Big Data in Shipping - Challenges and Opportunities

Ørnulf Jan Rødseth, MARINTEK, Trondheim/Norway, OrnulfJan.Rodseth@marintek.sintef.no
Lokukaluge Prasad Perera, MARINTEK, Trondheim/Norway, Prasad.Perera@marintek.sintef.no
Brage Mo, MARINTEK, Trondheim/Norway, Brage.Mo@marintek.sintef.no

Abstract

Big Data is getting popular in shipping where large amounts of information is collected to better understand and improve logistics, emissions, energy consumption and maintenance. Constraints to the use of big data include cost and quality of on-board sensors and data acquisition systems, satellite communication, data ownership and technical obstacles to effective collection and use of big data. New protocol standards may simplify the process of collecting and organizing the data, including in the e-navigation domain. This paper gives an overview of some of these issues and possible solutions.

1. Introduction

Shipping is currently on its way into its fourth technical revolution, sometimes called Shipping 4.0 or cyber-shipping. The first revolution was the transition from sail to steam around 1800, then from steam to diesel engines around 1910 and the third came with the introduction of automation and computerized systems around 1970. The current revolution is about digital data in all aspects of shipping operations and can be compared to what is called Industry 4.0 in the land industry, *Hermann et al. (2015).* Shipping 4.0 includes extensive use of new technology like Cyber-Physical Systems (CPS), Internet of Things (IoT) and the Internet of Services (IoS). This technology provides more intelligent on-board equipment with embedded computers that provide a plethora of new data and information as well as new shore services to use the data.

It is not obvious that this is really “big data”. Big data is often being defined as being high volume, high speed and/or high variety so that “conventional” data processing techniques are insufficient for efficiently using the data in analytics, decision-making or control, *De Mauro (2015).* In terms of processing power and computer capacity one can hardly say that this is the case in shipping today. While large sets of data from ships may be awkward or impossible to handle in some simple tools, it does not really pose big data challenges to today’s computer systems. However, the volume and the complexity of the information will require new methods and tools to enable the users to understand the information properly and, in that sense, can be said to be a form of a big data problem.

While the fourth shipping revolution provides a vast number of new possibilities for more advanced on-line control and off-line analytics, there are also a number of other problems than the sheer size of the data sets. This paper explains some of the issues that have emerged in our work on big data and gives some suggestions as to how they can be solved.

2. Increase in availability of data

There are several trends that contribute to the availability of more data about ship and ship systems, i.e. navigation and automation systems. All of these trends are connected to Shipping 4.0, but materialises in quite different ways and with very different impact on how easy it is to use the data. This section will give an overview of some of the main data sources used today.

2.1 Bridge data network, mandatory and special purpose instrumentation

The ship bridge equipment is normally interconnected with digital interfaces from the IEC 61162 family of standards, *IEC (2007-2015).* This makes it relatively easy to get access to measurements from the navigation sensors and equipment. A steady but slow increase in carriage requirements from IMO and flag states increases the available data from the bridge.

In addition, some special ships will require special instrumentation relevant for their operations. This may include wave radars, oil spill detectors, high accuracy inertial navigation sensors etc. These may or may not be accessible via standard interfaces.

The voyage data recorder (VDR) is sometimes used as a collection point for such data, but one should be aware of limitations in the number of data points the VDR collects and how often the data is registered.

2.2 Conventional automation

Automation systems on most ships collect large amounts of data, but of variable quality. Even on a relatively old bulk ship the number of input/output points may be several hundreds. On modern and more complex ships the number of data points can be several tens of thousands.

There are two main problems with this type of data: The first is to get access, as the data is normally only available within closed and vendor-specific systems. The second is the quality of the data. As the data is used in closed loop control or in alarm limit supervision, sensors may not be particularly accurate. For older automation systems, it may also be a problem that sensors are defect or disconnected. The automation system rarely provide quality attributes on the data so the user will need to check if the measurement are sensible and if values are stuck or fluctuates too much.

Sometimes, modern integrated bride systems are equipped with both navigation and automation systems integrated into a single platform. This can more easily facilitate new cyber-physical systems. These systems also can collect significant amounts of data and that may introduce the same types of problems as discussed in this paper.

