Development of a maritime safety management database using relational database approach

Siyu Xu and Hao Hu*

Department of Transportation, Shipping and Logistics
and
State Key Laboratory of Ocean Engineering,
Shanghai Jiao Tong University,
800 Dongchuan Road, Shanghai, 200240, China
Email: xusiyu@sjtu.edu.cn
Email: hhu@sjtu.edu.cn
*Corresponding author

Abstract: Many types of maritime incident databases have been established that allow people to learn from past incidents and develop corresponding mitigation measures. However, our investigation of multiple international and national databases shows that most existing databases only record basic information regarding incidents in a single table. Lots of useful information is not included in the database (i.e., limited extension of the database). Meanwhile, some basic information is recorded tautologically (i.e., data redundancy). In this paper, two widely used databases are taken as examples, the Global Integrated Shipping Information System and the Lloyd's List Intelligence, to explain these common problems of existing databases. To overcome these limitations and improve the efficiency of data maintenance, this paper develops a relational maritime safety management database. The entity-relationship model is first used to depict the inter-related semantic information surrounding maritime incidents, and a relational database model is subsequently formed. Microsoft Access is employed to implement the proposed database, and a database application is also designed to demonstrate the utility of the database. Our preliminary study shows that the proposed database is implementable and has potential usage for both industry and academic research.

Keywords: maritime incident; data maintenance; Global Integrated Shipping Information System; GISIS; Lloyd's List Intelligence; LLI; relational database; entity-relationship model; Microsoft Access.

Reference to this paper should be made as follows: Xu, S. and Hu, H. (2019) 'Development of a maritime safety management database using relational database approach', *Int. J. Shipping and Transport Logistics*, Vol. 11, No. 4, pp.334–353.

Biographical notes: Siyu Xu is a Research Associate at the Departmet of Transportation, Shipping and Logistics, Shanghai Jiao Tong University. Her research interests are maritime safety and application of IT to transportation.

Hao Hu is a Professor at the Department of Transportation, Shipping and Logistics and State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University. His research areas include water transport, project management and transportation planning.

This paper is a revised and expanded version of a paper entitled 'Development of a maritime incident database using a relational database management system' presented at The 96thAnnual Meeting of Transportation Research Board, Washington, DC, 8–12 January 2017.

1 Introduction

Characterised by its large capacity and economies of scale, maritime transport is a primary mode of international transport and accounts for approximately 90% of global trade volume (Hong, 2012). Even with the advancement of maritime technology and management, maritime transport is still recognised as a high-risk mode of transport. Every year, many incidents take place at sea, often with serious consequences for people, ships or the environment. In fact, people today are less tolerant of maritime incidents than ever before. For example, the Deepwater Horizon oil spill in the Gulf of Mexico in 2010 put more than 400 species that live in the Gulf Island and marshlands at risk. More recently, the sinking of the Sewol off the southern South Korea shore in 2014 resulted in a death toll of 304 lives, and the capsizal of the Eastern Star on the Yangtze in 2015 incurred 442 fatalities. These incidents have all raised great concerns in the public. Maritime safety has become a priority for government authorities, and thus, maritime incident analysis has become a significant area of study (IMO, 2011). Many efforts have been made by researchers to investigate the causal mechanisms surrounding maritime incidents and the remedial measures that should be applied to prevent these incidents in the future (Li et al., 2013; Uğurlu et al., 2015; Edwards and Kauffman, 2016; Zhang et al., 2016a). As compared to experts' judgments and simulation experiments, data from real incident cases are the most important basis for maritime incident analysis performed by researchers.

The International Maritime Organization (IMO) has issued a series of international standards and recommended practices for a safety investigation into maritime casualties to collect incident data (IMO, 2008). Generally, maritime incident data include primary and secondary data sources (Mazaheri et al., 2015). The primary data are provided by those directly involved in an incident, such as the crew and passengers, or by monitoring equipment installed onboard a ship (Tsai, 2016), such as the voyage data recorder (VDR), the automatic identification system (AIS), vessel traffic services (VTS), very high frequency (VHF) communication systems and the electronic chart display and information system (ECDIS). Secondary source data are the data processed from the primary sources through multi-information fusion and managed in various databases. These databases underlie maritime incident analysis, and they can generally be classified into three types: public databases maintained by the IMO, commercial databases maintained by classification societies and national databases maintained by government agencies (Huang et al., 2013).

A deep investigation of numerous international and national maritime incident databases has been conducted; therein we take two most well-known databases as examples, the Global Integrated Shipping Information System (GISIS) and the Lloyd's List Intelligence (LLI), to describe the common issues haunting the existing databases. To facilitate direct reporting of incident information by member states, IMO (2009) launched the GISIS in 2005. ‘marine casualties and incidents (MCI)’ is one of its 22 modules. In the MCI module, an incident record is composed of four parts: an incident summary, reporting forms, investigation reports and analyses. Notably, as defined by the IMO circulars MSC-MEPC.3/Circ.3, reporting forms include ten annexes and focus on collecting information regarding different aspects of an incident. In the GISIS, basic or advanced searches to find a particular type of incidents based on certain criteria are allowed, and the search results can be visualised on a map. Unlike GISIS, which is a public resource, the LLI, formerly known as the Lloyd’s Marine Intelligence Unit (LMIU) is a private commercial database maintained by Lloyd’s Register. In this database, the insurance channel and the law and regulation channel both include a module called ‘Casualties’ that collects all types of worldwide maritime incidents, with varying severities, for merchant ships over 100 gross tons. This module contains detailed records on serious casualties and displays the distribution of casualties according to the casualty type, the ship type and the gross tonnage for the last six months. Each incident can be fully described with 33 fields, and by clicking the name of a ship or one of its ownerships, one can reach other interfaces that present more details concerned. Table 1 summarises the characteristics of the GISIS and the LLI.

Table 1 Summary of characteristics of the GISIS and the LLI

Items	*GISIS*	*LLI*
Data sources	Member states directly report incident information to the GISIS (mandatory/passively)	Collect from multiple suppliers, such as port agents and rescue centres, and combine with terrestrial and satellite AIS tracking (actively)
Reliability/accuracy	Average	Good
Comprehensiveness	Average Cover over 9,800 records dated from 1973	Good Cover over 86,400 records dated from 1965
Update frequency	Average	Good
Availability	Public	Private
Recording principle	Record incidents from the perspective of the incident	Record incidents from the perspective of the ship
Definition of seriousness	Four categories: • Very serious casualties • Serious casualties • Less serious casualties • Marine incidents	Two categories: ‘seriousness’ involves only what may affect the ship and not the ship’s crew or passengers
Causes of incidents	Analyse multiple aspects, including people, ships, cargoes and the environment	Record at most three causes, which actually are confused with casualty type
Consequences of incidents	Qualitative: consequences to ships Quantitative: consequences to people and the environment	Qualitative: fatality indicator, pollution indicator and seriousness indicator Quantitative: consequences to people

Like other structured databases, the GISIS and the LLI both report maritime casualties in a set of data fields that cover worldwide incidents. In both databases, data fields concerning the incident and the ship are encompassed in two separate tables, which are correlated by the ship's unique identifier, i.e., the IMO number or the ship name. This reveals that both the GISIS and the LLI have applied some principles of relational database technology where data records are distinguished while relate to each other. Nevertheless, the GISIS and the LLI only take advantage of preliminary knowledge about the relational database. Much information is isolated and tautologically recorded. The GISIS records casualties from the perspective of the incident but cannot expose the incident history of each ship intuitively. Contrastingly, the LLI lists the incident history of each ship but fails to uncover other possible ships involved in the same incident. The problems indicate that these databases can be further improved in the organisation of data. In this respect, the prevailing relational database technology is a promising tool. These problems could be solved if with the systematical use of the relational database approach.

