both days of the conference. Different stakeholders from the maritime domain were participating in the workshops. These included representatives from Finnish and Swedish shipping companies, representatives from Finnish and Swedish maritime authorities and investigation bodies, administrators of near miss databases, and researchers from multiple countries. An insurance company representative gave a presentation at the IMISS conference, but did not participate in the workshop.

The themes of the first day workshops were the barriers, the benefits, and the future development of near miss reporting. The results of these workshops can be found in [81]. On the second day, the discussion was aiming to find how different stakeholders can benefit from accident/incident modeling. The second problem was to identify factors affecting the quality of accident/incident data and to suggest improvements. Unfortunately many participants were not familiar with what accident models mean. This might be because "modeling" is mostly done within science. Nevertheless, the outcomes of such models should be known to all participants as they are often open for the community. This can also be one of the reasons why shipping companies do not benefit more from the reporting process. Data that they provide are used by consulting companies and sold to the same shipping companies. Later, discussion was directed to data quality analysis. All participants were aware that data have to be improved and that this is an important issue. It was recognized that mistakes cumulate during the complete reporting process, i.e., from reporting, collecting information, converting those into databases fields and storing them. Regarding the quality of the data, the main factors summarized from day two workshops were:

- Under-reporting; the recorded data does not represent reality
- Each database has its own taxonomy
- Changes of taxonomies during time
- Too strict or too wide categorization
- Missing data – empty fields
- Incorrect data
- Inadequate search engine
- Restricted or denied access

4.2 Discussion on how to improve data

This report has described various sources for marine traffic accident modeling. Whether a database can be seen as useful for extracting information for accident modeling depends on its structure and interface, but most importantly on the data it contains. The former can be improved by technology advancements including better search engines. Taxonomies can even be omitted if searching is possible by word. This is important as different databases have different classifications of events. Another important factor is allowing for multiple causes and entering information in sequence of events. A correct format and the fulfillment of important fields can be forced, but this does not mean that valid data will be entered. Mistakes cannot be avoided when people are populating the database, they can only be decreased. Also, databases typically change after some time when it is necessary to implement new points of view, e.g. when human factors were found to be the most dominating factor as accident causes or when new regulation is forced. Some databases just made abandon the old data and start populating the updated database with new data, while other transfer old data into a new system. From accident modeling perspective one should be aware of changes and be careful if going to combine old and new data. Having a new system does not equal improving the quality of the database. As already stated, much depends on data stored inside the database. Databases give a framework for the data they require, however they are often led by what one wants to include rather than by rationalizing what can be obtained in practice. To overcome this, the end users and investigators should cooperate when designing and populating a database.

To analyze the data quality, we have to study reporting practices in the maritime industry. Reporting is an important part of the ISM code requirements. In an ideal situation, all accidents and incidents would be reported. However, in practice this is not the case. By comparing data from different databases [119], it was found that the number of unreported accidents makes roughly 50% of all occurred accidents. The problem is the worst if the casualty occurred in international waters. A rough rule of thumb is that only 1 % of all maritime casualties are reported in such case [117]. This is quite a high figure which influences the outcome of any accident assessment and must be considered when interpreting the results. Under-reporting was also studied in [126]. Not perceiving the value of reporting and time were found to be important barriers to incident reporting. In [117], a reason for not reporting was found to be the crews believing that the incident will remain as a black mark no matter how blameless they really are. It was also stated that the crews believe that owners do not want to know about the incident. In [86], similar barriers to reporting in Finnish companies were found. Thus, to increase reporting, its purpose and the results have to be made clear to the crew. A general argument of increasing safety is perhaps too abstract, especially as a direct risk-reducing effect of reporting is challenging to measure.

There is an opinion that much more can be accomplished by training than by reporting. This is not discussed here further, but one should note that in either case the top management plays a key role. Every reaction to a report is feedback to the crew and clearly states whether management commitment is true or false. If a report results in a change and an improvement, it is more likely that reporting will be perceived by the crew as a positive and effective matter. This is a good ground for safety culture development. When a system works within a company it is expected that data will flow to external databases run by authorities or third parties. As some near miss databases are not available to shipping companies unless they contribute to the database with their own reports, a trust in the maritime industry does not exist. So to improve the reporting process, openness and trust have to be improved. Near misses databases where the ship owners share information are showing that there is a place for optimism.

Reporting might also be increased by having an easy to use system. Contribution should be facilitated by using computers, mobile phones and cameras, rather than a paper format.

So far it has been discussed how to increase the number of reports, which is important if the data were to reflect the reality. However, having more cases in the data does not automatically imply that the data is adequate for quantitative modeling. If one wants to model accidents, the logical way is to start from causes. It is questionable whether causes are assigned correctly in different databases, and whether they ever can be. Reason behind this is that casualties do not partition themselves into neat categories [117]. Thus databases with 'no taxonomy' would be a better option. The process should be uniform and standardized between different countries. Also, much depends on the investigator and his/her perspective and accumulated experience. Biases can be avoided if two investigators with different backgrounds can work on each case. Further, investigators should have basic knowledge on quantitative risk analysis to distinct what data are important and what are not relevant at all.

Many of the existing data sources contain errors. This is especially true for traffic data such as the IMO number. Latitude and longitude, if known at all, are in many cases wrong. Many fields in databases are empty. If data from navigational equipment is not saved in time, necessary information is lost. Investigation should start as soon as possible to preserve all evidence and to prevent any changes in witnessing. Bridge team should be trained how to save navigational data recorded 24 hours prior to an accident.

Summarizing all aforementioned, it can be said that an effort should be made to report all accidents and to support the reporting of near misses. In that case, we will have better models and thus a more realistic picture of the accidents and safety level of the marine traffic. Correct data should be entered into databases, which can be partly forced by a suitable technical design and partly by training the people who populate the databases. Different databases should find a standard way of assigning causes and allowing multiple causes and sequential descriptions. Database administrators should try to check all available sources. Also, if an event should be included in two databases, it must not occur that one database contains the information on the event while the other does not.