2.3 New cyber-physical systems

Modern ships make use of more advanced equipment with built-in sensors and control systems. This includes engines and power generation, winches with torque-control, advanced dynamic position systems, new navigational sensor systems and so on. The systems have computers and sensors for condition monitoring, closed loop control or for detecting alarm conditions. This type of integration of physical systems with computer control is one of the main characteristics of Industry 4.0: Cyber-Physical Systems (CPS).

The problems of using data from these systems are similar to that of automation: Difficulty of access and suitability for use outside the system for which they are intended. For CPS, general quality may be less of an issue as most data points are used in active control or monitoring of the equipment and errors in sensors will be detected rapidly.

2.4 Ship performance monitoring

Later years, particularly when the oil prices were high, saw a substantial increase in ship instrumentation for performance monitoring and optimization. This typically included shaft torque meters, fuel mass-flow meters, improved environmental sensors, trim measurements etc. This instrumentation was fitted explicitly to provide data for improved ship performance and one can expect that both access to and quality of data is good, dependent on how and by whom the instruments were fitted. In some cases, instruments and services may be provided by third parties and this may reduce access to the information.

2.5 Ship reporting

Ships are sending a number of operational and administrative reports to shore. This includes various mandatory port and port state reports, noon at sea reports, technical maintenance reports etc. Many of these can be very valuable as input in data analysis as they represent operational decisions by the crew. These reports are normally generated from manual input and this makes them susceptible to data entry errors. Some reports also have impact on the economic performance of the ship so figures may be tweaked to avoid penalties or additional costs.

This again means that one needs to be careful about quality control and verification of data in such reports. On the other hand, these reports also represent the crew's opinion of the state of affairs and general operational outlook, which may be used to quality control and verify other automatically acquired data as well as to complement more technical measurements.

2.6 External ship monitoring – AIS and VTS

Another growing source of ship data is external monitoring. Most coastal states operate automatic identification system (AIS) base station networks along their coast. This will monitor ship traffic within range of the VHF radios on the base stations. Low Earth Orbit satellite AIS receivers are used to extend detection beyond the range of the coastal stations. AIS receivers can provide very valuable data on ship movements. Under way, ships will send AIS data quite frequently, normally minimum each 10 seconds. Data that is transmitted is position, speed, course, true heading and rate of turn. Less frequent messages will transmit ship static data, draught and other voyage related information.

In addition to coast state authorities, there are also several private providers of AIS data, either using shore stations or satellite. Thus, access to AIS-data is normally not a problem, but quality and prices may vary. In most cases one will get "filtered" data so that one also needs to consider any implications on data quality from the filtering. Filtering is normally some form of decimation of data in high density reporting areas or interpolation in areas without registrations.

In addition to AIS, coastal states also operate vessel traffic services (VTS) that monitor ship traffic with radar and CCTV systems in certain areas. In some cases, they may also use synthetic aperture radar or other forms of satellite sensors to improve local situation awareness. Neither of these data is normally available to the public.

2.7 Weather data

Forecasted and historic weather data is available for free or for payment, dependent on the level of detail one requires. The free information is relatively low resolution, but can give good high-level information about environmental effects on the ship. The data is less useful close to the coast or in areas where weather phenomena are influenced by smaller scale geographic features, e.g. close to islands or within narrow ocean currents.

2.8 Port call data

Port and ship agents are collecting valuable information on ship movements close to and in ports. This is mostly operational data that can give very accurate information on port delays, time taken for loading and discharge etc. The data is often made available to the ship and ship owner as a statement of facts or a port log. The data tend to be accurate as it is often used as basis for calculating port fees, performance claims and other commercially important data.

3. Obstacles to efficient collection of ship data on board

The previous section gave an overview of commonly used data sources on or off the ship and some of the problems encountered when trying to access and use the data. Table I summarises the main problems and links them to the data sources in question. The following sub-sections will give more details about each of the acquisition problems.

