2018 IEEE 5th International Conference on Data Science and Advanced Analytics

Data Fusion to Describe and Quantify Search and Rescue Operations in the Mediterranean Sea

Katherine Hoffmann Pham
Stern School of Business
New York University
New York, USA
khof@nyu.edu

Jeremy Boy
Global Pulse
United Nations
New York, USA
jeremy@unglobalpulse.org

Miguel Luengo-Oroz
Global Pulse
United Nations
New York, USA
miguel@unglobalpulse.org

***Abstract*—The Mediterranean Sea is the stage of one of the biggest humanitarian crises to affect Europe. Since 2014, thousands of migrants and refugees have died or gone missing in dangerous attempts to cross into the continent. However, there is relatively little structured information available on how they attempt the crossing. Such information could be used to better target maritime rescue efforts or to anticipate smuggling patterns, which could potentially save lives. In this article, we provide an overview of data sources available for the study of migration in the Central Mediterranean. We describe how these data can be structured, combined, and analyzed to provide quantitative insights on the situation in the region. We define a *quantified rescue* framework for fusing different data sources around individual rescue operations, and we explore the potential of machine learning to perform automated rescue detection based on vessel trajectory information. We conclude with technical research questions, and potential policy and operational implications related to the use of these data sources.**

***Index Terms*—Social good, computational social science, migration, trajectory data, machine learning, artificial intelligence**

I. Introduction and Context

The world is experiencing an unprecedented wave of maritime migration. At the peak of the European refugee crisis in 2015, the United Nations Refugee Agency (UNHCR) estimated that over one million people had arrived in Europe by sea [1]. This included refugees escaping political persecution and conflict, as well as economic migrants hoping to find a better life and to support dependents back home [2]. Over the course of 2016, arrivals shifted away from the Eastern Mediterranean (the Greece–Turkey route) toward the Central Mediterranean (the Libya–Italy route). They are now moving toward the Western Mediterranean (the Morocco–Spain route) after Italy's recent refusals to accept migrant and refugee ships, and the adoption of a more aggressive role by the Libyan Coast Guard [3]–[6]. In this article, we focus on the Central Mediterranean route, as it has been by far the deadliest since the beginning of the crisis. The International Organization for Migration (IOM) estimates that over 14,000 people have died while trying to cross along this route since 2014 [7].

Search and rescue (SAR) missions in the Central Mediterranean were initially led by the Italian authorities as part of *Operation Mare Nostrum*, which took place from October 2013 to October 2014. This operation was replaced in November 2014 by the joint EU *Operation Triton*, which was later supplemented by the anti-smuggling *Operation Sophia* in June 2015. Since mid-2015, a number of non-governmental organizations (NGOs) have joined the effort, launching privately-funded rescue missions operating off the Libyan coast and collectively patrolling the sea in search of ships in distress. Overall, hundreds of SAR missions have been conducted to assist tens of thousands of people, often in numerous operations per day.

The Libyan coast is a transit hub for migrants and refugees from countries as diverse as Eritrea, Nigeria, and Bangladesh. While some leave their countries of origin with the deliberate goal of reaching Europe, others reside in Libya for up to several years before being displaced by violence, instability, or human trafficking [2].

From holding centers along the coast, smugglers pack vessels with migrants and direct them toward international waters. Rubber rafts are generally preferred, as they are inexpensive to replace if lost or destroyed, and they can carry as many as 200 people at once. Re-purposed wooden fishing boats are also used, which have been known to carry over 800 people at a time. Overcrowded and poorly equipped, these vessels have little chance of arriving in Italy or Malta on their own. They usually proceed until they run out of fuel, begin to leak, or are rescued. The more fortunate ones are equipped with a satellite phone, which can be used to request help from the Maritime Rescue Coordination Center (MRCC) in Rome. Others depend solely on being spotted by passing ships and SAR aircraft. Less fortunate vessels are found only days later, and some simply disappear at sea.