4.3 Discussion on how to improve modeling

So far the various data sources and the shortcomings of data have been described but little is said about how the data is utilized in building a quantitative model. Further, some of the data deficiencies might be compensated or taken into account by choosing the modeling approach carefully. In this chapter, a brief introduction to building models from data is described, followed by a discussion on how to improve the validity of the models given the data deficiencies.

Depending on the problem to be modeled and the data available, the data can be utilized in the quantitative model construction in many ways. Especially in engineering and nature sciences, there are often the laws of physics and other knowledge on the phenomenon to be modeled available. In this case, the mathematical or probabilistic representation of the model is known and data is used for determining the unknown model parameters, for example. Sometimes there is no knowledge on the dependencies between the model variables or even on the correct variables involved and the complete model has to be learned from the available data. However, the dataset might be so large or complex that humans cannot construct the correct model from it. In this case, machine learning techniques (see e.g [127]) could be used in learning the model automatically from the data.

In a case where the data available contains information on relevant variables and the quantity of the data is not a problem either, values might still be missing from the dataset. The selection of the approach to handling missing data depends on the way the data is missing and it should be done with care. Missing data can be tackled for example using only the cases with complete data, deleting case(s) or variable(s), applying imputation methods, or using model-based procedures [128]. Imputation means estimating the missing values based on the other values in the dataset. For example, the averages of the other available valid values could be used, or a value of a similar or almost similar case could be substituted. Model-based procedures include methods such as the EM-approach [129]. EM-approach performs two steps iteratively: Finding a model that maximizes the likelihood of the given data values (E-step), and finding values for the missing data that maximize the likelihood of the model found on E-step (M-step). The values from the M-step are then used on the next E-step.

When the amount of data is limited, one could apply Bayesian approach to constructing a probabilistic model. In the Bayesian approach, in addition to the data, prior knowledge on the possible model is taken into account (e.g. [130]). The influence of the prior knowledge in finding the best model is following the well-known Bayes theorem:

$$P\ (M|D) = \frac{P(D|M)P(M)}{P(D)} = \frac{P(D|M)P(M)}{P(D)} \tag{1}$$

, where $P(M|D)$ is the probability of a model M given the data D, $P(D|M)$ is the likelihood of M, i.e., the probability of observing the data D given model M, and $P(M)$ is the prior probability of the model M. Depending on the problem, the prior probability could describe knowledge such as expert opinion or the laws of physics, or it could be based on a somehow related other dataset. For example, worldwide accident statistics could serve as prior knowledge when modeling the accident risks in the Gulf of Finland. A fully Bayesian approach means not choosing one model for describing the problem, but instead using the distribution $P(M|D)$ over all model candidates. On the other hand, if one wants to select only one model, the mode of $P(M|D)$ could be used.

Another way to overcome completely missing data on certain variables or otherwise deficient data is to combine multiple data sources which are somehow connected. Using multiple datasets together when building a model for e.g. classifying ships into accident-proneness categories might produce a more accurate model than using only one of the datasets alone. As an example, one could assume both the VTS violation report data and accident data provide information on ship's safety. Some of the variables in these datasets are common, but they also have unique variables. Further, there might be ships that appear in both datasets, and then again ships that are included in only one of them. Ship's safety level could be seen as a hidden variable that cannot be observed, but which is shared between the datasets. Combining multiple sources for machine learning has many names depending on the dependencies of the datasets and the learning task, for example multi-view learning, multi-task learning, transfer learning, co-training, and domain adaptation [131, 132, 133]. It should be kept in mind that whenever multiple data sources are combined, in order to have a valid model, the combination t should be done with extreme care [134].

5. Conclusions

Various data related to marine traffic and the accidents on the sea exist and the amount of data seems to grow in the future. Typically, the data has been collected for other purposes than for providing input to quantitative models. On the other hand, data is necessary for modeling and building a new database from an existing data is not an unusual practice in research. However, it is time consuming and does not present a practical solution.

This report examined possible sources of input data for quantitative marine traffic accident models. Summary of the data sources studied in this report is given in Table 7. These data sources differ in the scope and purpose and they all have their strengths and weaknesses. However, using any of them as the only source of input to a quantitative model seems risky, and if factors such as underreporting, errors and missing fields are not considered, the models may produce completely unreliable results.

To improve the models' validity, researchers need to decrease the uncertainty in the data. Double checking between two and more databases is necessary prior to a model population. Also, using the data together with prior knowledge might help. Combining multiple related data sources when learning the model from data could also be utilized.

It is much an easier task to point what is wrong and much harder to suggest sound improvements possible to implement. To fulfill the task promised in the report title, a more detailed study with expert involvement is needed. Additionally, as 'a quantitative marine accident model' can mean many things, the true feasibility of different data sources cannot be determined without applying the data to the modeling and then validating the results. In the end however, all improvements in the data or its handling will not matter, if the databases stay unavailable to the modelers and further indirectly to the stakeholders making the decisions based on the models.

Table 7. Summary of the feasibility and the drawbacks of data covered in the report

	Feasibility for accident modeling	**Drawbacks**
TRAFFIC DATA	past ship trajectories and routes can be extracted from data; can be used in dynamic ship traffic system reconstruction provides information also on “safe ships”, not only on ships in accident or incidents	contain errors and missing fields; large amount of data points - difficult to maintain and store
INVESTIGATION DATA	can be used for accident description; can be used for the analysis of causes	more severe accidents are investigated in more detail; accidents with no injuries are underreported; biases might be present during investigation; data have to be extracted from (long) text format; not all data can be summarized in the report; reports are often in a national language
ACCIDENT DATABASES	information is provided in categorical or numerical format which can be analyzed statistically; factual data about ship is available; field for narrative part exist; establishment of common taxonomy in EMCIP	different taxonomies are still an issue; often do not take into account multiple causes or describe the accident chains; missing fields and errors in data; do not contain all accidents which belong to scope of the database; fixed categorization; changes during time
NEAR MISS DATA	more data compared to accident cases; provide valuable insight into causes; can be used for analysis of barriers	no traffic data; no factual data except ship type; lot of "unimportant" cases to analyze; for some databases access has to be granted by stakeholders
INSPECTION AND CONTROL DATA	first-hand information whether a vessel is at risk; technical issues and safety management system well covered; data give insight on management commitment to safety	data is confidential and not available except in case of PSC data; not all ships in one area are inspected in the same time interval, e.g. one year; typically there is more data on the “risky” ships as the inspections are conducted more frequently on them; human factors are not checked sufficiently