Table I: Common data acquisition problems

Type of data acquisition / Problem area	1. Bridge	2. Automation	3. CPS	4. Performance	5. Reporting	6. Monitoring	7. Weather	8. Port data
Context dependent data quality	X	X	X					
Safety and security considerations	X	X	X					
Complex phenomena are difficult to measure	X			X			X	
Non-automated data entry create errors					X	X		X
Commercial issues may induce wilful errors in reports					X			X
Bad quality instrumentation, no quality attribute on data		X						
Proprietary and costly interfaces to data		X	X					
Ownership of derived data			X	X				
Lack of standards		X	X	X				
Artefacts in AIS base station and satellite reception						X		
Cyber attacks	X				X	X		

3.1 Context dependent data quality

The specific acquisition context of sensor data is more or less a problem for all measurements that are taken from one system and used in another. One example is different sensors of similar types, where they are located and how they relate to each other: As an example, the ship normally has a number of position sensors, each with its own local position reference and individual quality attributes that may not be known outside the bridge system. Using raw position data from the navigation system can cause accuracy or consistency problems.

A similar problem occurs in sensors used in closed loop control or as alarm limit sensors. Basic sensor properties such as linearity, offsets, accuracy and stability will always be adapted to the task at hand and the fitting of a specific sensor is a trade-off between cost, robustness and fitness for purpose. For control and alarm applications, sensor artefacts can be adjusted in the control and monitoring software and the absolute quality of the measurement may be of lower importance. Unless the adjustment factors are known to external systems, use of these measurements outside their systems rapidly becomes problematic. Unfortunately, the context problem is not easy to solve and will require very complex attribute lists for any measurement that shall be used outside its original context. In practical terms, this may be too complicated to implement in a standardised way, *Rødseth (2016).*

3.2 Safety and security

Connections to the bridge or other data networks that provide safety or security related functions may be a safety or security hazard. Any physical connection into networks or systems can be a vector for error propagation or hostile attacks. Thus, such connections are often not allowed by flag state or class authorities. On bridge networks, there have been developments in safer Ethernet networks and firewall technology for connecting to external networks *IEC (2015), Rødseth and Lee (2015).*

For automation networks there may also be possibilities for using safe connections via firewalls and gateways. Unfortunately, standards for this have not yet been developed so this type of solutions can become expensive (see sec. 3.9).

3.3 Measuring complex external phenomena

Complex phenomena such as environmental impact on the ship performance are difficult to measure accurately and care must be taken when such data is used in calculations. Speed through water, wind speed, waves and other similar data vary significantly around the ship and it is generally difficult or impossible to estimate overall impact just from one measurement point, which is normally what one has on most ships.

In addition, the fact that the sensors are mounted on or outside the hull contributes to errors. They are susceptible to damage, coating by foreign matter or general degradation that make the measured values less reliable.

3.4 Non-automated data entry create errors

Experience shows that manual entry of data into report forms, computer systems or AIS transceivers is an important source of errors. For AIS, this will manifest itself particularly in voyage destination, ship draught and sailing mode, but it also sometimes shows up as illegal or wrong ship identity.

For AIS there are also issues related to the transmission of navigational data such as rate of turn and true heading that should be got from sensors external to the AIS. Many AIS transmitters are not connected to such sensors and send non-valid data or data derived internally from the position.

It is difficult to avoid this problem. One obvious solution is to use automated data entry, but this is often too costly if it requires physical connections to other systems. Extensive validity check is probably the most relevant solution in the short term. However, automated ship reporting is a prioritized solution in the e-navigation strategic implementation plan, *IMO (2014),* so one may expect some developments in this area in the coming years.

3.5 Commercial considerations may induce wilful errors in reports

Some reports from the ship to shore have commercial implications, e.g. for owner or charterer. This may be bunkering, fuel consumption and speed reports including reports on contractual performance, exceptions related to bad weather or ship safety issues. Traditionally, this has rendered some reports less reliable as the operator has had direct economic interests in reporting misleading data.

Today, the possibilities of tampering with such data is much lower as it is fairly straight forward to do consistency checks via AIS satellite, weather data or other generally available information. However, the problem does to some degree persist and needs to be taken into consideration when commercially sensitive data is used in analysis.

3.6 Low quality instrumentation – no quality attributes

Sensor data on older ships and sensor data that are rarely used operationally can sometimes be of dubious quality. This can be because the sensor has failed, it has been disconnected, pained over or from any number of other problems. This is a relatively common problem with data from the automation and alarm systems.

Most automation systems will not perform quality checks on the data or provide any other quality attributes to the measurements. When data is exported out of the system, it will be difficult to assess whether the data can be relied on or not.

Today, this problem needs to be handled by careful data quality checks in the systems that use the data. There is work under way to look at new interface standards (sec. 3.9) that may introduce improved quality attributes for automation data.