Some maritime incident databases may utilise the basic ideas of a relational database, but they are not in a systematic way, which may result in following defects:

1 Inefficient maintenance. Existing databases were not designed according to the relational database model. Information surrounding an incident is usually recorded in a single table, where much basic information is repeatedly recorded (i.e., data redundancy), while more relevant information for further reference is not available (i.e., limited extension of the database). Once the database needs to be updated, the efficiency of data maintenance could be undesirable, because the same piece of information needs to be modified in multiple places in one table instead of being updated synchronously.

2 Inefficient query. Tables recording different incidents are isolated from one another, making it difficult to form a network of information to be used in knowledge management. In other words, it is difficult to discover which ships are involved in the same incident, and meanwhile what is the incident history of each ship involved.

3 Inefficient analysis. Most incident databases are merely used to retrieve incident records; however, these databases lack the type of flexible analysis functions required by an intelligent database. Considering that databases will be updated over time, the traditional method that exporting incident data from the databases and analysing them externally makes it difficult to achieve real-time analytical results. Therefore, it is necessary to be able to conduct analysis automatically by the database itself so that the latest data can be fully used.

The purpose of this paper is to summarise the properties and capabilities of the existing maritime incident databases and to establish an improved relational maritime safety management database called the MSMDB using a systematic approach to improve the data maintenance, query and analysis. The entity-relationship (ER) model is employed to depict the inter-related semantic information surrounding maritime incidents and to provide a concise visualisation of concerned entities and their relationships. Full use of relational database technology is employed to accomplish the database design. With this design, incident information can be organised in a logical way that strengthens data independence while maintaining data correlation. The preliminary study of the proposed database finds that such a design can be used to reflect the real world more clearly and comprehensively and that such a design makes it possible to implement an intelligent database for use in knowledge management.

The remainder of the paper is organised as follows: Section 2 reviews the literature related to the development of maritime incident databases and their typical applications, as well as the limitations of existing databases. Next, the ER model of the proposed relational maritime incident database is developed and converted to the relational database model in Section 3. Section 4 presents an implementation of the database design using the software program Microsoft Access and imports empirical data to demonstrate a sample database application. Finally, conclusions and recommendations for future work are provided in Section 5.

2 Literature review

Incident databases are the most important sources of information for maritime incident analysis, and the goal of analysis can be better achieved by effect of scale of databases. In addition to the databases that exist at the global level, e.g., the GISIS and the LLI, more databases specific to a given country or region are also available. Maritime authorities, such as the National Transportation Safety Board in the US and the Marine Accident Investigation Branch in the UK, investigate maritime incidents involving ships sailing under their flags worldwide and any ship in their territorial waters. These organisations issue investigation reports that comprise the national maritime incident databases. Other incident databases include the Finnish DAMA database (Asbjørnslett et al., 2010), the Helsinki Commission (HELCOM) Database that gathers statistics for Baltic Sea incidents, the European Marine Casualty Information Platform (EMCIP) database, etc. (Ladan and Hanninen, 2012). Moreover, existing maritime incident databases are constantly improved. For example, the IMO (2014) made amendments to MSC-MEPC.3/Circ.3 to revise the online reporting procedures of the GISIS in 2013. Some advanced information technologies, such as the Geographic Information System (GIS), are also integrated into maritime incident databases to develop management information systems for safer maritime transport (Martin et al., 2004).

Applications of maritime incident databases are multifaceted, typically focusing on data processing, statistics and risk analysis surrounding maritime incidents. When people utilise incident databases, preliminary data processing such as data fusion or text mining is often required. Li et al. (2014a) applied the Dempster-Shafer theory (DST), a generalisation of the Bayesian theory, to combine evidence from different databases into a comprehensive result. Text-mining techniques were also utilised to replace human efforts in extracting the key information from investigation reports that are normally in text format (Mazaheri et al., 2015; Tirunagari et al., 2012). Apart from data processing, statistical methods are often used to summarise risk factors and to reveal potential mechanisms. Zhang et al. (2016b) explored ship incident frequency according to different incident types and consequences, and a series of statistical examinations were conducted to discover relationships between contributory factors and incident consequences. Using maritime piracy data from International Maritime Bureau (IMB), a division of

International Chamber of Commerce (ICC), a function based on binary choice model was developed to quantify how ship attributes and pirate characteristics determine the rate of success and degree of violence of piracy attacks (Wong and Yip, 2012). Based on incident databases, risk profiles can also be generated. Regarding the risk associated with specific ships, Li et al. (2014b) proposed a quantitative safety index for each worldwide ship using the binary logistic regression method. Regarding the risk of ocean regions, Huang et al. (2013) visualised the spatial distribution of worldwide maritime incidents in the GISIS between 2002 and 2011 and carried out hot spot analysis and buffer analysis on a GIS platform.

Although they are constantly improving, maritime incident databases still suffer from problems, such as the underreporting of incidents (Psarros et al., 2010), the lack of a unified recording standard, etc. In various studies, common methods used to estimate the actual number of maritime incidents include the conditional probability method, the capture-recapture method, the best case scenario method and the up-scaling of subset data method. It is estimated that approximately 50% of all maritime incidents that actually occur are underreported, and users of incident databases are suggested to assume a certain degree of underreporting and adjust their analyses accordingly (Hassel et al., 2011). As for the recording standards for incident data, different incident databases may suffer from a lack of the uniformity in recording methods and field taxonomies (Ladan and Hanninen, 2012). Far more importantly, the results of the analysis will be handicapped if the data are incomplete. Devanney (2008) outlined the problems with ship casualty data, e.g., the censored data, the hidden data that cannot be audited and confusing cause and effect relationships within data, and a public database of tanker casualties was developed as an example to demonstrate what a reasonable database should be like. Additionally, Sun et al. (2007) developed a management information system for maritime incidents based on SQL Server 6.0 and Delphi 6.0. The study described what types of information should be included and what functions were needed by system users, and how these requirements can be realised by computer programmes. Similarly, in order to facilitate an efficient maritime safety management for Taiwan, a web-based three-tier maritime casualty database structure is devised (Chen and Su, 2005). These studies only provided a function architecture to fill in the absence of the database necessary for maritime safety management but without reflecting the essence and superiority of the relational database approach in the development of the system. The standardisation of incident reporting with respect to road transport also has a certain relevance to maritime incident reporting. To standardise traffic crash location records, a five-element crash location description method based on the linear referencing system was defined and taught to practitioners for use in field operations (Zhang et al., 2012).