Acknowledgements

The study was conducted as a part of Competitive Advantage by Safety (CAFE) project, financed by the European Union - European Regional Development Fund – through the Regional Council of Päijät-Häme, City of Kotka, Finnish Shipowners' Association, and Kotka Maritime Research Centre corporate group: Aker Arctic Technology Inc., Port of HaminaKotka, Port of Helsinki, Kristina Cruises Ltd, and Meriaura Ltd. The authors wish to express their gratitude to the funders. In addition, the participants of the IMISS conference workshops are thanked for sharing their expertise. Olle Blåfelt from and Sari Talja and Kati Westerlund from Finnish Transport agency are warmly thanked for the interviews and material provided for this report. The comments and support from the colleagues at Aalto University's Marine traffic safety research group and at Kotka Maritime Research Centre are highly appreciated.

References

[1] **MARISEC**. *Value of volume of world trade by sea*. Online [cited 13.02.2012]. Available: http://www.marisec.org/shippingfacts/worldtrade/volume-world-trade-sea.php

[2] **MARISEC**. *BIMCO – ISF Manpower 2010 Update. The worldwide demand and supply of seafarers*. Onlne [cited 13.02.2012.]. Available: http://www.marisec.org/Manpower%20Study.pdf

[3] **Kujala, P., Hänninen, M., Arola, T. & Ylitalo, J.** *Analysis of the marine traffic safety in the Gulf of Finland*. Reliability Engineering & System Safety 94(2009)8, pp. 1349–1357.

[4] **Heijari, J. & Tapaninen, U.** (eds.) *Efficiency of the ISM Code in Finnish Shipping Companies*. Publications from the Centre for Maritime Studies A 52, Turku 2010. 57 p.

[5] **Attwood, D., Khan, F. & Veitch, B.** *Occupational accident models – Where have we been and where are we going*? Journal of Loss Prevention in the Process Industries 19(2006)6, pp. 664-682.

[6] **Hänninen, M.** *Analysis of Human and Organizational Factors in Marine Traffic Risk Modeling: Literature review*. Helsinki University of Technology, Department of Applied Mechanics, Espoo 2008. TKK-AM-4. 51 p.

[7] **Jalonen, R. & Salmi, K.** *Safety Performance Indicators for Maritime Safety Management. Literature Review*. Helsinki University of Technology, Department of Applied Mechanics, Espoo 2009. TKK-AM-9. 60 p.

[8] **Fujii, Y. & Shiobara, R.** *The Analysis of Traffic Accidents*. Journal of Navigation 24(1971)4, pp. 534-543.

[9] **Hänninen, M. & Kujala, P.** *Modeling of Collision and Grounding Risks in Marine Traffic. Literature review (in* Finnish). Helsinki University of Technology, Ship Laboratory, Espoo 2007. M-299. 65 p.

[10] **Mazaheri, A. & Ylitalo, J.** *Comments on Geometrical Modeling of Ship Grounding*. 5th International Conference on Collision and Grounding of Ships ICCGS 2010, 14.-16. June 2010, Espoo, Finland. TKK-AM-16. p. 235-243.

[11] **Montewka, J., Hinz, T., Kujala, P. & Matusiak, J.** *Probability modeling of vessel collisions*. Reliability Engineering & System Safety 95(2010)5, pp. 573-589.

[12] **Goerlandt, F. & Kujala, P.** *Traffic simulation based ship collision probability modeling*. Reliability Engineering & System Safety 96(2011)1, pp. 91-107.

[13] **Hänninen, M. & Ylitalo, J.** *Estimating ship-ship collision probability in the Gulf of Finland*. 5th International Conference on Collision and Grounding of Ships ICCGS 2010, 14.-16. June 2010, Espoo, Finland. TKK-AM-16. p. 250-255.

[14] **Ståhlberg , K.** *Estimating Deformation Energy in Ship-Ship Collisions with Stochastic Modeling*. Master's thesis. Espoo 2010.

[15] **Ståhlberg, K., Goerlandt, F., Montewka, J. & Kujala, P.** *Uncertainty in Analytical Collision Dynamics Model Due to Assumptions in dynamic Properties*. 9th International Navigational Symposium on Marine Navigation and Safety of Sea Transportation, TRANS-NAV 2011, Gdynia, Poland.

[16] **Goerlandt, F., Ståhlberg, K. & Kujala, P.** *Influence of impact scenario models on collision risk analysis*. Accepted to Ocean Engineering.

[17] **Lecklin, T., Ryömä, R. & Kuikka, S.** *A Bayesian network for analyzing biological acute and long-term impacts of an oil spill in the Gulf of Finland*. Accepted to Marine Pollution Bulletin.

[18] **Ver Hoef, J.M. & Frost, K.J.** *A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska*. Environmental and Ecological Statistics 10(2003)2, pp. 201-219.

[19] **Larsen, O. D.** *Ship Collision with Bridges; The Interaction between Vessel Traffic and Bridge Structures*. IABSE, Zürich 1993. 131p.

[20] **Harati-Mokhtari, A., Wall, A., Brooks, P. & Wang, J.** *Automatic Identification System (AIS): Data Reliability and Human Error Implications*. Journal of Navigation 60(2007)3, pp. 373-389.

[21] **House, D. J.** *Seamanship Techniques: Shipboard and Marine Operations*. 3rd edition. 752 p. Butterworth-Heinemann 2004.

[22] **IMO.** *Automatic Identification System (AIS)*. 2011 (online) [cited 31.10.2011]. Available:
http://www.imo.org/OurWork/Safety/Navigation/Pages/AIS.aspx

[23] **Ristic, B., La Scala, B, Morelande, M. & Gordon, N.** *Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction*. 11th International Conference on Information Fusion. IEEE, 2008. pp.1-7.