3.7 Proprietary and costly interfaces to data

Most parties accept that the ship owner is the owner of any data measured or otherwise produced on-board. However, as long as these data are not readily available via standard interfaces, the owner or operator is likely to pay a high price for getting access to them. It will normally require retrofitting of special hardware and software that together with the required service personnel can be expensive.
This is less of a problem for systems that use established interface standards, such as the bridge data network. Unfortunately, standards are missing for many ship systems and in particular for interfacing to automation systems. This may change with recent standards initiatives (sec. 3.9).

It is particularly difficult to avoid these costs for older ships that are being retrofitted, but it should be kept in mind for new-builds where new interface technology and costs are easier to negotiate with the yards.

3.8 Ownership of special or derived data

Some systems may be delivered with special restrictions on the use of data internal to the system. In some cases, condition based maintenance and general monitoring is provided as a service to the ship owner where the owner does not at all have access or ownership to the underlying data. Similar third party services are sometimes also used for various ship or fleet monitoring or optimization functions. This may also imply restrictions on ownership of the data.

Even if the ship-generated measurements are accessible to the ship owner, there may also be derived and intermediate data that is not. These may be the results of various calculations and analysis performed on the ship or on shore.

If third party services are used for monitoring of ship or fleet, one should consider how ownership of collected and derived data could be ensured. This is not as clear-cut as for general on-board data, but will be something that can be negotiated with the service provider.

3.9 Lack of interface standards

The lack of open interface standards is an important problem for many types of data acquisition although most pronounced for interfaces to automation and embedded control systems. Bridge and navigational systems have some established standards that are suitable for open interfacing (sec. 2.1) and some types of ship reporting is using standard formats, *Rødseth (2016).* Lack of interface standards means that all data acquisition interfaces will need to be built to purpose and that the configuration of the interface most likely will be different for all implementations, even on almost identical ships (sec. 3.7). This is costly, time consuming and error prone. Arguably, one can say that this is one of the fundamental obstacles to more efficient data analytics on ships. There is work underway to develop interface standards for automation systems. *ISO (2015a, 2015b)* has proposed two new standards to cover respectively digital interfaces and data tagging for shipboard use. It remains to be seen when these standards will be finished and accepted in the market.

3.10 Artefacts in AIS base stations and satellite reception

Data acquisition from AIS shore systems is limited to when the ship is within range of a base station. Reception will also be hindered by radio shadows. When reception is possible, the ship’s transmission rate is highly variable and dependent on ship speed and manoeuvres.

Fig. 1: North Europe view of AISSat1 and AISSat2 orbits over a 24-hour period

Satellite reception is limited by the orbits and radio equipment of the satellite. Norway currently operates two satellites in polar orbits (see Fig. 1). These give around eight minutes observation windows approximately each 45 minutes north of 75º, but this rapidly decreases to about one window each two hours around the south part of Norway. At Equator, one will get two observations per day, one for each of the satellites. The satellites are also subject to a number of radio-technical and atmospheric disturbances that further decreases probability of observations, *Rødseth et al. (2015).* This is particularly a problem near areas with high ship traffic density due to cross-talk between ships. Fig. 2 shows a plot of satellite observations of a ship operating around 80º North over a period from January 2014 to November 2014. The y-axis shows time between observations in hours. The high peaks (long observation drop-outs) are typically during port calls while the reduction in time between observations from July 2014 is due to the start of the testing of the second AIS satellite. AISSat2, *Rødseth et al. 2015*. In the period up to June 2014, only AISSat1 was operational.