Despite this substantial flow of human traffic and extensive SAR efforts, there is little systematic analysis of maritime movement patterns, beyond the processing of migrants who arrive on shore. Information is decentralized across a diverse range of actors, including: Italian authorities such as the Navy and the Coast Guard; NGOs operating rescue boats, such as Médecins Sans Frontières (MSF) and Migrant Offshore Aid Station (MOAS); and international migrant and refugee organizations such as IOM and UNHCR. A better understanding of movement patterns may be helpful in anticipating arrivals; identifying areas where increased search and rescue patrols are needed; galvanizing international governments to act on the

978-1-5386-5090-5/18/$31.00 ©2018 IEEE
DOI 10.1109/DSAA.2018.00066

current crisis; and understanding the ever-evolving situation.

Here, we describe our efforts to collect and combine data on rescues in the Central Mediterranean. We compile "ready-made" and "custom-made" data sources (as proposed in [8]) to provide a comprehensive view of the situation. *Ready-made* data refers to large-scale, big data sources that are adapted for purposes other than the one(s) they are originally collected for. *Custom-made* data refers to data that are expressly collected to study the phenomenon of interest.

This article is based on work we have previously published in [9], and is organized as follows. In Section II, we describe two ready-made big data sources—Automatic Identification System (AIS) data on ship movements, and Broadcast Warning data on maritime distress calls—which, to our knowledge, have not yet been used for large-scale quantitative analyses of rescue efforts. We also present information on NGO activities found in social media. In Section III, we describe custom-made official and unofficial data sources, including IOM and UNHCR statistics, as well as information compiled by journalists. In Section IV, we introduce the *quantified rescue*, a framework for fusing these different data sources to provide a holistic perspective on rescue operations, and in Section V, we explore how machine learning might support automated detection of rescue operations. Finally, in Sections VI and VII we conclude with a discussion on policy and operational implications of this work, as well as on open technical challenges.

II. "Ready-made" Big Data Sources

One immediate source of data on vessel movement comes from Automatic Identification Systems (AIS), which ships use to track and report their positions in order to avoid collisions. With AIS, it is possible to monitor e.g., the behavior of rescue ships off the coast of Libya, or to identify the involvement of merchant ships in SAR operations. Broadcast Warnings can then be used for additional context. These data contain information about vessels requesting assistance at sea. Further, as a growing number of NGOs have entered the response, Twitter provides a rich source of narrative, photographic, and video information on the crisis.

A. Automatic Identification System Data

The richest source of information on sea movement comes from vessel AIS data. All cargo ships over 500 tons, all internationally-traveling ships over 300 tons, and all passenger ships are required to be equipped with an AIS transceiver capable of transmitting (1) *static information*, such as self-reported ship name, size, country of registration, and up to three ship identifiers[1]; and (2) *dynamic information*, such as latitude, longitude, course over ground, speed, and the self-reported destination of the ship [10], [11].

[1] A 9-digit Maritime Mobile Service Identity (MMSI), a 7-digit International Maritime Organization (IMO) number, and a variable-length alphanumeric call sign.

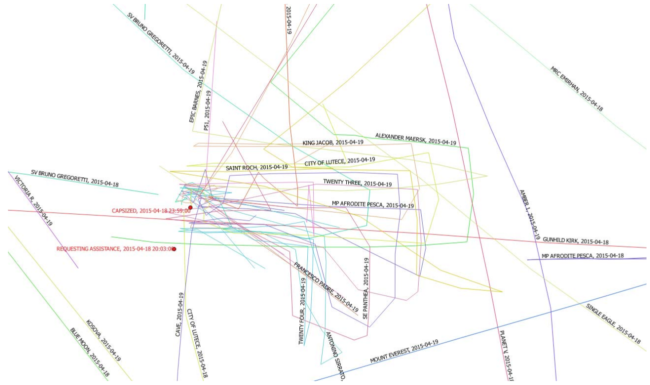


Fig. 1. An example of low-frequency AIS trajectory data illustrating shipping traffic around an incident on April 18, 2015, in which a commercial ship (King Jacob) collided with a migrant boat, leading to an estimated 800 deaths [19].

AIS transmissions can be collected by anyone equipped with a suitable land or satellite-based receiver. As a result, various organizations aggregate, share, and resell the data. Leading providers include: MarineTraffic, ExactEarth, OrbComm, FleetMon, VesselFinder, and VesselTracker [12]–[17]. For a full review, see [18]. Working with data from MarineTraffic, ExactEarth, and OrbComm, we have found that AIS information can be differentiated along four key axes.

