Overview of AIS dataset

Introduction

Maritime transportation plays a central role in global trade since around 80% of the global trade by volume is carried by sea. This growth of global trade increases the demand for more vessels with larger cargo capacity [1] and requires different solutions to ensure safety at sea.

The development of the Automatic Identification System (AIS), through which nearby ships exchange information on their location and navigation status frequently utilizing a radio signal, facilitates safety and traffic management. AIS data can potentially enhance and/or replace manual lookout and surveillance at sea and port [1].

AIS data is collected at several places around the world from land-based stations and satellites. Allowing for real-time geo-tracking and identification for equipped vessels, the AIS data promises to map and describe certain marine human activities. So, AIS data provides a big data source of unrivaled quality above and beyond its original application for collision avoidance.

Nowadays, AIS information is used to serve various purposes and facilitates the work of people in various occupations, such as Port Authorities, Ship Owners, Managers, Agents and Brokers, Vessels’ crews, Rescue teams, Hotels and Tour operators, Passengers or recreational sailors, Environmental Protection agents, Researchers [2].

The number of applications for this data is enormous, and some apps are already built based on it. For official statistics, AIS data could be used, inter alia, as the backbone for the maritime transport statistics – to achieve overall goals of sustainable transport which aims to secure inclusive, resilient and low-carbon transport solutions. UNCTAD, for example, already provides statistics on port calls building on AIS data provided by MarineTraffic. In addition, research on the use of AIS for official statistics is part of the ESSnet Big Data project [3].

Description

AIS was originally developed by IMO (International Maritime Organization) in 2004 as an outcome of amendments to the International Convention SOLAS (Safety of Life at Sea) in 2002. IMO is UN agency established in 1948 aims to improve maritime safety and security.

IMO’s mission statement:

"The mission of the International Maritime Organization (IMO) as a United Nations specialized agency is to promote safe, secure, environmentally sound, efficient and sustainable shipping through cooperation. This will be accomplished by adopting the highest practicable standards of maritime safety and security, efficiency of navigation and prevention and control of pollution from ships, as well as through consideration of the related legal matters and effective implementation of IMO’s instruments with a view to their universal and uniform application."

There are various public references that describe in detail about AIS dataset. See description from Marine Traffic FAQ webpage.

The Automatic Identification System (AIS) is an automated, autonomous tracking system which is extensively used in the maritime world for the exchange of navigational information between AIS-equipped terminals. Thanks to it, static and dynamic vessel information can be electronically exchanged between AIS-receiving stations (onboard, ashore or satellite).


Content of AIS

AIS is a collaborative, self-reporting system, which allows vessels to periodically broadcast their identity, navigation and position data and other characteristics. AIS uses Global Positioning Systems (GPS) in conjunction with shipboard sensors and digital Very High Frequency (VHF) radio communication equipment to automatically exchange navigation information electronically [4].

Currently, marine vessels use AIS in coordination with VTS to monitor vessel location and movement primarily for traffic management, collision avoidance, and other safety applications [4].

AIS has been made compulsory for international commercial ships with gross tonnage (GT) of 300 or more tonnes, and all passenger ships regardless of size. AIS reporting requirements are described in Regulation 19 of Chapter V of the International Convention for the Safety of Life at Sea (SOLAS) [5].

The technology works with transponders which automatically broadcast information at regular intervals by VHF transmitter. Vessel identifiers such as the vessel name and VHF call sign are programmed in during initial equipment installation and are included in the transmittal along with location information originating from the ship’s global navigation satellite system receiver and gyrocompass [4].

There are 27 different AIS messages types containing different types of information. AIS broadcasts voyage related information (including ship location, speed, course, heading, rate of turn, destination and estimated arrival time) as well as static information (including ship name, ship MMSI ID, message ID, ship type, ship size). Dynamic information such as the positional aspects (current latitude and longitude) is automatically transmitted, depending on the vessels’ speed and course. While the vessel is on the move this information is transmitted every 2 to 10 seconds and while a vessel is anchored every 3 minutes.

AIS data comprised of the following three information categories:

1. Static data (information on ship characteristics)
2. Dynamic data (information on ship position and movements) and
3. Voyage-related data (information on a current voyage).
The terrestrial-based AIS station or onboard transceiver could typically cover about 15-20 nautical miles (nm), depending on many factors such as transceiver type/location and weather conditions. In open seas, satellite-based receivers provide an efficient supplement while terrestrial stations are out of range [6].

In the following table, the typical AIS message types are presented [7].

### Table 1. TYPICAL AIS MESSAGE TYPES

<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>Position report</td>
<td>(Assigned) Scheduled position report, or response to interrogation</td>
</tr>
<tr>
<td>4</td>
<td>Base station report</td>
<td>Position, UTC, date and current slot number of base station</td>
</tr>
<tr>
<td>5</td>
<td>Static and voyage related data</td>
<td>Scheduled static and voyage related vessel data report</td>
</tr>
<tr>
<td>6-8</td>
<td>Binary related message</td>
<td>Binary communication</td>
</tr>
<tr>
<td>10-11</td>
<td>UTC related message</td>
<td>Request/Response-to UTC/date</td>
</tr>
<tr>
<td>12-14</td>
<td>Safety-related message</td>
<td>Communication/Acknowledgement/Broadcast safety data</td>
</tr>
<tr>
<td>15</td>
<td>Interrogation</td>
<td>Request for special response</td>
</tr>
<tr>
<td>21</td>
<td>Aids-to-navigation report</td>
<td>Position and status report for aids-to-navigation</td>
</tr>
<tr>
<td>27</td>
<td>Position report for long-range applications</td>
<td>Class A and Class B “SO” shipborne mobile equipment outside base station coverage</td>
</tr>
</tbody>
</table>
AIS Vessel Type and Group Codes used by the Marine Cadastre Project (2018)