In general, as high-quality data are the backbone of scientific researches (Dobler, 1994; Wan, 2015), many types of maritime incident databases are becoming available and greatly stimulating evidence-based research. Whereas the deficiencies of databases are likely to negatively affect the results of applications that use them, great efforts have been made to improve maritime incident databases from many different perspectives. Almost all the existing academic studies focus on eliminating the underreporting of incidents, standardising recording methods, etc., but many overlook the necessity of improving the storage structure of incident information. Actually, creating a reasonably structured database model based on relational database technology can make maritime incident databases fundamentally more practical.

3 Design of the relational maritime safety management database

The relational database, first defined by Codd (1970) has been the predominant type of database for many years. It is based on the relational model, whose central data description construct is the relation, i.e., the table. Data are stored in multiple tables, where records are distinguished from each other through the primary key and relate to the corresponding records in other tables through the foreign keys. The relational database has good data independence and low data redundancy and manages data in a systematic and interrelated manner. A relational database design normally includes six steps: requirements analysis, conceptual database design, logical database design, physical database design, implementation and operational maintenance. In particular, the conceptual and logical database designs incorporated in this section are critical to the overall design of a system.

3.1 Conceptual database design with the ER model

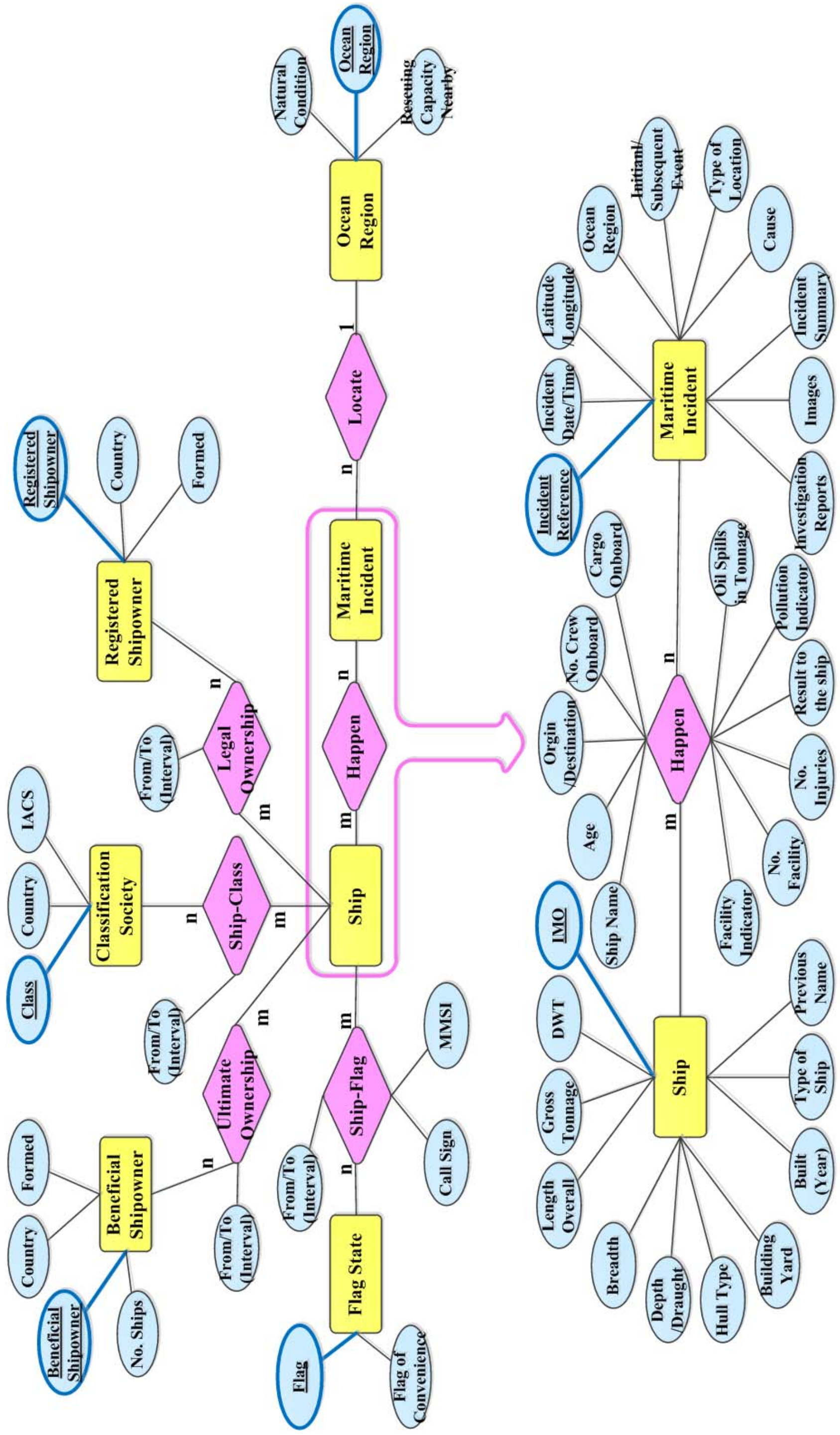
The proposed relational maritime incident database aims not only to record maritime incident information comprehensively but also to improve data independence and logic while maintaining the correlations among data to facilitate the synchronous maintenance, more considerate query and statistics gathering of incident data. These objectives will guide the whole design process. Conceptual database design is the initial design phase in which a well-known semantic model, called the ER model, is used to describe and abstract the data gathered from the real world (Chen, 1976). In software engineering, the ER model is an invaluable abstract data model used to define a data structure that can be implemented in a database, typically a relational database.

An ER model consists of entity types that classify the things of interest and identifies relationships between these entity types. In grammatical terms, entities are the equivalent of grammatical nouns, e.g., ships, incidents, or ocean regions. An entity can be defined by means of its properties, called attributes. Relationships are the equivalent of verbs or associations, e.g., the act of happening, or being a member of a group. A relationship can be defined based on the number of entities related to it, known as the cardinality. As shown in Figure 1, the ER model, using the Chen notation, pictorially presents the entities relevant to a maritime incident and how they relate to one another. According to Chen's diagramming technique, the ER diagram consists of entities, relationships and attributes, which are represented by yellow rectangles, pink diamonds and blue ovals, respectively.

3.1.1 Entities – yellow rectangles

An entity is an object in the real world that is distinguishable from other objects, either a physical object or a concept (Ramakrishnan and Gehrke, 2000). In the context of a maritime incident, the maritime incident and the ships involved are the most important entities, and they are emphasised with a box and analysed in more detail at the bottom of the diagram.