First, AIS readings vary in temporal frequency. Different providers offer ship position readings at frequencies ranging between approximately 2 minutes and one hour. There can also be variation *within* data sources, which appears to be correlated with ship speed and activity type.

Second, AIS readings vary in geospatial coverage. Land-based receivers typically detect offshore signals within 15–20 nautical miles, though their range can extend as far as 60 nautical miles [20]. For further coverage, satellite receivers are necessary. Coverage can vary by AIS provider, and is typically considered a premium feature.

Third, AIS providers sample their data for purchase in different ways. Companies vary in the availability of historical (as opposed to real-time) data, and in how data can be selected from their databases. For example, data can be obtained at the ship level; for a latitude/longitude bounding box; or for the whole world. Although many providers tag trajectories with calculated features (e.g., they may infer when a ship is visiting a port) we have found variation in the cost and feasibility of rendering data according to more advanced queries (e.g., selecting all AIS readings within a two-week timeframe for all ships passing through a given port).

Fourth, AIS providers vary in the extent to which their data are structured and machine-readable. Some provide data in a standard format that has been heavily pre-processed (e.g., ship-level readings with latitude, longitude, speed, course over ground, etc.). Others provide raw data, which requires substantial investment to structure and decode. In particular, AIS data are identified according to 27 different types, which generally correspond to data features such as the source (e.g., terrestrial or satellite), the reporting object (e.g., ship or base station), and whether the data are static or dynamic.[2] The data are transmitted in "encapsulated information sentences", which are strings coded according to a standard established by the National Marine Electronics Association: NMEA 0183. These coded strings substantially decrease the size of transmissions (and their required storage space), but they need to be individually decoded with the help of a library such as LibAIS [22], as they are not human-readable (e.g., `Km@@TJ0N?7>7n04`).

In addition to this decoding process (where necessary), typical preprocessing includes removing readings on land, and correcting identifiers mistakenly used for multiple ships (creating the impression that the same ship is in two locations at once), as well as multiple identifiers used for the same ship. More sophisticated preprocessing can also include: connecting points to create trajectories, imputing data to cover gaps between points, and compressing repetitive data into a more economical representation. For example, the "trajectories" of anchored ships may create the impression that a ship is traveling around a point, and can typically be replaced with two readings representing the start and the end of the anchored period.

Finally, AIS readings can sometimes be inaccurate or missing. For example, deceptive ships might turn off their AIS transceivers, or they may deliberately "spoof" their location. Further, the quality of self-reported information like ship name and destination can be variable. For example, some ships list destinations like "SAR zone". These data quality issues can present a significant risk to the use of AIS data; a review by Robards et al. (2016) found that some researchers drop 28–74% of records in their datasets [23]–[25].

AIS data have been used in a wide range of practical applications, including monitoring fishing activity and tracking maritime trade flows [26]–[29]. However, we are aware of only one publication that uses AIS data to track migration, with a limited sample of data [30]. We initially gathered AIS data in the hopes of identifying migrant vessels' trajectories as in [30], but found that such vessels were not detectable at scale. In practice, most migrant rafts and boats do not emit AIS signals, either because they are not equipped with transceivers, or because these transceivers are deliberately turned off or manipulated (as described above). Consequently, we shifted our focus to studying migration from the perspective of search and rescue activity. We located the Maritime Mobile Service Identities for a number of NGO rescue ships, as well as for several navy and coast guard ships.

B. Broadcast Warnings

Broadcast Warnings are short, text-based radio alerts used to update ships on nearby activities, dangers, and emergencies. The warnings contain semi-structured information on events

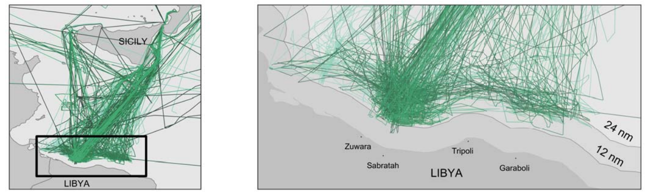


Fig. 2. Aggregate AIS data showing the high concentration of rescue vessel movements off the Libyan Coast.