[24] **Zhu, F**.*Mining ship spatial trajectory patterns from AIS database for maritime surveillance*. 2nd IEEE International Conference on Emergency Management and Management Sciences (ICEMMS), IEEE, 2011. pp. 772-775.

[25] **Friis-Hansen, P., Ravn, E. S., Hartmann, J. P. & Sørensen, A.** *FSA of the Navigational Safety in Baltic West*. In: Society of Naval Architects of Japan (ed.). Proceedings of the 3rd International Conference on Collision and Grounding of Ships ICCGS 2004, Izu, Japani 25.-27.10.2004. Society of Naval Architects of Japan. pp. 71-80.

[26] **Rambøll.** *Navigational safety in the Sound between Denmark and Sweden (Øresund); Risk and cost-benefit analysis*. Rambøll Danmark A/S, 2006.

[27] **van Dorp, J.R. & Merrick, J.R.W.** *On a risk management analysis of oil spill risk using maritime transportation system simulation*. Annals of Operations Research 187(2011)1, pp. 249-277.

[28] **Chang, S.J.** *Development and Analysis of AIS applications as an Efficient Tool for Vessel Traffic Service*. OCEANS'04. MTTS/IEEE TECHNO-OCEAN'04. IEEE 2004. pp. 2249-2253.

[29] **Sølvsteen, C**. *Analysis of AIS Data Quality* (online) [cited 24.2.2012]. Danish Maritime Safety Admnistration. Available: http://www.boos.org/fileadmin/user_upload/AnnualMeetings/Sopot_2009/Claus_Solvsteen_AIS_Data_Quality_bm4.pdf

[30] **Aarsæther, K.G. & Moan, T.** *Estimating Navigation Patterns from AIS*. The Journal of Navigation 62(2009)4, pp. 587–607.

[31] **Posti, A., Häkkinen, J. & Tapaninen, U.** *Promoting information exchange with a port community system* (in Finnish). Turku, Finland: University of Turku, 2010. 117 p. Publications from the Centre for Maritime Studies University of Turku, B 175. Available: http://www.merikotka.fi/mopo/tiedostot/Posti_Hakkinen_Hyle_Tapaninen_Satamayhteison_informaatiokeskus_tiedonvalityksen_tehostajana.pdf

[32] **Rautiainen, P. & Rinta-Keturi, I.** *PortNet 2. Preliminary feasibility study* (in Finnish). Helsinki, Finland: Ministry of Transport and Communications, 2005. AINO publications 3/2005. 63 p. AINO publications 3/2005.

[33] **European Maritime Safety Agency.** *SafeSeaNet* (online). 2011 [cited 22.12.2011]. Available: http://www.emsa.europa.eu/operations/maritime-surveillance/safeseanet.html

[34] **European Maritime Safety Agency**. *First semester 2010 SSN report (January-June)* (online). Lisbon, 15 September 2010 [cited 22.12.2011]. 8 p. Available: http://www.emsa.europa.eu/documents/statistical-reports.html

[35] **European Maritime Safety Agency.** *Sharing the European Vessel Traffic Image and Beyond* (online).May 2010 [cited 22.12.2011]. Available: http://www.emsa.europa.eu/operations/maritime-surveillance/safeseanet/download/375/217/23.html

[36] **European Maritime Safety Agency.** *Long Range Identification & Tracking [LRIT] – Integrated Maritime Data Environment* – IMDatE (online). 2011 [cited 22.12.2011]. Available: http://www.emsa.europa.eu/imdate.html

[37] **Kotka Maritime Research Centre.** MIMIC - Work Package 2 [online]. n.d. (cited 1.11.2011). Available: http://www.merikotka.fi/mimic/index.php/en/work-packages/wp2

[38] **Finnish Transport Safety Agency**. *Seaman statistics 2010* (in Finnish and Swedish). Helsinki, Finland: Finnish Transport Safety Agency, 2011. 11 p. Trafin julkaisuja 5/2011.

[39] **Finnish Transport Agency's Hydrographic Office.** *Nautical charts* (online). [cited 04.01.2012]. Available: http://portal.liikennevirasto.fi/sivu/www/e/transport_network/nautical_charts

[40] **Finnish Meteorological Institute**. *Weather monitoring and statistics* (in Finnish) (online). [cited 22.12.2011]. Available: http://ilmatieteenlaitos.fi/saaseuranta-ja-tilastot

[41] **Finnish Meteorological Institute.** *Oceanographic services* (online). [cited 22.12.2011]. Available: http://en.ilmatieteenlaitos.fi/oceanographic-services

[42] **Hänninen, S., Nyman, T., Rytkönen, J., Sonninen, S., Jalonen, R., Palonen, A. & Riska, K.** *Risk factors of the maritime traffic in the Gulf of Finland. Preliminary Study (in Finnish).* VTT products and production, Espoo 2002. BVAL34-021198. 106 p.

[43] **Ministry of the Environment, Finnish Environment Institute & Finnish Meteorological Institute**. *The Baltic Sea portal: Ice situation* (online).

[cited 23.12.2011] Available:
http://www.itameriportaali.fi/en/itamerinyt/en_GB/jaatilanne/

[44] **Suominen, M.** *Statistical modeling of ice induced loads on a ship hull in the Baltic Sea*. Master's thesis. Aalto University School of Engineering, Espoo, 2011. 80 p.

[45] **Ice Atlas,** *Climatological ice atlas for the Baltic Sea, Kattegat, Skagerrak and lake Vänern (1963-1979)*. Swedish Meteorological and Hydrological Institute & Institute of Marine Research, Norrköping, 1982.

[46] **Laiho, A.** *Ship accident analysis 2001-2005* (in Finnish). Helsinki, Finland: Finnish Maritime Administration, 2007. 23 p. Merenkulkulaitoksen julkaisuja 5/2007.

[47] **Kallberg, V.-P.** *Accident statistics for different modes of transport - preliminary survey* (in Finnish). Helsinki, Finland: Finnish Transport Safety Agency, 2011. 37 p. Trafin julkaisuja 1/2011.

