

# **AegeaNET Syros AIS Dataset for Vessel Traffic Monitoring**

Troupiotis-Kapeliaris A., Spiliopoulos G., and Zissis D.

*Intelligent Transportation Systems Lab, University of the Aegean*

---

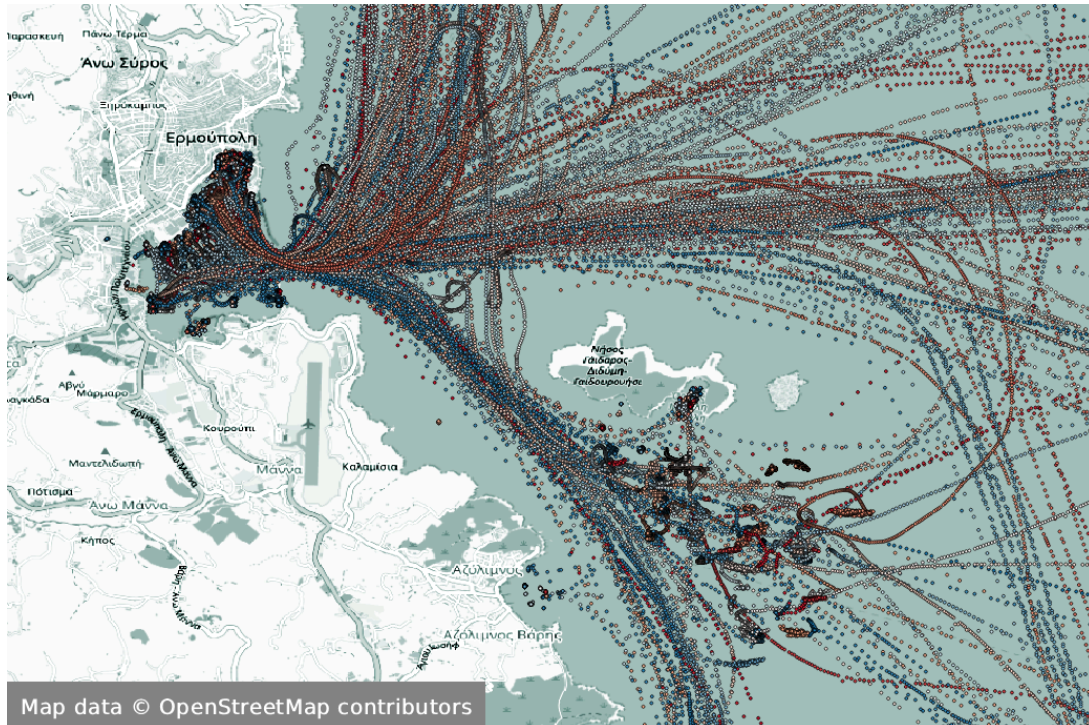
The Automatic Identification System (AIS) is a collaborative monitoring system designed for collision avoidance in maritime navigation. Originally designed for larger vessels, it is now used by different types of vehicles, from tankers to fishing boats and pleasure crafts. Essentially AIS works on the principle that vessels regularly transmit their current position, along with other characteristics like speed-over-ground, heading and a unique identifier<sup>1</sup>. The abundance of AIS messages have allowed for Geographic Information System (GIS) applications and trajectory analytics to model and describe vessel traffic, extract useful patterns and provide predictions on future trends.

To that end, the **AegeaNET** is an open sensor network composed of AIS and ADS-B receivers, strategically deployed throughout the Aegean Sea. AegeaNET is an academic initiative by the **Intelligent Transportation Systems Lab (ITS Lab)** of the University of the Aegean (Greece). This network facilitates real-time tracking of maritime activity, providing critical data essential for navigation and safety.

We present a publicly accessible dataset originating from a receiver on the island of Syros, covering a large area in the middle of the Aegean sea, including most of the Cyclades. The duration of the dataset is a full three (3) month period, from July to September 2024. The positional messages of all vessels are included in a single AIS file, sorted in ascending order using their respective timestamp. An initial cleaning was performed to remove messages with erroneous or missing critical fields. Finally, for the purpose of anonymity, a masking mechanism was applied on top of the vessel identification field (MMSI). You can find more on the masking process and the dataset characteristics in the following sections.

---

<sup>1</sup> More on AIS data and its applications can be found in the following article  
Yang, Dong, et al. "How big data enriches maritime research—a critical review of Automatic Identification System (AIS) data applications." *Transport reviews* 39.6 (2019): 755-773.  
<https://doi.org/10.1080/01441647.2019.1649315>



*The dataset is comprised of vessel tracks from a terrestrial AIS station, part of the AegeNET network, located on the island of Syros, Greece.<sup>2</sup>*

## MMSI masking

To maintain the anonymity of the dataset, a hashing mechanism is employed, producing deterministic results while preventing reverse engineering of the original identifiers through a randomly selected salt.