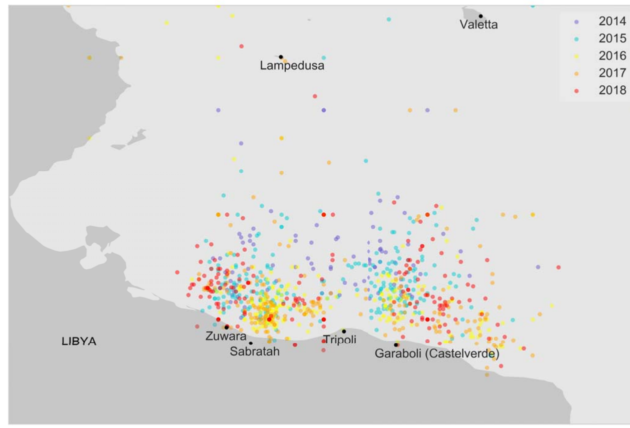


Fig. 3. Broadcast warnings off the coast of Libya, for years 2014–2018.

such as military exercises, objects adrift at sea, and vessels in distress. The text of the warning typically references the name of the sea or ocean; the closest country; the date, time, and location; and contact information for rescue coordination centers (if applicable). A sample warning is shown in Fig. 4. These warnings are produced by the World-Wide Navigational Warnings Service (WWNWS), and their raw text can be retrieved from the US National Geospatial-Intelligence Agency [31].

The primary difficulty in using Broadcast Warnings is filtering for records of interest. Broadcast Warnings are released for five different areas, with HYDROLANT warnings covering the Mediterranean and much of the Atlantic Ocean [32]. We have found it helpful to further filter these data by smaller geographic zones [33], and to restrict the results to keywords of interest.[3]

Another challenge involves obtaining structured features from the text. The wording of Broadcast Warnings is systematic enough that standard text processing methods apply. Regular expressions matching can be used to build features of interest. For example, we constructed variables for whether one or more vessels were in need of help, the number of people on board, the status of the boat (e.g., "requesting assistance",

[2] For our purposes, the relevant dynamic data appeared to be of types 1–3 (Class A position reports), 18–19 (Class B position reports), and 27 (long-range/satellite position reports). The class refers to the type of transceiver; Class A corresponds to larger commercial ships, while Class B typically corresponds to smaller commercial or recreational ships [21].

[3] Specifically, zones 52–56 cover the Mediterranean. We also limited our search to warnings mentioning a rescue organization such as the MRCC or the Libyan Coast Guard, the Eastern Mediterranean Sea, and the words "person", "people", and/or "assist". Finally, we eliminated warnings containing phrases like "cable repair", "hazardous operations", "fishing gear adrift", etc.

```
HYDROLANT 2050/2018 (56)
(Cancelled by HYDROLANT 2291/2018)
EASTERN MEDITERRANEAN SEA. LIBYA. DNC 09.
VESSELS REQUESTING ASSISTANCE, 200 PERSONS ON BOARD,
IN 33-06N 014-07E. VESSELS IN VICINITY REQUESTED
TO KEEP A SHARP LOOKOUT, ASSIST IF POSSIBLE.
REPORTS TO MRCC ROME, INMARSAT-C: 424744220,
PHONE: 3906 5908 4527, 3906 5908 4409,
FAX: 390 6592 2737, 3906 5908 4793,
E-MAIL: ITMRCC@MIT.GOV.IT.
REPORTS TO LIBYAN NAVY COAST GUARD,
PHONE: 2182 1444 9488, 2189 2517 9382,
E-MAIL: LIBYAN.NAVAL.COMMS.CENTRE@GMAIL.COM.
( 241012Z JUN 2018 )
```

Fig. 4. A sample Broadcast Warning.

"capsized", "taking on water"), and pairs of latitude/longitude coordinates (though these are often approximate or self-reported).

Our analysis shows that during peak periods of migration in the Central Mediterranean, Broadcast Warnings contain extensive information on migrant boats in distress. Between 2014 and 2018, we found an average of 20 relevant warnings per month. All together, we were able to generate a dataset of 1,094 geotagged warnings, of which 614 had associated information on the number of people involved. These warnings revealed an evolution in the spatial distribution of rescues over time, with calls for help growing increasingly close to shore in 2016 and 2017 (see Fig. 3).