[48] **Heiskanen, M.** *Accident analysis 1990-2000, Groundings and collisions with vessel* (in Finnish). Helsinki, Finland: Finnish Maritime Administration, 2001. 66 p. Merenkulkulaitoksen julkaisuja 7/2001.

[49] **Helsinki Commission.** *Report on shipping accidents in the Baltic Sea area for the year 2006*. 2007. 27 p.

[50] **Helsinki Commission.** *HELCOM Map and Data Service* (online). 2011 [cited 9.5.2011]. Available: http://maps.helcom.fi/website/mapservice/index.html

[51] **Helsinki Commission**. *Helcom: Accidents and response* [online]. n.d. Available <URL: http://www.helcom.fi/shipping/accidents/en_GB/accidents/>

[52] **Kujala, P., Hänninen, M., Arola, T. & Ylitalo, J.** *Analysis of the marine traffic safety in the Gulf of Finland*. Reliability Engineering and System Safety, 2009. Vol. 94:8. p. 1349 – 1357.

[53] **Salmi, K.** *Targeting accident prone ships by their behavior and safety culture*. Aalto University School of Science and Technology, Espoo 2010. TKK-AM-14. 44 p. Available: <URL:http://appmech.tkk.fi/julkaisut/TKK-AM-14.pdf/

[54] **European Maritime Casualty Information Platform.** *EMCIP Reporting* (online). 2011 [cited 05.07.2011]. Available:
http://emcipportal.jrc.ec.europa.eu/EMCIP-Reporting.341.0.html

[55] **Correia, P.** European Marine Casualty Information Platform a common EU taxonomy. 5th International Conference on Collision and Grounding of Ships ICCGS 2010, 14.-16. June 2010, Espoo, Finland. TKK-AM-16. p. 13-17.

[56] **Caridis, P.A.** *CASMET. Casualty Analysis Methodology for Maritime Operations*. National Technical University of Athens, Athens 1999. C01.FR.003. 40 p.

[57] **European Maritime Safety Agency.** *Bertranc Final Report* (online). [cited 05.07.2011]. Available: http://www.emsa.europa.eu/implementation-tasks/accident-investigation/item/516.html?cuscat=141

[58] **Finnish Transport Safety Agency & Statistics Finland**. *Maritime traffic accident statistics 2010* (in Finnish) (online) [cited 23.8.2011]. Available: http://www.trafi.fi/filebank/721-Vesiliikenneonnettomuuksien_vuositilasto_2010.pdf0

[59] **IMO.** MSC.255(84). Adoption of the code of the international standards and recommended practices for a safety investigation into a marine casualty or

marine incident (Casualty Investigation Code). International Maritime Organization, London 2008.

[60] **Accident Investigation Board Finland.** *Role and function.* n.d. (online) [cited 13.5.2011]. Available: http://www.onnettomuustutkinta.fi/2579.htm

[61] **Accident Investigation Board Finland**. *Marine.* n.d. (online) [cited 20.5.2011]. Available: http://www.onnettomuustutkinta.fi/en/Etusivu/Tutkintaselostukset/Vesiliikenne

[62] **Marine Accident Investigation Branch.** *Business Plan 2011-2012* (online). [cited 09.12.2011]. Available: http://www.maib.gov.uk/resources/other_useful_documents.cfm

[63] **Marine Accident Investigation Branch.** Annual Report 2001. (online). [cited 09.12.2011]. Available: http://www.maib.gov.uk/publications/annual_reports.cfm

[64] **Marine Accident Investigation Branch.** Home (online). [cited 23.01.2012]. Available: http://www.maib.gov.uk/home/index.cfm

[65] **Information Commissioner's Office.** Freedom of Information Act. (online). [cited 23.01.2012]. Available: http://www.ico.gov.uk/Global/faqs/freedom_of_information_act_for_the_public.aspx

[66] **Marine Accident Investigation Branch.** Annual Report 2009. (online). [cited 09.12.2011]. Available: http://www.maib.gov.uk/publications/annual_reports.cfm

[67] **Gould, K. S., Røed, B. K., Koefoed, V. F., Bridger, R. S. & Moen, B. E.** *Performance-Shaping Factors Associated With Navigation Accidents in the Royal Norwegian Navy*. Military Psychology 18(2006)Suppl., p. 111-129.

[68] **Grabowski, M., Merrick, J. R. W., Harrald, J. R., Mazzuchi, T. A. & van Dorp, J. R.** *Risk Modeling in Distributed, Large-Scale Systems*. IEEE Transaction on Systems, Man, and Cybernetics – Part A: Systems and Humans. 30(2000)6, p. 651-660.

[69] **de la Campa Portela, R.** *Maritime Casualties Analysis as a Tool to Improve Research about Human Factors on Maritime Environment*. Journal of Maritime Research 2(2005)2, p. 3-18.

[70] **Kiuru, H. & Salmi, K.** *Accident analysis; the tool for RISK evaluation*. Helsinki University of Technology, Espoo 2009. TKK-AM-11. 37 p. Available: http://appmech.tkk.fi/fi/julkaisut/TKK-AM-11.pdf

[71] **Hopper, A.G., Dean, A.J.** *Safety in shipping – learning from experience*. Safety of Science, Vol 15, Issues 4-6, Nov 1992, pp. 249-271

[72] **Pillay, A., Wang, J.** *Technology and Safety of Marine Systems*. Elsevier Ocean Engineering Series, Vol 7, 2003., pp. 9-28

[73] **Ellis, J.** *Analysis of accidents and incidents occuring during transport of packaged dangerous goods by sea*. Safety Science, Article in Press, Corrected Proof, Available online 27 April 2011

[74] **Antão, P., Guedos Soares, C.** *Causal factors in accidents of high-speed craft and conventional ocean-going vessels*. Reliability Engineering&System Safety, Vol 93, Issue 9, Sept 2008., pp. 1292-1304

[75] **Hetherington, C., Flin, R., Mearns, K.** *Safety in shipping: The human element*. Journal of Safety Research, Vol 37, Issue 4, 2006, pp. 401-411

[76] **Reason, J.** *Human Error*. Cambridge University Press, New York 1990. 302 p.