Figure 1 The ER diagram of the relational maritime incident database (see online version for colours)



One certainly cannot claim that the flag is a possible cause of a maritime incident, but it is one quite important risk factor. The flag may be chosen as a 'proxy' for other variables that cannot be easily measured, such as crew composition, crew training and others. Similarly, the classification society and the country of ownership of a ship may also be regarded as 'proxy' variables, and they have proven to be statistically significant factors affecting maritime incident frequencies (Li et al., 2014c; Psaraftis et al., 1998a). Therefore, the flag state, classification society, registered ownership and beneficial ownership, i.e., the real ownership, are considered as the primary entities most closely related to the ship in terms of maritime safety. Additionally, the ocean region is identified as the major entity related to the maritime incident itself because maritime administrations are quite concerned with the conditions found at incident sites, such as natural conditions and traffic conditions. In accordance with the *World Casualty Statistics* by Lloyd's Register, the ocean regions of the world are divided into 31 zones. Due to their unique geographical environments, different regions will have different effects on the safety of a ship (Yin, 2013). Similarly, LLI classifies the locations of casualties into 34 regions, and this taxonomy will be employed in this paper to associate maritime incidents with spatial information that can be used in further reference.

3.1.2 Relationships – pink diamonds

With the relevant entities identified, the next step is to recognise the logical connection among these entities, i.e., the relationships. The relationship that an entity has with another entity is usually realised by the primary key and the foreign key. When a primary key migrates to another table, it becomes a foreign key in the other table. With respect to maritime incidents, it is not difficult to determine relationships among entities; however, a heavy emphasis is placed on determining the cardinality of such relationships. Because the relationships in the established ER model are all binary, they can be referred to as being one-to-one (1:1), one-to-many (1:n) or many-to-many (m:n) and labelled in the ER diagram.

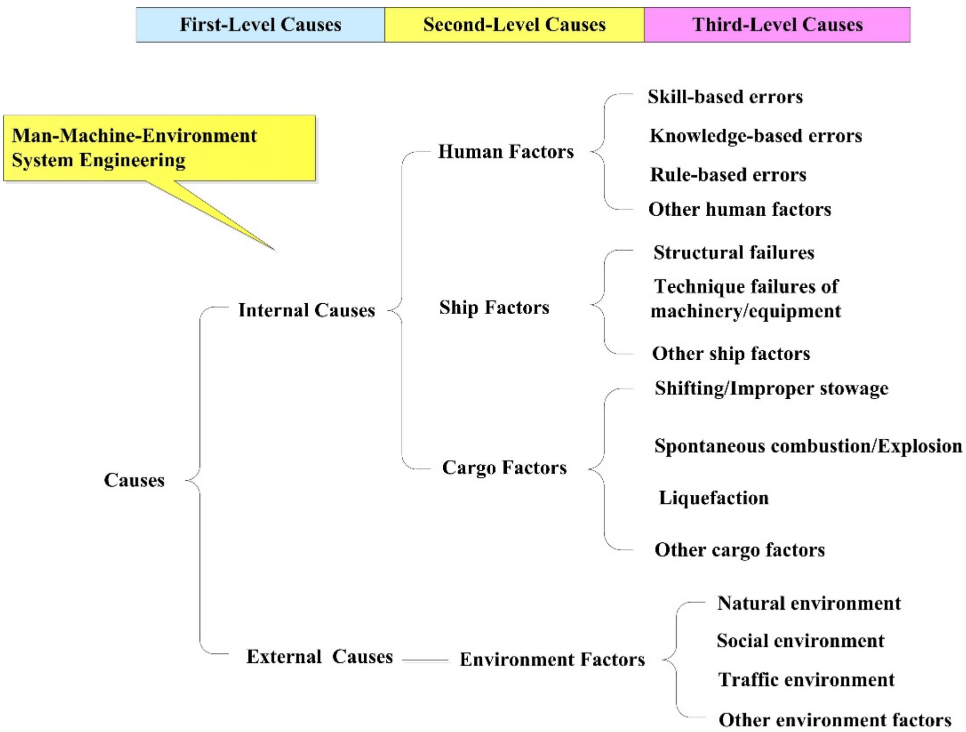
A ship may change its flag, classification and/or ownership status several times during its lifetime, and any given ship may also be involved in more than one incident; hence, the relationships between the ship and other connected entities are all many-to-many relationships. Additionally, any casualty must be located in a region, while any region may have observed more than one casualty, so the relationship between the ocean region and the maritime incident can be defined as a one-to-many relationship.

3.1.3 Attributes – blue ovals

Apart from entities and relationships, the other essential elements of an ER diagram are the attributes used to describe entities and relationships, i.e., database users registered these data fields to record a casualty. A comprehensive set of attributes is included to describe the maritime incident, the ship involved and their relationship because the choice of attributes reflects the level of detail at which one wants to represent information about entities and relationships. These attributes are designed on the basis of the recording format of several widely used maritime incident databases, especially the GISIS and the LLI. Modifications are also made to improve the recording patterns of some data fields, such as the incident causes and consequences. When identifying attributes, the attribute domain is simultaneously specified for each attribute, including the allowable set of values for the attribute, its size and its format.

As modern maritime technology tends to provide more safeguards against possible errors, the occurrence of a maritime incident requires an even larger combination of factors (Psaraftis et al., 1998b). Hence, a systematic record of factors contributing to casualties is necessary. From the perspective of man-machine-environment system engineering (Rasmussen, 1983), an improved taxonomy designed for collecting causal factors is presented in Figure 2. For example, environmental causes are deconstructed into natural conditions (e.g., berthing with tidal conditions (Lalla-Ruiz et al., 2016a), social environment, traffic environment (e.g., traffic management in waterways (Lalla-Ruiz et al., 2016b) and other environment factors o that maritime incidents can be investigated in more detail.

Figure 2 Alternative contributing factors of a maritime incident (see online version for colours)



Additionally, a systematic method is employed to record incident consequences, taking into account the consequences to people, ships and the environment. Without defining the seriousness of incidents, this paper only uses the 'facility indicator', the 'pollution indicator' and the 'result to the ship' (i.e., total loss/unfit to proceed/remains fit to proceed) attributes of an event to qualitatively indicate the incident results for the convenience of data retrieval. Details about consequences to people and the environment are also collected quantitatively. Moreover, since data fields can hardly provide a detailed context for each incident, the MSMDB also allows users to attach investigation reports and incident images to enhance the amount of contextual information available about any given incident.