C. Twitter

Finally, Twitter contains a surprising amount of real-time information on the Central Mediterranean migration crisis. In the early stages of the crisis, when easily-accessible and standardized datasets were limited, this was one of the most comprehensive sources of information on rescues. For example, the Italian Coast Guard's tweets often included detailed information on the number of migrants involved in a rescue, the name of the rescue ship, and coordination with other rescue operations.[4]

However, compiling structured rescue data from Twitter is difficult due to variations in tweet structure, according to the organization that produced the tweet; the need to translate foreign-language tweets; and the lack of georeferenced information. Nevertheless, Twitter has been useful for monitoring the situation at large.

III. "Custom-made" Reports and Official Statistics

A. Official Sources

1) Guardia Costiera: Since 2016, the Italian Coast Guard has produced aggregate monthly reports on rescue operations, which is a more consistent source of information than their Twitter feed. Available from the "Attività SAR Immigrazione" section of their research page [34], these PDF reports include data on the number of operations coordinated by the MRCC and the number of people rescued. Critically, they also include information on the role of different actors in the rescues, highlighting the changing contribution of Italian, European, and NGO forces. The Italian Coast Guard reports show the growing role of NGOs in the rescue efforts over time, concurrent with a decline in operations by the Italian Navy (see Fig. 5). They also show how the MRCC has coordinated the majority of rescues in the SAR zone off the Libyan Coast, making it the most authoritative source of information on the situation. It is possible to download and process these reports to obtain a longitudinal view of operations.

2) The UN Refugee Agency and the International Organization for Migration: Two United Nations (UN) organizations—UNHCR and IOM—have taken the lead in monitoring the crisis. UNHCR offers a data portal which highlights arrivals in Europe by sea with a focus on three destination countries: Spain, Italy, and Greece [1]. Specifically, the available data include: (1) Daily arrivals by sea in Italy and Greece since mid-2015; (2) Monthly arrivals by sea for all three countries since 2014; and (3) Overall arrivals for all three countries broken down by country of origin for 2016–2017. Internally, arrivals are also recorded by port at the level of individual rescue ships, in coordination with the Italian authorities.

IOM offers two additional data portals. First, IOM's ongoing Missing Migrants project tracks migrant deaths and disappearances around the world since 2014, in an impressive effort to manually compile information from government agencies, NGOs, and news sources. A full dataset is available from IOM's online portal, which also links to a "regions in focus" page on the Mediterranean [7]. These data are geotagged and associated with a regional category, as well as estimates of the number of dead, missing, and survivors, and a breakdown of the number of men, women, and children involved. They also contain qualitative descriptions of the cause of death and an estimated rating of the quality of the information source, with a link to the original report. Second, IOM's Displacement Tracking Matrix (DTM) project attempts to quantify origin-destination flows of refugees and internally displaced people in high-priority settings around the world [35]. The quality and specificity of DTM data varies by context; for the Mediterranean, data is currently available in the form of quarterly reports from 2016 onward [36]. These reports include: (1) Monthly arrivals by country of origin for Italy, Greece, and (more recently) Spain, broken down by gender and whether an individual is an accompanied or unaccompanied minor; and (2) Daily arrivals by land and sea, broken down by country for Spain, Greece, Italy, Malta, Cyprus, and a number of other landlocked transit countries.

B. Unofficial Sources

1) WatchTheMed: WatchTheMed (WTM) is an organization that advocates on behalf of migrants and refugees in the Mediterranean. Inspired by individuals such as Eritrean priest Don Mussie Zerai, who had been personally receiving phone calls from refugees and migrants in need of help since

[4] For example: "@guardiacostiera, 14 Oct 2016: #CentraleOperativa #GuardiaCostiera coordina 2 operazioni #SAR: unità @jugendrettet salva 138 #migranti su 1 gommone e su 1 piccola barca."