[77] **Tan, A.H.** *Text mining: The state of the art and the challenges*. Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases. 1999. pp. 65-70.

[78] **Grech, M., Horberry, T. & Smith, A.** *Human Error in Maritime Operations: Analyses of Accident Reports Using the Leximancer Tool*. Human Factors and Ergonomics Society 46th Annual Meeting (HFES), Baltimore 30.9.-4.10.2002. Human Factors and Ergonomics Society. p. 1718-1722.

[79] **Harrald, J. R., Mazzuchi, T. A., Spahn, J., Van Dorp, R., Merrick, J., Shrestha, S. & Grabowski, M.** *Using system simulation to model the impact of human error in a maritime system*. Safety Science 30(1998)1-2, p. 235-247.

[80] **ForeSea**. Purpose & Joining. (online) [cited 29.6.2011]. Available: http://www.foresea.org/about.aspx

[81] **Storgård, J., Erdogan, I. & Tapaninen, U.** *Incident reporting in shipping - Experiences and best practices for the Baltic Sea.* Publications from the Centre for Maritime Studies, University of Turku A59/2012.

[82] **Interview with O. Bråfelt, Insjö/ForeSea administrator**. *IMISS Conference, 1-2 September 2011, Espoo, Finland*

[83] **Ishikawa, K.** *Introduction to Quality Control*. Productivity Press, Tokyo 1990. 435 p.

[84] **Köhler, F.** *Barriers to Near-miss Reporting in the Maritime Domain*. Chalmers tekniska högskola, 2010. Master thesis

[85] **Zachau, J**. *Near-Misses and Accidents in Proactive Safety Work*, 2008. Swedish Maritime Administration report series B 2008-6

[86] **Lappalainen, J., Vepsäläinen, A., Salmi, K., Tapaninen, U.** *Incident reporting in Finnish shipping companies*. WMU Journal of Maritime Affairs, Vol 10 (2), 2011, pp. 167-181

[87] **Erdoğan, I.** *Best Practices in near-miss reporting. The role of near-miss reporting in creating and enhancing the safety culture*. Chalmers University of Technology, Sweden, 2011. Master thesis

[88] **The Finnish Transport Agency**. *Vessel Traffic Service*. 10.8.2011 (online) [cited 20.9.2011]. Available: http://portal.liikennevirasto.fi/sivu/www/e/professionals/vts/vts

[89] **Ministry of Justice.** *Vessel Traffic Service Act 5.8.2005/623* (in Finnish) (online) [cited 13.2.2012]. Available: http://www.finlex.fi/fi/laki/ajantasa/2005/20050623

[90] **The Finnish Transport Agency**. *GOFREP*. 10.8.2011 (online) [cited 18.10.2011]. Available: http://portal.liikennevirasto.fi/sivu/www/e/professionals/vts/gofrep

[91] **Westerlund, K.** *The Risk Reducing Effect of VTS in Finnish Waters*. EfficienSea Deliverable No. D_WP6_5_01. Finnish Transport Agency 2011. 49 p.

[92] **IMO.** *Port State Control*. n.d. (online) [cited 20.10.2011]. Available: http://www5.imo.org/SharePoint/mainframe.asp?topic_id=159

[93] **Paris MoU.** *Memorandum of Understanding* 23.6.2011 (online) [cited 20.10.2011]. Available: http://www.parismou.org/Organization/2010.12.27/Memorandum_of_Understanding.htm

[94] **Paris MoU**. *Paris Memorandum of Understanding on Port State control Including 33rd Amendment*. Paris MoU 2011.

[95] **Paris MoU**. *Paris Memorandum of Understanding on Port State control Including 29th Amendment*. Paris MoU 2007.

[96] **Paris MoU.** *Inspection types*. 1.9.2011 (online) [cited 20.10.2011]. Available: http://www.parismou.org/Inspection_efforts/Inspections/Inspection_types/

[97] **Paris MoU.** *Inspection database search*. 15.7.2011 (online) [cited 24.10.2011]. Available: http://www.parismou.org/Inspection_efforts/Inspections/Inspection_database_search/

[98] **European Maritime Safety Agency**. *THETIS* (online). 2010 [cited22.12.2011]. Available: https://portal.emsa.europa.eu/web/thetis/home

[99] **Mejia, Jr, M.Q., Cariou, P. & Wolff, F.C.** *Vessels at risk and the effectiveness of Port State Control inspections*. Working paper, 2010.

[100] **Knapp, S. & Franses, P.H.** *Econometric analysis on the effect of port state control inspections on the probability of casualty: Can targeting of substandard ships for inspections be improved?* Marine Policy 31(2007)4, pp 550-563.

[101] **Li, K.X., Yin, J., Yang, Z. & Wang, J.** *The Effect of Shipowners' Effort in Vessels Accident: A Bayesian Network Approach*. The Fourth International Forum on Shipping, Ports and Airports (IFSPA). Chengdu, Sichuan, China, 2010.

[102] **Räisänen, P.** *Influence of Top Management on Safety Culture*. In: Heijari, J. & Tapaninen U. (eds.). Efficiency of the ISM Code in Finnish Shipping Companies. Publications from the Centre for Maritime Studies A 52, Turku 2010. 57 p.