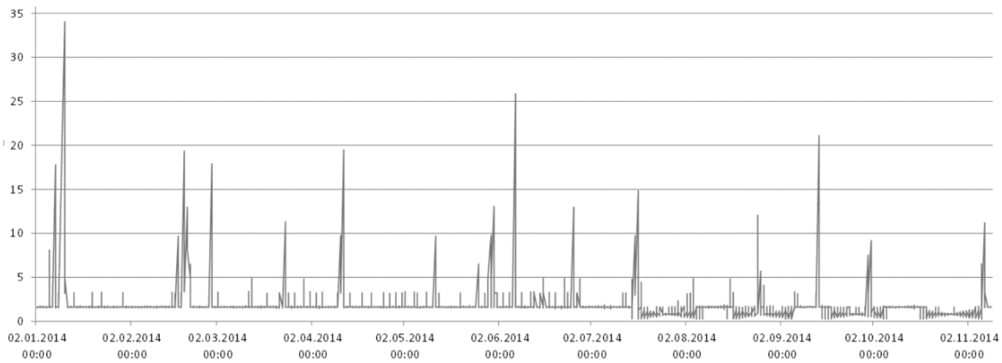

Fig. 2: Hours between AIS observations

The irregular reception pattern means that data may have to be carefully interpolated and filtered to avoid sampling artefacts in calculations that are dependent on regular time series. One also needs to consider the sources of AIS data before assumptions on acquisition frequency, probability or data quality can be made.

3.11 Cyber-security challenges

Cyber security is not considered a major factor in general data acquisition on-board, but one should be aware that certain data could be jammed or spoofed from cyber-attacks. Examples have shown that even GPS position signals can be spoofed. All data that relies on wireless transmissions from other ships or shore are susceptible to spoofing or jamming. As examples, this applies to AIS or radar targets acquired by the ship and sent to shore. It will also apply to AIS signals data collected by shore or satellite systems.

All reports sent from the ship to shore (or vice versa) can also be intercepted and tampered with. Any application of ship data needs to consider if this problem is serious enough to include in the analysis of errors sources, *Rødseth and Lee (2015)*.

4. Maritime big data management and applications

As the previous sections have shown, there are many sources of data on the ship, potentially high volumes of data and correspondingly many possibilities for errors. Data management on the ship and on shore has to be performed along three axes as shown in Fig. 3.

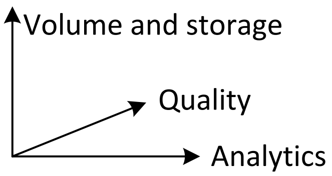

Fig. 3: Three dimensions of big data management

Volume and storage management manages large volumes of data in a structured collection on board and/or on shore. Volume management will also have to consider the cost and capability of moving large sets of data between ship and shore. The science of examining these data with the purpose of drawing meanings can be categorised as data analytics. Quality control of the data needs to be considered by either or as a separate activity and is critical for the correctness of the analysis results.

4.1 Volume and storage management

Sensors and data acquisition systems collect large quantities of ship performance and operational data. Some systems, in particular bridge and performance monitoring can be equipped with several sensors that often collect similar types of parameters. The same is the case for AIS observations where different base stations can receive simultaneous data from the same ship. Many sensors, including AIS, also collect data with a time resolution that is often much higher than required for ship performance analysis.

Generally, measured data can be functionally or temporally redundant in many situations and those situations should be identified during the volume and storage management process. Most data sets can normally be reduced to a much smaller data set by removing redundancy. This significantly improves the data handling and communication process, *Perera and Mo (2016)*. However, extensive knowledge about the structure of the data sets and the intended analysis is required to avoid removing important data.

There are also ways to reduce the amount of data using statistical methods that remove some of the data details but focuses on keeping the main components and the variation in the data, e.g. principal components analysis, PCA. Data processing of this kind may both reduce the size of data sets and filter out noise.

Virtually all methods that reduce the volume of the data will also remove information, typically about phenomena that is not related to the principal interest of investigation or main statistical trends in the data. Removing data items that have been registered by different sensors can, as an example, remove information about the relative reliability of each sensor.

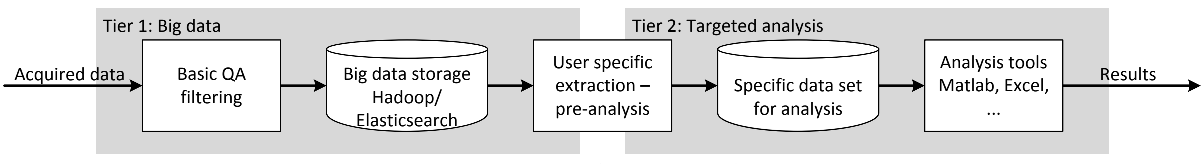

Fig.4: Proposed system for ship performance data

Data management needs to consider how data is stored and structured, particularly where one has to handle large data sets efficiently. In light of the conflict between having consistent and easy to manage data set and not losing details in the data sets that may be of use later, we are currently looking into a two-tier storage system as illustrated in Fig. 4. The lowest tier will use big data technology to store as much data as possible, after initial bad data removal and careful data reduction. The second tier will use specific extraction, structures and tools for the analytic problem at hand.