3.2 Logical database design: ER model to relational model

When designing a relational database, the objective of the logical database design is to map a conceptual data model onto a relational data model, i.e., to translate the ER model into relation schemas. The relation schema describes the column headers for the table, which become the recording format of maritime incidents. A relation schema can be expressed as: name of the relation (attribute 1, attribute 2, …, attribute n). Attribute 1 in each relation is underlined to indicate that it is the primary key. In accordance with the ER model of the developed maritime incident database, relation schemas are derived as follows:

1 Entities
   - ship [*IMO*, dead weight tonnage, gross tonnage, length overall, breadth, depth, draught, hull type, building yard, built (year), type of ship, previous name]
   - maritime incident (*incident reference*, incident date, incident time, latitude, longitude, ocean region, type of location, initial event, subsequent event, cause, incident summary, images, investigation reports)
   - flag state (*flag state*, flag of convenience)
   - classification society [*class*, country, International Association of Classification Societies (IACS)]
   - beneficial shipowner [*beneficial shipowner*, country, formed (year), number of ships]
   - registered shipowner [*registered shipowner*, country, formed (year)]
   - ocean region (*ocean region*, natural condition, traffic condition, rescuing capacity nearby).

2 Many-to-many relationships:
   - happen (*IMO*, *incident reference*, ship name, age, origin, destination, number of crew members on board, particulars of cargo onboard, facility indicator, number of fatalities (dead or missing), number of injuries, result to the ship, pollution indicator, oil spills in tonnage)
   - ship-flag (*flag state*, *IMO*, from, to, call sign, MMSI)
   - ship-class (*class*, *IMO*, from, to)
   - ultimate ownership (*beneficial shipowner*, *IMO*, from, to)
   - legal ownership (*registered shipowner*, *IMO*, from, to).

An entity can be mapped to a relation in a straightforward way. Each attribute of the entity becomes an attribute of the table, and the primary key is underlined. With respect to relationships, a typical way to deal with the one-to-many relationship is to insert a foreign key into the table that represents the 'many' side of the relationship. And a many-to-many relationship needs to be transformed into two one-to-many relationships, which can be achieved by creating an additional relation. The attributes of such relation contain the primary keys of participating entities and the descriptive attributes of the relationship.

After transformation, relation schemas are checked against normalisation and integrity constraints to make sure all the relations are structurally correct. First, attributes in each relation are validated using the rules of normalisation to eliminate non-atomic values and data redundancy. Functional dependencies and the primary key of each relation are used in the process of normalisation. To achieve a balance between minimal data redundancy and maximum accessing efficiency, each relation conforms to the rules of third normal form, 3NF.

In addition, to prevent the database from becoming incorrect, invalid or inconsistent, the relational data model is examined against integrity constraints, including entity integrity, referential integrity and user-defined integrity. In particular, as entity integrity specifies, the primary key must not bean empty set of attributes. Assuming that ships without an IMO number replace this field with the ship name, the IMO number can be regarded as the primary key of the ship relation.

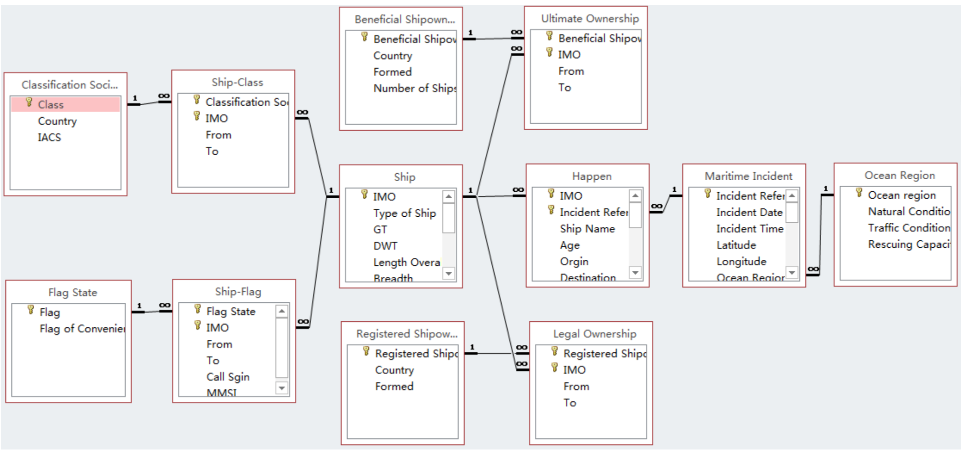
4 Implementation of the maritime safety management database

After completion of the basic design of the relational maritime incident database MSMDB, this section will implement the database and provide a sample database application to demonstrate that the developed database is implementable and has the potential for practical use.

4.1 Fundamental table structures

With the basic database design in hand, Microsoft Access, a well-known relational database management system (RDBMS), is utilised to implement the design and to execute a sample application that query the database. An Access database contains six types of objects: tables, queries, forms, reports, macros and modules. Tables are the ultimate data structures in which data are stored. The structures of the database tables are specified according to the previously designed relation schemas. These tables are then correlated through the primary and foreign keys. The relationships between pairs of tables created in Access are shown in Figure 3.

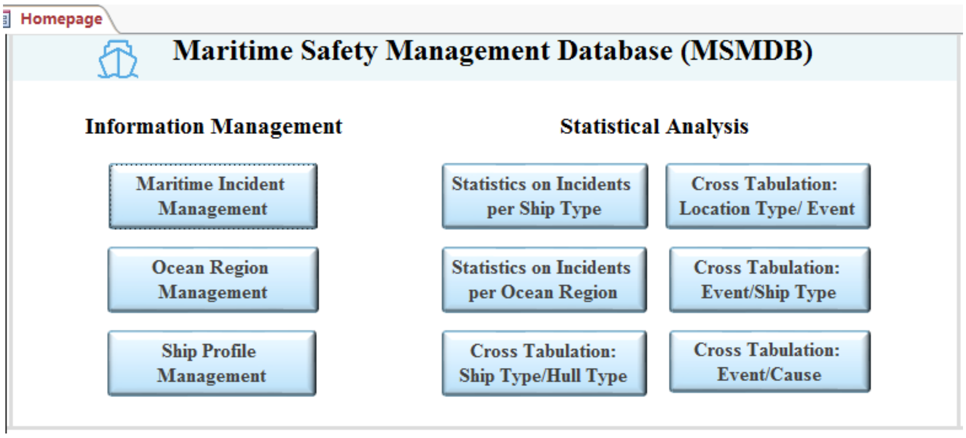
Figure 3 Establishment of tables and relationships in Microsoft Access (see online version for colours)



4.2 Practical queries and forms for database applications

In addition to tables, queries and forms are also utilised to build the target database and to create a practical application. Using the aforementioned objects, the proposed database management system launches two modules, an information management module and a statistical analysis module, as shown in Figure 4. Furthermore, we extract 200 serious incidents that occurred in the period 2012–2015 from the GISIS and import them into the MSMDB to concretely demonstrate the design approach. In the future work, more past incident entries will be imported into the relational database using efficient data migration techniques, and further evidence-based research can be conducted to derive statistically significant conclusions.

Figure 4 Homepage of the database application (see online version for colours)



Module one: information management

The information management module systematically manages the information about maritime incidents, ocean regions, ships, crewmembers etc., including the following three functions. The first function (Figure 5), maritime incident management presents a thorough description of incidents, like the spatial-temporal properties, and the causes and consequences. Moreover, it provides an access to the corresponding ship profile in each incident by correlating the ship entity and the maritime incident entity. In this case, for a ship whose information has been stored in the system, the administrator simply needs to input its IMO number in order for the system to automatically match this number with the archived ship information, which relieves the user of the burden of data entry.