[103] **Assuranceforeningen Skuld.** *An introduction to P&I Insurance for Mariners*. (online) [cited 15.12.2011] Available: http://www.skuld.com/Publications/Introduction-to-PI/

[104] **Alandia Marine.** Home & Statuatory Accident Insurance. (online) [cited 28.10.2011] Available: http://www.alandia.com/avdelning_text.con?iPage=271&m=1171

[105] **UK P&I Club.** *Loss Prevention – Risk Management Advice*. (online) [cited 14.12.2011] Available: http://www.ukpandi.com/loss-prevention/risk-management-advice/

[106] **UK P&I Club.** *Loss Prevention- Claims Analysis Advice* (online) [cited 14.12.2011] Available: http://www.ukpandi.com/loss-prevention/claims-analysis-advice/

[107] **UK P&I Club.** *Member Services- Ship Inspection* (online) [cited 14.12.2011] Available: http://www.ukpandi.com/fileadmin/uploads/uk-pi/LP%20Documents/Ship_Inspection_Report.pdf

[108] **The Swedish Club.** *Publications – Standard forms - Surveys.* (online) [cited 15.12.2011] Available: http://www.swedishclub.com/main.php?mcid=1&mid=26&pid=17153&tid=17153

[109] **The Swedish Club.** *Instructions for surveyors.* (online) [cited 22.3.2012]. Available: http://www.swedishclub.com/main.php?mcid=1&mid=180&pid=17153&tid=17153

[110] **UK P&I Club.** *Loss Prevention – PEME.* (online) [cited 14.12.2011] Available: http://www.ukpandi.com/loss-prevention/peme/

[111] **Det Norske Veritas.** *Casualty Information* (online) [cited 13.2.2012]. Available: http://exchange.dnv.com/CasualtyInformation/

[112] **IHS.** *IHS Maritime Intelligence & Publications: Fairplay* (online [cited 13.2.2012]. Available: http://www.ihs.com/products/maritime-information/index.aspx

[113] **Psarros, G., Skjong, R., Strandmyr Eide, R.** *Under-reporting of maritime accidents.* Accident Analysis & Prevention, Vol 42, Issue 2, March 2010, pp. 619-625

[114] **Deverill, M., Mclure, J.** *Developing a GIS for global vessel movements.* CARIS conference 2005., Halifax, Nova Scotia. (online) [cited 8.12.2011] Available: http://www.marinebiodiversity.ca/COINAtlantic/reference/applications/LloydsMIU.ppt

[115] **Toffoli, A., Lefèvre, J.M, Bitner-Gregersen, E., Monbaliu, J.** *Towards the identification of warning criteria: Analysis of a ship accident database.* Applied Ocean Research, Vol 27, Issue 6, Dec 2005, pp. 281-291

[116] **Papanikolau, A., Eliopoulou, E., Mikelis, N., Aksu, S., Delautre, S.** *Casualty analysis of tankers.* Learning from Marine Incidents III, London, UK, 2006

[117] **Devanney, J**. *Uses and Abuses of Ship Casualty Data*, 2009. Center for Tankship Excellence

[118] **IHS.** *IHS Fairplay Global Maritime Statistics & Forecasting reports* (online) [cited 24.2.2012]. Available: http://www.ihs.com/products/maritime-information/statistics-forecasts/

[119] **Hassel, M., et al**. *Underreporting of maritime accidents to vessel accident databases.* Accident Analysis & Prevention 46(2011)6, pp. 2053-2063.

[120] **IHS.** *IHS Fairplay World Casualty Statistics* (online) [cited 23.3.2012]. Available: http://www.ihs.com/products/maritime-information/statistics-forecasts/world-casualties.aspx

[121] **Class Direct Live – Lloyd's Register**. *About ClassDirect Live* (online) [cited 11.1.2012] Available: https://www.cdlive.lr.org/mainmenu.asp?showPage=What_is.htm

[122] **Pomeroy, R.V**. *Reliability and risk assessment.* Lloyds Register of Shipping. Presentation at technical committee meeting, 28th April 1982

[123] **French Ministry for Transport -DAM/SI-Saint-Malo**. *Equasis.* 2011 (online) [cited 28.10.2011] Available: http://www.equasis.org/

[124] **French Ministry for Transport -DAM/SI-Saint-Malo**. *Equasis – Main concers.* 2011 (online) [cited 28.10.2011] Available: http://www.equasis.org/EquasisWeb/public/About?fs=HomePage

[125] **French Ministry for Transport -DAM/SI-Saint-Malo**. *Equasis – About*. Equasis 2011 (online) [cited 28.10.2011] Available: http://www.equasis.org/EquasisWeb/public/About?fs=About&P_ABOUT=ProvidersNotConnected.txt

[126] **Psarros, G., Skjong, R., Strandmyr Eide, R**. *Under-reporting of maritime accidents*. Accident Analysis & Prevention, Vol 42, Issue 2, March 2010, pp. 619-625

[127] **Marsland, S.** *Machine learning: an algorithmic perspective*. Chapman & Hall/CRC, USA, 2009. 406 p.

[128] **Hair, J..F, Anderson, R.E., Tatham, R.L. & Black, W.C.** *Multivariate Data Analysis*. Fifth edition. Prentice-Hall, New Jersey, 1998. 730 p.

[129] **Dempster, A.P., Laird, N.M. & Rubin, D.B.** *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society, Series B 39(1977)1, pp. 1-38.

[130] **Sivia, D. S. & Skilling, J.** *Data Analysis : A Bayesian Tutorial*. Oxford University Press, Oxford, 2006. 259 p.

[131] **Tripathi, A.** *Data fusion and matching by maximizing statistical dependencies*. Doctoral dissertation. University of Helsinki. Helsinki, Helsinki university Print, 2011. 198 p. Available: http://urn.fi/URN:ISBN:978-952-10-6750-1

[132] **Blum, A. & Mitchell, T.** *Combining labeled and unlabeled data with co-training*. In: Proceedings of the eleventh annual conference on Computational learning theory,ACM, 1998. pp. 92-100.

[133] **Ben-David, S., Blitzer, J., Crammer, K. & Pereira, F.** *Analysis of Representations for Domain Adaptation*. In: Advances in neural information processing systems 19, MIT Press, 2007. pp. 137-144.