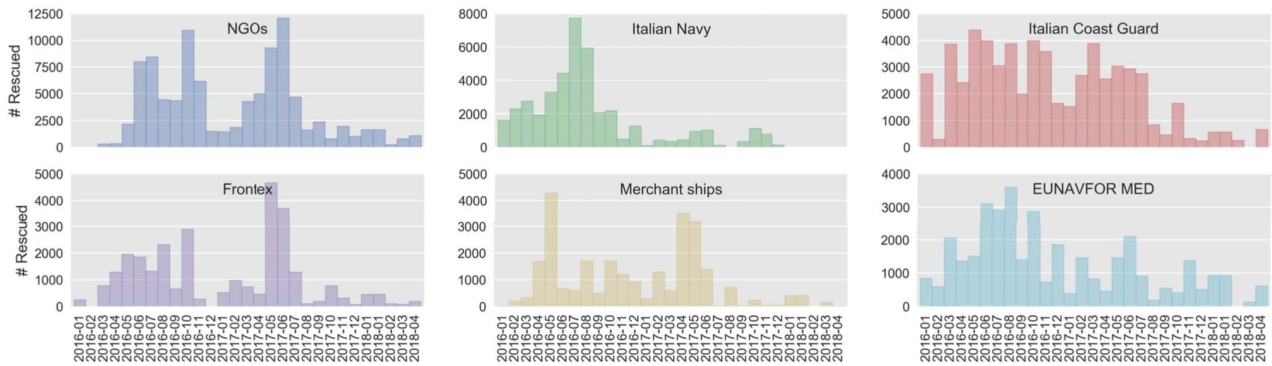


Fig. 5. The number of people rescued by actor and month, for rescues coordinated by the MRCC. Note that the y-axis scale varies by actor.

2004, WTM created a volunteer emergency hotline called AlarmPhone in 2014 [37], [38]. Upon receiving a call for help from a migrant or refugee vessel, AlarmPhone attempts to liaise between the vessel and rescue organizations, track the vessel's status, and organize a rescue. Once the vessel is brought to shore, the organization documents the incident in a case report.

WTM has produced 829 incident reports since 2011, including 269 related to the Western Mediterranean route, 143 related to the Central Mediterranean route, and 410 related to the Eastern Mediterranean route (see Fig. 11) [39]. Although there is variation in the level of detail and the number of incidents described within each report, all provide helpful context. Often, they explicitly describe the involvement of other organizations, such as the MRCC or NGOs. However, to make the most of the information provided, it is often necessary to extract structured data (e.g., timestamps, the number of people involved, and the precise location of the incident), which can be tedious and time-consuming.

2) Rescue NGOs: Some rescue organizations also share detailed data. MSF provides an interactive map showing operations by ship and date, along with weather conditions, the length of the operation, and the number of people involved [40]. Xchange, a migration research group associated with MOAS, has produced reports featuring data from MOAS rescue ship trajectories, and even a sample GPS trajectory from a migrant and refugee boat [41], [42].

3) Journalists and Advocates: Finally, journalists have also played a critical role in documenting the crisis. For example, in 2013 a group of European journalists created the Migrants' Files. This project represented an early effort to compile systematic data on migration deaths and costs for routes into Europe [43]. The Migrants' Files' "Counting the dead" project compiled data on dead and missing migrants from 2000 to mid 2016, using an "open-source intelligence" approach which coded data from available news sources [43]. This effort was a predecessor to IOM's Missing Migrants project; it provided geolocated incident-level information on: the number of dead and missing, the route, the cause of death, a description of the incident, and the source. Similarly, the German newspaper der Tagesspiegel has compiled a list of *"33,293 Registered Asylum-Seekers, Refugees, and Migrants Brought to Death as a Result of Fortress Europe's Restrictive Policies"* from 1993 to 2017 [44]. Forensic Architecture has also produced detailed reports on the situation in the Central Mediterranean, which include narratives on the "left-to-die boat" (a boat that remained adrift off the Libyan coast for two weeks in 2011 due to the negligence of responsible authorities); the risks of allowing commercial ships to conduct migrant rescues; and the growing political hostility toward rescue organizations operating in the region [19], [45], [46].

IV. The Quantified Rescue

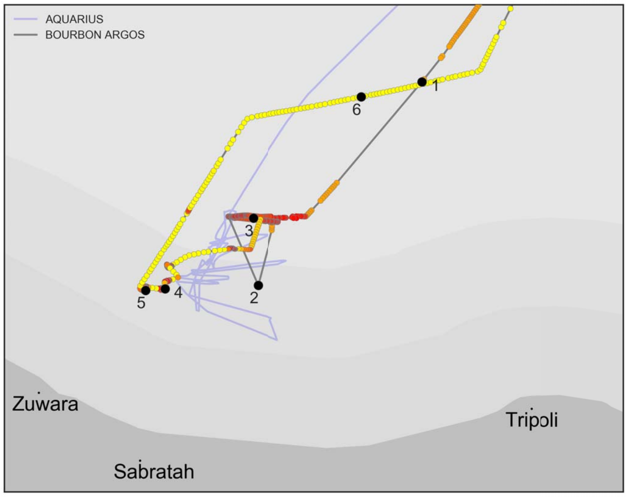
A. Working Definition

We define the *quantified rescue* as a framework for fusing multiple datasets around an individual maritime SAR operation. This allows us to build coherent narratives involving geospatial data, descriptions of rescue activities, key events, and even photos or videos of the situation at sea. One advantage of quantified rescues is that they describe operations without being restricted to the efforts or perspective of a single organization.