Compared to the first function, the two other functions are more innovative. They are capable to manage incidents from multiple perspectives. As illustrated in Figure 6, 'Ocean Region Management' can provide both the basic information of a region and the incident records corresponding to each region. This function is helpful for coastal states who want to improve the safety level of nearby ocean regions, and ships passing through these ocean regions can learn about potential risks from past incidents and raise their vigilance accordingly. Another function, the 'ship profile management' is designed to document the flag, class and ownerships of each ship during different periods, as well as the crewmember list and incident history of each ship. This function may serve as a reference in estimating the risk level of a given ship.

Figure 5 Interface of the ‘maritime incident management’ function (see online version for colours)

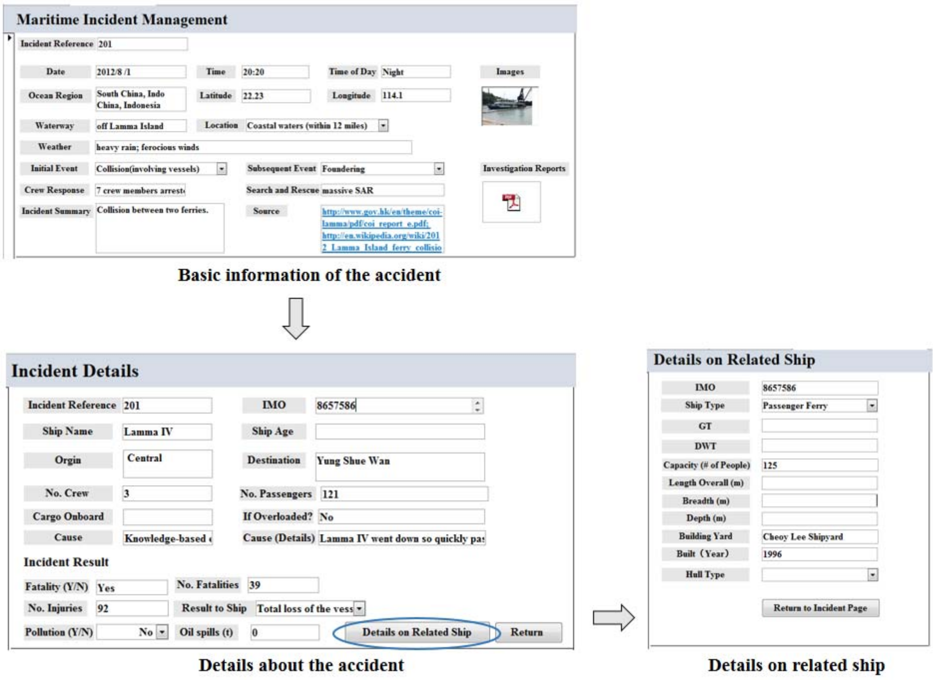
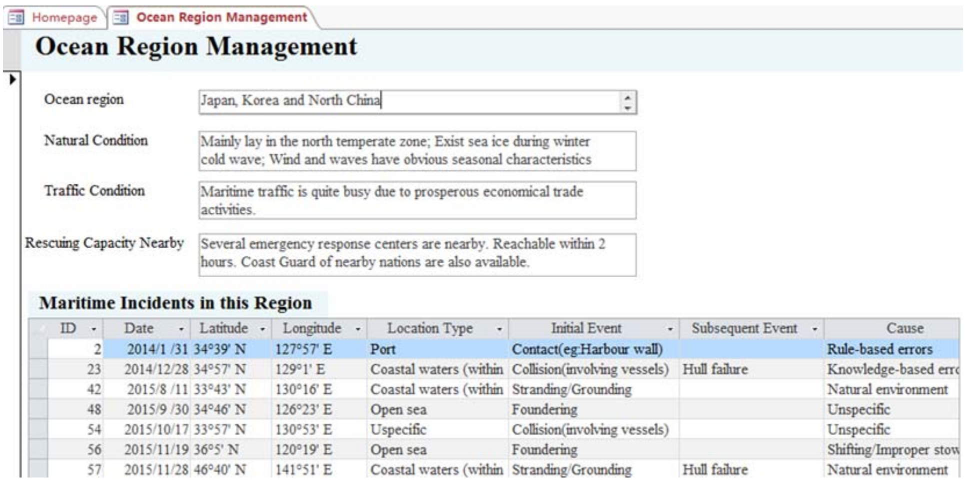


Figure 6 Interface of the ‘ocean region management’ function (see online version for colours)



Module two: statistical analysis

Programmed with the structured query language (SQL), the statistical analysis module includes two types of queries as sample applications, i.e., aggregate queries and crosstab queries.

Aggregate queries count the incidents grouped by a specific data field, e.g., the ship type or the ocean region, and perform descriptive statistics on the incident results. The function 'statistics on incidents per ship type' counts the incidents during a specified period of time according to the type of ship and reports the age distribution of the ships involved in such incidents and the cumulative incident consequences. The function 'statistics on incidents per ocean region' counts the number of incidents, and the number of ships involved in such incidents, according to the ocean region. Moreover, it considers the seriousness of incidents by counting the number of ships involved in incidents that produced fatalities or in incidents where the ship was totally lost and calculates the ratio of the number of ships involved in each of these types of incidents to the overall number of ships involved in all types of incidents in the region. In this case, one would be able to identify the 'hot spots' more easily by reasoning that if two regions have a similar quantity of incidents, but one reflects more serious consequences, the region with incidents that have more serious consequences is more likely to be a 'hot spot'.

In addition to the two aforementioned aggregate queries, the system also provides four crosstab queries. Crosstabs display the joint distribution of two or more variables, and they are usually represented in the form of a contingency table in a matrix. Crosstab queries are targeted at finding patterns or regularities between two types of data fields, e.g., the relation between the incident type and the cause, and at deriving design modifications and administrative suggestions. As incident records accumulate, this function can spot the incipient shifts of the trend in a timely fashion.

4.3 Case study: benefits of the MSMDB

In the era of big data, great progress has been made in data storage and analysis technologies. However, in the field of maritime safety management, maritime incident databases still have great room of improvement. In this paper, two examples are introduced to intuitively present the gap of some existing maritime incident databases and demonstrate the benefits of the proposed MSMDB.

1 Realisitic example: room of improvement of the maritime incident database.

   Firstly, we list one actual example about ferry incident recording that may be intuitive to indicate the room of improvement of the maritime incident database.