4.2 Ship-shore communication

The digital connectivity of ships is increasing, in particular for advanced and well-managed ships. Land based mobile communication systems is also more accessible to ships at a relatively low costs and high capacities. However, high sea satellite communication is in many cases a limiting factor for transfer of data between ship and shore.

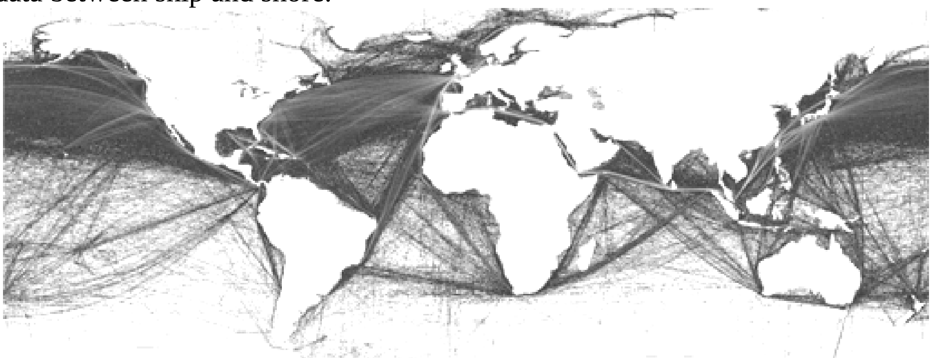

Fig.5: Ship density in the world (B.S. Halpern, from T. Hengl; D. Groll / Wikimedia Commons)

Operators of telecommunication satellites will point their beams to areas where there are customers, and many areas of the deep seas are not prime locations for high throughput services (see Fig. 5). This means that for most ships, there will be limitations in maximum available bandwidth and costs of transmissions, at least in certain areas. This also means that the data management system needs to consider how the available communication capacity can be used to best possible effect.

This process can be integrated with analytical processes, e.g. to develop hierarchical performance indicators where higher level indicators give an overview of status while it is still possible to drill down into more details if and when that is required. This principle is used in technical condition and safety indexes, Rødseth *et al. (2007)*, and in the Shipping KPI system, *Rialland et al. (2014)*. This principle can save processing and transmission costs and time as only selected data elements needs to be handled at a given abstraction level and time.

Perera and Mo (2016) have also discussed possibilities for using principal component analysis for data compression. Here, a volume saving of 30% at an information loss of about 6% has been demonstrated. This will come in addition to any conventional data compression technique used in such system.

For all data reduction approaches one should also consider to keep a more complete record of measurements on-board for later off-line analysis. Much of the value of the big data approach is that one can use novel and deep investigations on large data sets collected over time to investigate phenomena that was not even thought about at the time of data collection. To do this successfully, one will have to be careful about how data is filtered.

4.3 Data quality

Several data quality issues were discussed in sec. 3 and one can identify four main groups of quality problems that can occur individually or in combination:

1. Unreliable data: The data input is completely or partly faulty and there is no clear or consistent relationship between data value and the observed phenomenon.
2. Context specific artefacts: The input value needs additional and possibly unavailable information to be interpreted correctly, e.g., the exact location of the sensor on the ship.
3. Temporal artefacts: The data input frequency creates problems due to too high, too low or too irregular sampling frequency. This is also related to the Nyquist–Shannon sampling theorem.
4. Non-proportional data: The input data is not proportional to the phenomenon observed. This may be due to problems with linearity, offsets, hysteresis, saturation, etc.

Each of these classes of problems can use many specific and diverse methods to filter out errors and correct for artefacts. However, a common situation is that the actual cause of data degradation is not known and one may have to use general statistical methods to try to detect measurement artefacts.
As discussed in *Perera and Mo (2016),* systematic data reduction can be used for quality enhancements. Reducing redundancy in measurements will be an opportunity to detect and remove data that is outside expected values or areas. This could be based on mathematical models of the physical system, empirical or statistical methods or on various forms of machine intelligence. In all situations, the quality and level of details of the methods will determine what information is removed and lost. Lost information may or may not be acceptable for general data acquisition. Even directly faulty data can contain information, e.g. on sensor reliability. However, data that is outside expected ranges should always be flagged as such so that it is not used in analysis without proper precaution. Data analytics in itself, when having access to data that has redundancy in the systems states they observe, can also be used for sensor fault detection.