[134] **Druzdzel, M.J. & Díez, F.J.** *Combining knowledge from different sources in causal probabilistic models*. The Journal of Machine Learning Research 4(2003), pp.295-316

Appendix A: Data fields in DAMA accident database

Field	Format	Field	Format
Case number	number	Country	text
Ship name	text	Waters	cat
Home port	text	Voyage phase	cat
Nationality	text	Working ac.	cat
Type of ship	cat	wind direction	cat
Constuction year	number	wind force	cat
Renovation year	number	sea	cat
Material	cat	visibility	cat
GRT	number	light	cat
DWT	number	cargo	cat
Length	number	Pilot onboard	y/n
Classification society	text	2. ship name	text
Year	number	2. ship nation	text
Month	number	Loss/ damage severity	cat
Day	number	evacuated	y/n
Time of event	number	Hull damage	y/n
Day of the week	number	Hull damage severity	cat
Event #1	cat	Damage length	Number
Event #2	cat	Damage width	Number
Event #3	cat	Damage depth	Number
Cause #1	cat	Hull damage location y	cat
Cause #2	cat	Hull damage location z	cat
Cause #3	cat	Hull damage location x	cat
Cause #4	cat	Death persons	Number
Departure port	text	Injured persons	Number
Destination port	text	Oil pollution	Number
Latitude	number	Bridge manning	Free text
Longitude	number	Damages	Free text

Appendix B: Categorization of accident causes in DAMA

	External factors
A01	Storm, nature disaster etc.
A02	Drift caused by wind, current etc. or other maneuvering challenges
A03	Collision with a floating object not detected or avoided on time
A04	Failures in aids to navigation or external safety equipment
A05	Error in navigation chart or publication
A06	Technical fault of the other ship (incl. Tug)
A07	Another ship acting wrongly
A08	Technical fault in loading, unloading or bunkering equipment. Faults in pier struc-
A09	Faults in using loading, unloading or bunkering equipment. Faults in using port or
A10	Blow-up or other external factors in oil drilling
A11	Difficult ice conditions
A12	Icing
	Ship structure and layout
B01	Insufficient structural strength
B02	Impaired structural strength due to welding, corrosion etc.
B03	Loss of stability due to ship structure
B04	Inadequate maneuvering capabilities
B05	Equipment layout / placement in the machinery room caused a leak or a fire hazard
B06	Bad placement or layout of cargo or storage area
B07	Bad placement or layout of other areas than the bridge
B08	An area difficult to access for cleaning, maintenance or inspection
B09	Other factors related to ship structure or maintenance
	Technical faults in ship equipment
C01	Technical fault in navigation equipment
C02	Technical fault in steering equipment
C03	Technical fault in propulsion system
C04	Technical fault in auxiliary system
C05	Technical fault in anchoring equipment / deck machinery
C06	Technical fault in control/remote control/automatic control/warning system
C07	Technical fault in cargo handling equipment
C08	Technical fault in backup systems/ inert gas system/halon system
C09	Technical fault in drilling equipment
C10	Other technical fault

	Factors related to equipment usage and placement
D01	Impractical bridge design, equipment missing or wrongly placed
D02	Poor user interface design or placement
D03	Placement of an equipment not suitable for operating
D04	Unsuitable/poor/worn equipment, equipment difficult to use
D05	Other equipment design /operation factors, man-machine interface problems
	Cargo, cargo and fuel handling and related safety equipment
E01	Autoignition of cargo/fuel
E02	Inert gas system or other fire/explosion prevention system missing
E03	Irregular stability (wrongly placed cargo, missing ballast etc.)
E04	Cargo not properly secured
E05	Liquid cargo leak (barrels, containers, tanks etc.)
E06	Leaks in cargo or fuel pipes / hoses
E07	Other factors related to cargo or fuel
	Communication, organizing, instructions and routines
F01	Missing/incomplete general instructions
F02	General procedures not known/inadequately trained
F03	Missing / incomplete safety instructions
F04	Safety instructions known but not followed
F05	Welding safety instructions not followed
F06	Welding lead to fire despite following safety instructions
F07	Emergency equipment testing and test instructions not followed
F08	Personal protective equipment not used
F09	Inadequate level of organizing/instructions/competence
F10	Inspection/maintenance instructions not followed
F11	Stability not known / no accepted stability calculations
F12	Inappropriate management style, people problems etc.
F13	Undermanning (missing helmsman, lookout etc.)
F14	Unclear task responsibilities
F15	Bridge routines not defined or defined poorly
F16	Bridge routines not followed
F17	Nautical charts/other publications outdated
F18	Mistakes in cooperation with a tug, land organization etc.
F19	Other factors related to organization, safety rules, routines or communication
	Persons, situation assessment, actions
G01	Unqualified for the task (education, degrees etc.)
G02	Inexperience (work experience, water area familiarization, equipment usage etc.)
G03	Poorly planned task/action (cargo, night navigation, route plan, anchoring etc.)
G04	Available means for receiving a warning inadequately utilized
G05	Alternative navigational systems not used. Lights, buoys etc. Wrongly assessed
G06	Available navigational aids or publications not used
G07	Position not fixed correctly
G08	Misunderstanding of the other vessel's movement or intentions
G09	Misunderstanding of own vessel's movement (wind, current etc.)
G10	Tried to perform the task in unfavorable conditions
G11	Did not stay at the correct side of the waterway/water area

G12	Situational speed too high
G13	Sickness, fatigue, excessive workload etc.
G14	Fell asleep on watch
G15	Alcohol or other intoxicant usage
G16	Other human failures

Aalto-ST 11/2012

The report describes various data sources and their utilization in quantitative marine traffic accident modeling. The primary interest is on the data sources that cover the Gulf of Finland and which could be useful in modeling human and organizational causes in ship collisions and groundings. The sources are analyzed considering the nature, quantity, quality and availability of the data, and if the data is feasible to quantitative accident modeling. It is found that the data sources differ in the scope and purpose and they all have their strengths and weaknesses. The existing sources are not perfect and using any of them as the only source of input to a quantitative model seems risky. The report is a part of the research project Competitive Advantage by Safety (CAFE). CAFE is funded by the European Regional Development Fund, the City of Kotka, Finnish Shipowners' Association, Kotka Maritime Research Centre corporate group: Aker Arctic Technology Inc., the Port of HaminaKotka, the Port of Helsinki, Kristina Cruises Ltd, and Meriaura Ltd., and the project partners.

Leverage from the EU
2007–2013

European Union
European Regional Development Fund

CAFE } COMPETITIVE ADVANTAGE BY SAFETY

ISBN 978-952-60-4599-3 (pdf)
ISSN-L 1799-4896
ISSN 1799-490X (pdf)

Aalto University
School of Engineering
Department of Applied Mechanics
www.aalto.fi