   Ferries provide a crucial mode of transportation for many in the waterside regions, especially in archipelagic nations like Indonesia and in river delta nations like Bangladesh. Because of the large number of passengers transported, it is of great importance to improve the ferry safety. Empirical ferry incident data play a crucial role in ferry safety research (Golden and Weisbrod, 2016). The Worldwide Ferry Safety Association (WFSA) has compiled a maritime incident dataset that consists of 25 data fields, including the ferry information, the circumstances regarding the incident etc., to gather incident entries and stored them in an Excel worksheet (Golden, 2015). It is a typical example about recording the incident data in a single table. Much basic information, such as the profile of the flag states and the introduction of ocean regions, is repeatedly recorded (i.e., data redundancy); while more relevant information for further reference, like the AIS information, is difficult to be added into the database (i.e., limited extension of the database). In addition, once a certain data field needs to be updated, e.g., the rank of the flag defined by the Paris MOU ('white, grey and black list'), the efficiency of data maintenance may be undesirable for the reason that the same piece of information needs to be modified in multiple places in one table instead of being updated synchronously.

2 Simplified example: high efficiency of the proposed MSMDB approach.

Secondly, we use a small simplified example to explain the disadvantages of recording incident data in a single table and highlight the benefits of employing the relational database approach in developing a maritime safety management database to reduce the data redundancy and strengthen the expansibility of the database. Figure 7(a) and Figure 7(b) show the incident dataset before/after using the relational database approach, respectively. In Figure 7(a), the indicators 'developed/developing country' and 'open registry (Y/N)' have appeared three times regarding the flag state 'Panama', which results in data redundancy. Therefore, we establish another table that dedicatedly records information related to the entity 'flag state' as shown in Figure 7(b). By this means, data redundancy is reduced in the original table, and the database is easier to be extended to include more information for further risk analysis, such as the indicators 'rank of the flag state', 'excess factor', etc. (Paris MOU, 2017).

Figure 7 (a) Table structure without using the relational database approach (b) Table structure using the relational database approach (see online version for colours)

Reference ID	IMO	Ship Name	Date	......	Flag State	Developed/Developing Country	Open Registry (Y/N)
1	9035802	K.PRIDE	2015/05/14	......	Panama	Developing Country	Y
2	9314038	KAIROS	2015/05/18	......	United Kingdom	Developed Country	N
3	8408715	QI YUAN	2015/05/20	......	Panama	Developing Country	Y
4	9438391	PINE 5	2015/06/12	......	Panama	Developing Country	Y

(a)

Reference ID	IMO	Ship Name	Date	......	Flag State
1	9035802	K.PRIDE	2015/05/14	......	Panama
2	9314038	KAIROS	2015/05/18	......	United Kingdom
3	8408715	QI YUAN	2015/05/20	......	Panama
4	9438391	PINE 5	2015/06/12	......	Panama

Foreign key

Primary key

Flag State	Developed/Developing Country	Open Registry (Y/N)	Rank	Inspections 2014-2016	Excess Factor	......
Panama	Developing Country	Y	White	6082	-0.45	......
United Kingdom	Developed Country	N	White	1260	-1.73	......

(b)

In summary, information regarding different aspects of an incident can be stored separately and yet still remain correlated through common fields shared by the tables with the use of the relational database. This design approach allows one to keep a reasonable level of data redundancy and to implement information maintenance functions more accurately and efficiently, including addition, deletion and update of maritime safety data. Moreover, the improved MSMDB is capable to be expended to incorporate more related data and forms a more comprehensive network of information for the purpose of risk analysis in the future.

5 Concluding remarks

This paper developed a relational maritime safety management database called the MSMDB and provided a sample database application to interact with the database. Three important conclusions are drawn as follows.

- Generally, three types of maritime incident databases exist: public databases, commercial databases and national databases. Based on the extensive exploration of numerous international and national maritime incident databases, we take two widely used databases, the GISIS and the LLI, as examples to explain some common limitations that exist in most databases in detail. By analysing the data semantics of maritime incidents in a structured way, an improved relational maritime incident database is developed to address the shortcomings of the aforementioned systems. In addition, a useful database application is designed as an example that demonstrates the utility of the proposed database.
- The relational database is the predominant type of database in use today. However, only few maritime incident databases, like the GISIS and the LLI, have partly utilised the basic ideas of the relational database, and many databases still follow a straightforward approach to record incidents in a single spreadsheet, even the Excel spreadsheet (Golden, 2015). In this research, the ER model is applied to help understand the data semantics of maritime incidents in a structured way. A standard relational database MSMDB is then developed to record maritime incidents in a logical way, which can help reduce data redundancy in maritime incident recording and facilitate the extension of the database. Moreover, the developed database may be used not only to collect incident information but also to conduct knowledge management in an intelligent way.
- The MCI module in the GISIS is widely used in maritime safety management and in academic studies. Established in 2005, the incident reporting format of this module has been constantly improved, and its data fields are well defined. However, few efforts have been made to improve the organisation of data stored in this database or the statistical functions provided by this database. Therefore, we suggest that the IMO test the relational database approach to test its efficiency in managing maritime incidents and further improve the MCI module in GISIS. The anticipated efficiency may spur international and national authorities to redouble their efforts to fully investigate maritime incidents and publicly report on the investigations. Thus, evidence-based maritime safety management can be better realised by improvement of the data sources.

This paper is a pilot study towards the efficient data collection and maintenance, and demonstrates the potential for systematically designing the relational database model to manage maritime safety. Further work about acquisition of efficient data migration techniques, e.g., some extract, transform and load (ETL) tools and self-compiled batch processing tools will be developed so that historical incident data could be migrated more conveniently. And then evidence-based research using the proposed database could be conducted to derive statistically significant conclusions. Additionally, converting the proposed database to a web based one is also necessary to support online operations.

Acknowledgements

This paper is supported by Shanghai Natural Science Foundation (14ZR1422400).

References

Asbjørnslett, B.E., Hassel, M. and Hole, L.P. (2010) *A Comparative Study of Vessel Accident Databases from a Risk Management Perspective* [online] https://www.researchgate.net/publication/289240345 (accessed 9 May 2017).

Chen, P.P.S. (1976) 'The entity-relationship model – toward a unified view of data', *ACM Transactions on Database Systems (TODS)*, Vol. 1, No. 1, pp.9–36.

Chen, S.Y.H. and Su, K.K.M. (2005) 'A study of the development of Taiwan maritime casualty database system', *Journal of Marine Science & Technology*, Vol. 13, No. 3, pp.200–208.

Codd, E.F. (1970) 'A relational model of data for large shared data banks', *Communications of the ACM*, Vol. 13, No. 6, pp.377–387.

Devanney, J. (2008) *Uses and Abuses of Ship Casualty Data*, Center for Tankship Excellence [online] http://www.c4tx.org/pub/casdata.pdf (accessed October 2016).

Dobler, J-P. (1994) 'The requirement for the publication of detailed global marine casualties statistics', *Maritime Policy & Management*, Vol. 21, No. 1, pp.45–60.

Edwards, M. and Kauffman, M. (2016) 'Data tools to identify safety and environmental risks to the marine transportation system', *Transportation Research Record: Journal of the Transportation Research Board*, No. 2549, pp.101–110.

Golden, A. (2015) *Ferry Fatalities: Statistics and Causation of Major Accidents 2000–2014 (Excel Spreadsheet File)* [online] http://www.ferrysafety.org/news.htm (accessed 9 May 2017).