Quality management can also be performed independently of analytics and volume management. Sensor system knowledge, outlier detection, trending and statistical methods can be used to identify and possibly remove suspicious data elements.

4.4 Data analytics

Information is sometimes defined as data that answers a question about the world. There are two main ways to convert data to information as illustrated in Fig. 6. One can use knowledge about the world to formulate the question that the data provides the answer to (model based) or one can use various forms of discovery methods, e.g. based on statistics, correlation techniques or "machine intelligence" (MI), to discover new and unknown relationships in the world. Much of the ICT-oriented research on big data focus on the latter while much of what is currently going on in the shipping sector make use of the model based information extraction.

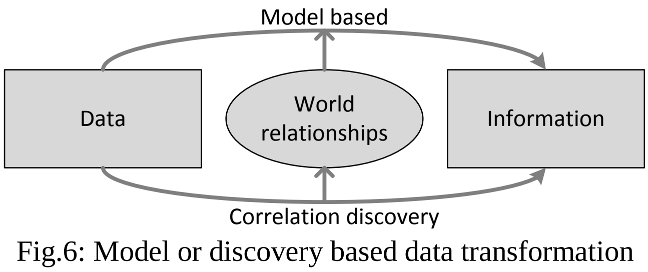

Fig.6: Model or discovery based data transformation

Model based extraction has strengths in having more possibilities for systematic analysis of data, including scientific hypothesis testing. The drawback is that the model or hypothesis must exist before the model can be used. Discovery based methods has the potential of finding new models and relationships, but with the danger that the model is not in fact valid, but only occurs as a random or otherwise correlated relationship between data.

In our case, initial studies on data analytics has been based on various empirical ship performance and navigation models. However, these conventional mathematical models often fail to handle large-scale data sets due to system-model uncertainties, sensor noise and fault conditions and complex parameter interactions. In some situations, these models cannot be adapted to predict actual ship navigation situations that are described by large data sets. Machine intelligence (MI) and statistical analysis are often used in such situations to overcome these data handling challenges. MI and statistical analysis can therefore be the more attractive approach in big data handling.

MI techniques have been used for data classification and regression analysis in various research and industrial applications. This can also be used under big data applications in shipping. Statistical techniques are often used to determine the relationships among the parameters in big data sets, e.g. through principal component analysis. These methods can be combined so that the parameter relationships among the classified data sets are investigated by statistics and the regression tools are used on these relationships to extract conclusions.

5. Operational time scale of big data applications

The new big data applications can be used in many different time scales from on-line decision support tools to long-term fleet statistics analysis. One can also have loops within this framework where results of long-term analysis is used to update control parameters in on-line decision support. We will not go into details on these issues in this paper, but only highlight some of the possibilities.

5.1 On-line ship decision support

Examples of more or less real-time applications are trim optimization, heavy weather decision support, engine optimization systems and so on. These will typically analyse immediate or short-term historical data to give decision support on current optimization possibilities, related to fuel consumption, ship movements or other ship parameters.

5.2 Ship performance optimization

Other applications will analyse data over longer periods to determine more general optimization possibilities. In addition to best practice operational guidance, this can also include technical condition monitoring and maintenance planning. One example of such a service is technical condition monitoring which uses monthly controlled measurements to assess the technical condition of ship machinery, *Rødseth et al. (2007).*

5.3 Fleet optimization

Even larger data sets can be used to compare performance between similar ships over longer periods. This can typically be used to determine statistical operational constraints for certain routes or more general best practice suggestions over the fleet. Other possibilities are analysis of technical performance of various technical and operational procedures, such as maintenance strategies, voyage optimization, hull coating performance etc. The Shipping KPI project is one example of such a service, *Rialland et al. (2014).*

5.4 Predictive analysis

Other interesting possibilities that opens up in a big data perspective is to use historical data to determine characteristics of new ships or new operational principles. Virtual prototyping and simulation of new ship designs in known and historically verified operational conditions is one interesting application of this, *Fathi et al. (2013).*

6. Conclusions

Big data processing for ships and shipping can tap a number of so far underused data sources and have a great promise in establishing more efficient technical solutions and operational regimes for ships and shipping. On the other hand, there are a number of error sources that can disturb the analytical processes and lead to erroneous conclusions. Therefore, various online and offline tools and techniques need to be developed to overcome such situations. This is particularly important for systems that operate on-line and that may be implanted as a part of modern integrated bridge systems. Systems using data collected over longer periods have more possibilities for filtering out bad data points through use of larger sets of historical data and using averaging over longer time periods to remove spurious bad points. However, even for online and real-time data handling there are new methods emerging that may promise better identification of sensor and acquisition fault situations. The quality of the data sets can be improved and instantaneous vessel performance can be observed at a much higher level of detail and with fewer errors.