Source
AIS message is regularly broadcasted by AIS responder which is mandatory for all passenger and cargo vessels over 299 Gross Tonnage (GT) that travel internationally. As for inland waterway – which is not covered by maritime regulations – in European countries this is required (although the level of information to be put in can differ between countries: for example not all countries require “Destination” due to business privacy). These messages can be received by surrounding ships, terrestrial-based AIS stations located along coast and satellite AIS stations. An individual with proper equipment can also receive and record AIS messages assuming that they are within the coverage range of AIS responder. Therefore, as a consequence, AIS messages are considered in the public domain and freely available. Nevertheless, AIS messages may track certain individuals living on their ships (i.e., yacht or ships on inland waterway) and it would cause some privacy concerns.

Data provider(s)
AIS messages can be seen as another type of sensor data that is regularly transmitted as part of maritime and inland waterway navigation. Individual AIS messages may contain useful information (i.e., static AIS information) however the most frequently used information is position report. This implies that large number of AIS messages must be recorded and stored and shorter interval transmission may yield more relevant data for analysis (i.e., AIS messages collected by Coast Guard). The increased use and demand of AIS data have enabled the rise of AIS data provider(s) which collect, store, analyze and disseminate AIS data. These data providers may be commercials and not-for-profit institutions. See below the cons and pros below in ESSNet project Big Data project.

Coverage Limitation
Due to the nature of AIS messages transmission (broadcasted by IAS responder), compiled AIS data by data providers/aggregators may be incomplete coverage in terms of time series, spatial or type of vessel. An analyst should take this coverage limitation when applying AIS data to specific purposes. For example, if someone analyses the time series impact of fishery in a certain region, they should ensure that AIS data covers adequate representation of fishing boats in the target area (and adequate time series). See below the cons and pros below in the ESSnet project Big Data project.

There are many AIS data providers which can provide real-time and historical terrestrial or satellite AIS data. However, the AIS data quality differs between the providers. There are private, public, commercial or free data providers. During ESSnet Big Data project, the coverage issue was investigated comparing commercial European dataset to public national AIS data Denmark, Greece and Poland and satellite data from Luxspace (LS). For the process of data comparison, a common assumption and approach was adopted. A common reference frame of ships was built to filter out existing maritime ships only [8]. Then, we defined the areas to be compared by selecting ports for the best coverage of AIS base stations. For each port, a rectangular area (latitude, longitude) was defined. A code developed in Scala and processed by Apache Spark was used. The results were saved in a Distributed File System of Hadoop at Sandbox. This code identifies the ships (based on MMSI) from the reference frame of ships in a defined area(s) for a defined time period. Then, it counts the number of appearances of each MMSI in that/those area(s) during that time period. It was run for both data sources [9], [10].

Quality consideration
When investigating the quality of AIS data it is important to keep in mind that:

- AIS is a radio signal, parts of the messages can get lost or scrambled due to factors such as meteorology or magnetics.
- Messages are transmitted encoded. As a result, an error in one transmitted 'byte' can result in an error in one or multiple fields in the decrypted message. Most of the time, these errors are detectable as the result yields an invalid variable, but sometimes they result in invalid variables. For instance, coincidentally the resulting MMSI (Maritime Mobile Service Identity) can be a technically valid, but incorrect MMSI, resulting from an erroneous detection. These errors can arise for every variable, so this can, for example, result in erroneous latitude and longitude, yielding faulty locations that are quite far away from the actual location of the ship. In turn, this can result in a very high journey distance of the ship.
- Receivers have timeslots in which data is received. In busy areas with many ships, not all data from all ships may fit into this time slot. Resulting in the loss of data on some ships in that time slot.
- Ships can turn off their AIS transponder resulting in the disappearance of a ship.
- AIS was intended originally for safety at sea, to warn nearby ships. As it was not meant for producing statistics, the (static) variables entered manually by the shippers are not always reliable.
• AIS receivers on land can only pick up signals within the range of about 40 sea miles. Therefore, land receivers have very limited coverage of signals transmitted from the sea which results in loss of information of ships on the open sea.

Moreover, AIS data may contain:

• Technical errors - related to dynamic data such as the position of ship, speed, course, rotation which comes from AIS device (sensors, cables, and antenna)
• Human errors or intentionally misreported data – related to static data (MMSI, IMO number, ship’s name, call sign, type, length) or voyage data (draught, destination) which are manually entered in the AIS device

While working with AIS data in ESSNet project, researchers noticed a lot of noise that prevented proper analysis. This was the basis for us to investigate the quality of these data. To provide an overview of different aspects of the quality of the DZ AIS data, we used a preliminary framework for national statistical offices to conceptualize the quality of big data.

This preliminary framework was developed building on dimensions and concepts from existing statistical data quality frameworks. It provides a structured view of quality at three phases of the business process:

• Input – acquisition, or pre-acquisition analysis of the data;
• Throughput – transformation, manipulation and analysis of the data;
• Output – the reporting of quality with statistical outputs derived from big data sources.

This Big Data Quality framework is a deliverable of the Big Data Quality Task Team from the UNECE/HLG project, *The Role of Big Data in the Modernisation of Statistical Production* [11].

References


