Key benefits are now being sought from Big Data in the maritime sector. Yet, as Maerospace's Eric Meger discusses, growing information is also presenting major obstacles

Big Data at sea

Maritime stakeholders have never had access to more data than they do today. Table 1 shows a sample of why we are truly in a Big Data world. Over the past decade, the Automatic Identification System (AIS) introduced digital ship tracking around ports and satellite AIS has since extended this to the globe. Over two billion location reports are transmitted every day to be detected by multiple systems and vendors. However, only a fraction of these are detected by each system and it remains a challenge to figure out what data actually matters.

Intelligence?

Data does not equal intelligence. This year we will see a huge array of data sources from companies like Orbcomm, exactEarth, Spire, and dozens of coastal systems. Before these data sources arrived, activity at sea remained hidden. Today we have the opposite problem, intelligent decision making is flooded with irrelevant data. Fortunately, advanced and predictive analytics are beginning to make sense of the chaos.

Every user group has distinct information requirements beyond raw data. Whilst navies need situation awareness, coast guards need interdiction support and fleet owners need confirmation of true positions of their assets and ways to optimise costs and revenues. In addition commodity traders need cargo status and import/export predictions, and shippers and consignees need to know when their cargo will be late. Each of these applications depends on the timely and accurate location of ships, but only if it is processed into actionable information.

Big data, big errors

Crucial limitations of the raw data are not obvious. Any user would look at a global map with AIS data (see Fig. 1) and conclude that we finally know where all the ships are. As valuable as the data is, it is still old, wrong and incomplete. As reported in the 16th edition of *Pan European Networks: Government,* the data from all these sources must be enhanced by analytics to make it current and complete. Otherwise, crucial data is not even considered in the decision process.

Eric Meger

For example, an assessment was recently carried out looking at tankers within a 3,500 mile (~5,630km) radius of Singapore (generally from the Persian Gulf to Japan) where it was found that 10.6% of tankers are missing from a typical display. Of the 89.4% remaining that were displayed, they were shown in old locations resulting in incorrect positions by over 100 miles. As a result, these statistics therefore mean that over 500,000 tonnes of oil at sea is carried on ships that are essentially invisible to users.

Intelligence key

Equally important to decision making is the need to separate normal behaviour from anomalous behaviour. When something unexpected happens with a vessel, timely detection is essential to making good decisions.

Anomalies occur for many reasons and can be classified as: intrinsic (e.g. with a latitude greater than 90°); contextual (e.g. a ship name does not match the Maritime Mobile Service