Offline data handling may require additional data management tools to save, search and recall the required data sets, where the overall vessel and fleet performance in detail can be analysed. Therefore, the development of both data management and analytics will play an important role in the field of big data in shipping.

Connectivity between ship and shore is still a factor for some ships, but this is getting less of a problem as communication technology develops and prices of bandwidth decreases. Even with limitations in connectivity, there are many promising applications for big data applications in ship and fleet analysis.

A major cost-increasing factor for big data applications on ships is the lack of technical standards for interfacing to the different data sources. This means that ad hoc and costly interface solutions have to be developed for each ship. In addition, it will also be costly to configure databases, create system models and link the different measurements to the relevant phenomena. Work is under way in this area, but this work need more support from owners, yards, system manufacturers and scientists.
Another main factor for successful application of big data applications in the maritime field is to have a good understanding of the possibilities and limitations of the different data sources and the technologies that are available to handle the different problems.

References

DE MAURO, A.; GRECO, M.; GRIMALDI, M. (2015), *What is big data? A consensual definition and a review of key research topic,* AIP Conf. Proc. 1644, pp.97–104.

FATHI, D.E.; GRIMSTAD, A.; JOHNSEN, T.A.; NOWAK, M.P.; STÅLHANE, M. (2013), *Integrated decision support approach for ship design,* IEEE Oceans, Bergen

HERMANN, M.; PENTEK, T.; OTTO, B. (2015), *Design principles for Industrie 4.0 scenarios: a literature review,* Report, TU Dortmund

IEC (2007-2015), *Maritime navigation and radiocommunication equipment and systems - Digital interfaces*, (Part 1 and 2: Serial line, Part 450 Ethernet; Part 460 Safety and security), IEC 61162

IEC (2015), *Maritime navigation and radiocommunication equipment and systems - Digital interfaces - Part 460: Multiple talkers and multiple listeners - Ethernet interconnection - Safety and security,* IEC 61162-460

IMO (2014), Sub-committee on Navigation, Communications and Search and Rescue, Report to the Maritime Safety Committee. NCSR 1/28, Annex 7: *Draft e-Navigation Strategy Implementation Plan*

ISO (2015a), ISO/NP 19847 *Shipboard data servers to share field data on the sea*, ISO/TC 8 / SC 6N 359

ISO (2015b), ISO/NP 19848 *Standard data for shipboard machinery and equipment of ship,* ISO/TC 8 / SC 6N 360

PERERA L.P.; MO, B. (2016). *Data compression of ship performance and navigation information under deep learning*, 35th Int. Conf. Ocean, Offshore and Arctic Engineering (OMAE 2016), Busan

RIALLAND, A.; NESHEIM, D.A.; NORBECK, J.A.; RØDSETH, Ø.J. (2014), *Performance-based ship management contracts using the Shipping KPI standard*, WMU J. Maritime Affairs 13(2), pp.191-206

RØDSETH, Ø.J.; STEINBACH, C.; MO, B (2007), *The use of technical condition indices in ship maintenance planning and the monitoring of the ship's safety condition*, Int. Symp. Maritime, Safety, Security and Environmental Protection, pp. 20-21.

RØDSETH, Ø.J.; KVAMSTAD-LERVOLD, B.; HO, T.D. (2015), *In-situ performance analysis of satellite communication in the high north*, IEEE Oceans, Genova

RØDSETH, Ø.J.; LEE, K. (2015). *Secure communication for e-navigation and remote control of unmanned ships*, 14th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Ulrichshusen

RØDSETH, Ø.J. (2016), *Integrating IEC and ISO information models into the S-100 Common Maritime Data Structure*, E-navigation Underway International, Copenhagen