Golden, A. and Weisbrod, R. (2016) 'Trends, causal analysis, and recommendations from 14 years of ferry accidents', *Journal of Public Transportation*, Vol. 19, No. 1, pp.17–27.

Hassel, M., Asbjørnslett, B.E. and Hole, L.P. (2011) 'Underreporting of maritime accidents to vessel accident databases', *Accident Analysis and Prevention*, Vol. 45, No. 6, pp.2053–2063.

Hong, N. (2012) 'The melting Arctic and its impact on China's maritime transport', *Research in Transportation Economics*, Vol. 35, No. 1, pp.50–57.

Huang, D.Z., Hu, H. and Li, Y.Z. (2013) 'Spatial analysis of maritime accidents using the geographic information system', *Transportation Research Record: Journal of the Transportation Research Board*, No. 2326, pp.39–44.

International Maritime Organization (IMO) (2008) *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)*, Resolution MSC. 255(84).

International Maritime Organization (IMO) (2009) *Global Integrated Shipping Information System (GISIS)*, Resolution A. 1029(26).

International Maritime Organization (IMO) (2011) *Strategic Plan for the Organization (for the Six-Year Period 2012–2017)*, Resolution A.1037(27).

International Maritime Organization (IMO) (2014) *Reports on Marine Casualties and Incidents*, MSC-MEPC.3/Circ.4/Rev.1.

Ladan, M. and Hanninen, M. (2012) *Data Sources for Quantitative Marine Traffic Accident Modeling*, Aalto University Publication Series [online] https://aaltodoc.aalto.fi/handle/123456789/3641 (accessed October 2016).

Lalla-Ruiz, E., Expósito-Izquierdo, C., Melián-Batista, B. and Moreno-Vega, J.M. (2016b) 'A set-partitioning-based model for the berth allocation problem under time-dependent limitations', *European Journal of Operational Research*, Vol. 250, No 3, pp.1001–1012.

Lalla-Ruiz, E., Shi, X.N. and Voß, S. (2016a) 'The waterway ship scheduling problem', *Transportation Research Part D: Transport & Environment*, In Press, Vol. 60, No. 1, pp.191–209.

Li, K.X., Yin, J.B. and Fan, L.X. (2014b) 'Ship safety index', *Transportation Research Part A: Policy and Practice*, Vol. 66, No. 1, pp.75–87.

Li, K.X., Yin, J.B., Bang, H.S., Yang, Z.L. and Wang, J. (2014c) 'Bayesian network with quantitative input for maritime risk analysis', *Transportmetrica A: Transport Science*, Vol. 10, No. 2, pp.89–118.

Li, Y.Z., Hu, H. and Huang, D.Z. (2013) 'Developing an effective fuzzy logic model for managing risks in marine oil transport', *International Journal of Shipping & Transport Logistics*, Vol. 5, Nos. 4–5, pp.485–499.

Li, Y.Z., Hu, H. and Huang, D.Z. (2014a) 'Data fusion of maritime incident databases with Dempster-Shafer theory', *Transportation Research Record: Journal of the Transportation Research Board*, No. 2426, pp.27–32.

Martin, P.H., LeBoeuf, E.J., Daniel, E.B., Dobbins, J.P. and Abkowitz, M.D. (2004) 'Development of a GIS-based spill management information system,' *Journal of hazardous materials*, Vol. 112, No. 3, pp.239–252.

Mazaheri, A., Montewka, J., Nisula, J. and Kujala, P.(2015) 'Usability of accident and incident reports for evidence-based risk modelling – a case study on ship grounding reports', *Safety Science*, Vol. 76, No. 1, pp.202–214.

Paris MOU (2017) *2016 Performance Lists Paris MOU* [online] http://www.cnss.com.cn/index.php?c=index&a=show&catid=330&typeid=3&id=276183&remains=true (accessed 9 May 2017).

Psaraftis, H.N., Caridis, P., Desypris, N., Panagakos, G. and Ventikos, N. (1998b) 'The human element as a factor in marine accidents', *IMLA-10 Conference*, St. Malo, France, pp.1–14.

Psaraftis, H.N., Panagakos, G., Desypris, N. and Ventikos, N. (1998a) 'An analysis of maritime transportation risk factors', *8th International Offshore and Polar Engineering Conference*, Montreal, Canada, 24–29 May, pp.477–483.

Psarros, G., Skjong, R. and Eide, M.S. (2010) 'Underreporting of maritime accidents', *Accident Analysis and Prevention*, Vol. 42, No. 2, pp.619–625.

Ramakrishnan, R. and Gehrke, J. (2000) *Database Management Systems*, McGraw-Hill, New York.

Rasmussen, J. (1983) 'Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 3 No. 3, pp.257–266.

Sun, H.N., Duan, X.F. and Liu, Z.J. (2007) 'SQL-based management system and implementation of maritime accident,' *Journal of Dalian Maritime University*, Vol. 145, No. 10, pp.120–121.

Tirunagari, S., Hanninen, M., Stanhlberg, K. and Kujala, P. (2012) 'Mining causal relations and concepts in maritime accidents investigation reports', *International Journal of Innovative Research and Development*, Vol. 1, No. 10, pp.548–566.

Tsai, C.L. (2016) 'An empirical study of the acceptance of electronic chart displays and information systems in the shipping industry', *International Journal of Shipping & Transport Logistics*, Vol. 8, No. 4, pp.425–441.

Uğurlu, O., Köse, E., Yıldırım, U. and Yüksekyıldız, E. (2015) 'Marine accident analysis for collision and grounding in oil tanker using FTA method', *Maritime Policy & Management*, Vol. 42, No. 2, pp.163–185.

Wan, Z. (2015) 'China's scientific progress hinges on access to data', *Nature*, Vol. 520, No. 7549, p.587.

Wong, M.C. and Yip, T.L. (2012) 'Maritime piracy: an analysis of attacks and violence', *International Journal of Shipping & Transport Logistics*, Vol. 4, No. 4, pp.306–322.

Yin, J.B. (2013) *Quantitative Risk Assessment for Maritime Safety Management*, PhD thesis, The Hong Kong Polytechnic University.

Zhang, B., Wang, X.S, Luo, Y., Chen, X.H., Chen, J. and Wang, S.R. (2012) 'Method of urban traffic crash location record description', *China Safety Science Journal*, Vol. 22, No. 2, pp.127–133.

Zhang, J.F., Teixeira, A.P., Soares, C.G., Yan, X.P. and Liu, K.Z. (2016a) 'Maritime transportation risk assessment of Tianjin Port with Bayesian Belief Networks', *Risk Analysis*, Vol. 36, No. 6, pp.1171–1187.

Zhang, L., Wang, H. and Meng, Q. (2016b) 'Evidence-based ship accident consequence analysis using heterogeneous-source accident reports', *Proceedings of the 95th Annual Meeting of Transportation Research Board*, Washington D.C., pp.1–18.