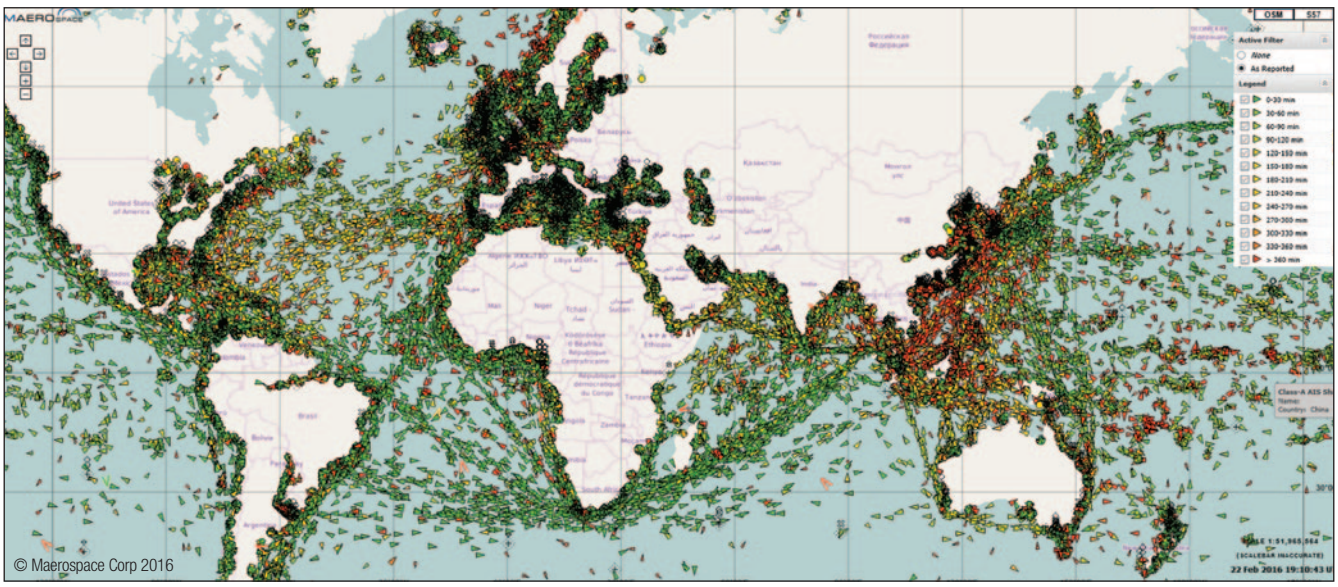

Fig. 1 Ships at sea

Maritime Big Data: daily activity to analyse

- 150,000 ships;
- Two billion transmissions;
- 100 million position reports;
- Thousands of coastal stations;
- Five million tonnes of oil at sea every day;
- 200 satellites;
- 243 flag states;
- 30 million containers;
- US $36bn (~€33bn) of exports carried per day; and
- 260,000 loaded containers at sea.

Identity code); and behavioural (e.g. a ship moving away from a destination port or toward a suspicious rendezvous). Detecting these anomalies and using them to make better decisions is at the heart of applying Big Data principles to the maritime domain.

Two approaches

Decision support in this context has two approaches: query-focused and monitor/alert-focused. Query-focused approaches are far more common in Big Data systems. Data is acquired by the system to be processed into precomputed results and stored in data warehouses powered by tools, such as 'Hadoop'. The goal is to make it possible for analysts to ask questions of the data and receive timely responses. The analyst works on varied problems and needs fast access to huge volumes of data to tap into whenever a new question emerges. This approach works best when the problems and questions change often and brand new insights are needed frequently.

By contrast monitoring/alerting systems depend on real-time data flows and anomaly detection. The questions to be addressed are much more standardised and operational. What ships will arrive at my port in the next 72 hours? Are there any ships about to enter a restricted area? When will my cargo arrive at its destination? Are there any delays in my supply chain? Users want real-time situation awareness and to be alerted when there is a material change, rather than the answers to brand new questions. For this type of decision support, real-time operations are essential and are best architected in different ways using different technologies.

Inaccurate ship location data can lead to difficulties in decision making

The maritime world is characterised by moving ships and cargo. As a result, a great many applications of tracking data fall into the monitor/alert category. Users need to be able to tag and track vessels of interest; monitor the status of regions, ships or cargo; be alerted to unexpected changes; and drill down to verify anomalies and find causes.

Knowing where to focus attention is the core benefit of such predictive analytics systems. Although some decisions can be automated, the vast majority require human judgement, for example: 'So the ship will be late … how late?'; 'Does that matter to me right now?'; 'Can I do anything to mitigate the delay?'; or 'Should I do anything to be ready in case the delay gets worse?'.

Right choice, right solution

Is this a distinction without a difference? When discussing complex sensor networks of any kind, the challenge is not just about software. The end-to-end system architecture is intimately involved in creating the information to be acted upon (i.e. intelligence). Data that does not lead to new decisions is just entertainment.

Organisations affected by the movements of ships and cargo should look carefully at their application and consider the most appropriate architecture; don't be seduced by buzzwords and the latest technology. Hadoop clusters, software as service models, data warehousing, Apache Spark and other terms may be important for your application, or they may not.

Whether your concern is with maritime security, optimising operations, or knowing about cargo arrivals, today's Big Data has to be turned into actionable intelligence before it is more than just a 'cool' display. That transformation requires a deep understanding of the true quality of the incoming data and a solution that is optimised properly for either querying or monitoring/alerting. Getting expert advice and continually asking how the solution will help make better decisions will keep you on the right track.

Eric Meger
Maerospace Corporation

www.maerospace.com