Generally the Maritime Mobile Service Identity (MMSI) is a 9-digit number to uniquely identify a vessel or station. For the creation of an MMSI number several factors are considered:

- The type of transmission: from a vessel, a coastal station, a ship group, a SAR transponder, or other sources.
- The country of origin of the vessel, represented by the Maritime Identification Digits -- MID (you can find more on this [here](#)).

Additionally, some MMSI values are reserved or not used by vessels, including:

- The MMSI must be exactly 9 digits long.
- 123456789 and 111111111 are reserved for testing purposes.

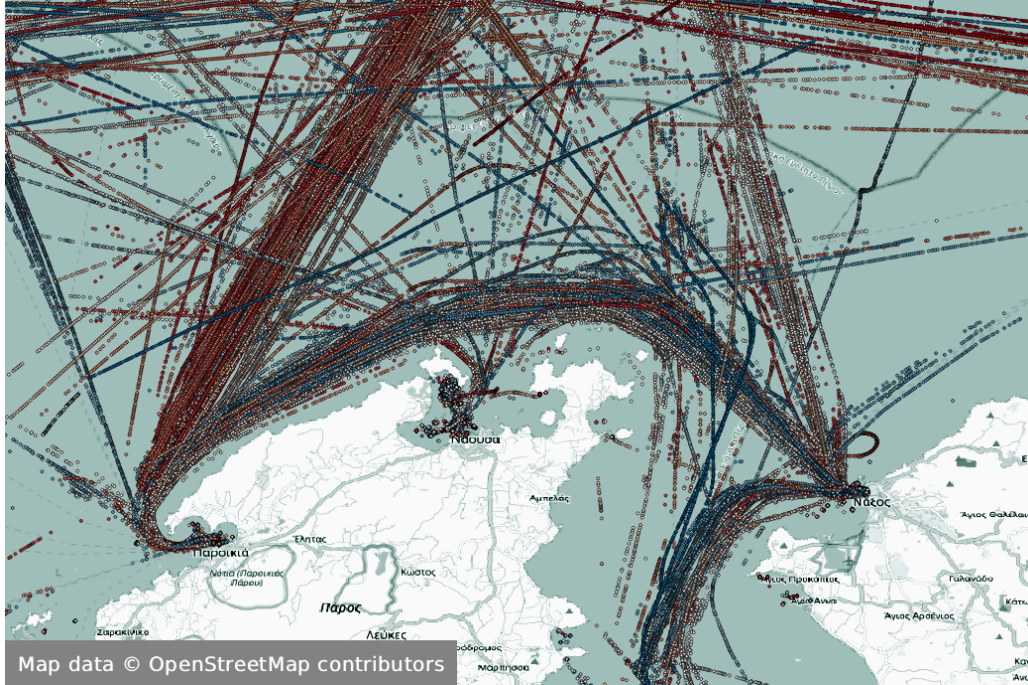
<sup>2</sup> Map data is available under the Open Database License.  
<https://www.openstreetmap.org/copyright>

- Numbers with repeated digits (e.g., 222222222, 333333333, etc.) are often used for testing.
- All regular vessel MMSIs have a prefix between 200-799, corresponding to their country of origin.

Following this logic, the masking mechanism returns only ids with a prefix between 200-799, avoids reserved numbers (as mentioned above) and remains deterministic while anonymizing the original MMSI. Finally, the resulting number does not retain the original country information, effectively hiding this detail.

Description of the dataset fields

Field	Description	Notes
MMSI	Vessel identification number	Masked, 1848 in total
UTC	Coordinated Universal Time of message	ISO 8601 format (UTC/Zulu), e.g., "2024-12-31T23:59 :59Z "
TIMESTAMP	Time of the AIS message in Unix epoch, based on UTC	In seconds
LON	Longitude of the vessel reported position in decimal degrees	Range: (24.6, 27.3)
LAT	Latitude of the vessel reported position in decimal degrees	Range: (35.8, 38.1)
SOG	Vessel's Speed-over-Ground	In knots
COG	Vessel's Course-over-ground	In degrees
HEADING	True heading of the vessel's bow	In degrees
STATUS	Navigational status code	Range: [0, 15]
STATUS_DESC	Description of navigational status	<a href="#">Reference</a>



The dataset allows for the extraction of common routes and patterns of movement in the Cyclades area, such as the traffic between the ports of Paros and Naxos.<sup>3</sup>

## License

The dataset is released using the following license:

*Creative Commons Attribution  
Non Commercial No Derivatives 4.0 International*



## Acknowledgements

This work was partially supported by the MUSIT Project through the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101182585; by the European Union's Horizon Europe programme under grant agreement No. 101092749 (project Critical Action Planning over Extreme-Scale Data – CREXDATA); and by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union under the NextGenerationEU Programme.

<sup>3</sup> Map data is available under the Open Database License.  
<https://www.openstreetmap.org/copyright>