Fig. 6 shows a quantified rescue. It displays the AIS tracks of two MSF boats—the Bourbon Argos and the Aquarius—from October 1–6, 2016, along with timestamped tweets from the @msf_sea Twitter account. According to MSF, on October 3 [47]:

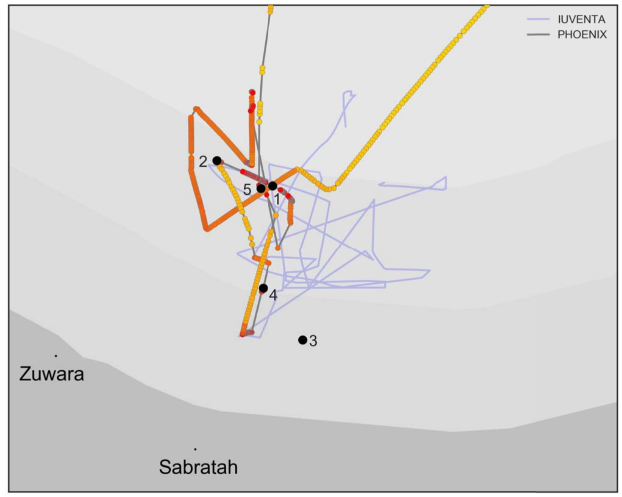
> . . . The 60 metre vessel [the Bourbon Argos] was . . . engaged in eight separate rescues, taking on board a total of 1,019 people. At the same time the Aquarius was involved in the painstaking rescue of a staggering 720 people from one very overcrowded wooden boat.

Fig. 7 shows another quantified rescue, coordinated by MOAS with the help of NGO Jugend Rettet. The figure displays the trajectories of boats from both organizations from October 10–15, along with key moments of the SAR operation, a nearby Broadcast Warning, and a tweet from the @moas_eu account. According to an October 13 press release [48]:



1 10-01 16:18 “@msf_sea: update: the bourbon argos is on its way to the sar zone and should be in position . . . ”
2 10-02 09:25 “@msf_sea: breaking: the #argos has just rescued 26 #people from this tiny speck of a wooden boat . . . ”
3 10-02 18:23 “@msf_sea: update: the #dignity1 and bourbon argos have transferred those rescued today to @savethechildren and @moas_eu . . . ”
4 10-03 11:15 “@msf_sea: then, just as the sun came up the #argos crew rescued this lilting and over crowded boat. followed by 4 more . . . ”
5 10-03 13:27 “@msf_sea: update: the rescues just keep coming on the bourbon #argos. with 7 boats rescued so far the deck is filling fast . . . ”
6 10-03 18:17 “@msf_sea: with 1019 men, women and kids on board there is not a centimetre spare on the #argos . . . ”

Fig. 6. An example of a quantified rescue involving a single ship. Each point represents an AIS position reading, with point colour illustrating speed in nautical miles per hour (red represents slower speeds). The top track shows the Bourbon Argos, while the underlying (light blue) track shows the Aquarius.



1 2016-10-12 14:00 Migrant/refugee boat departs Sabratha
2 2016-10-12 19:00 Phoenix receives call from MRCC
3 2016-10-12 20:24 “LIBYA. DNC 09. VESSEL, 125 PERSONS ON BOARD, REQUESTING ASSISTANCE IN 32-57N 012-29E AT 121601Z OCT. VESSELS IN VICINITY REQUESTED TO KEEP A SHARP LOOKOUT, ASSIST IF POSSIBLE. REPORTS TO MRCC ROME, INMARSAT-C: 424744220, PHONE: 3906 5908 4527 OR 3906 5908 4409 . . . ”
4 2016-10-12 21:20 Migrant/refugee boat is located
5 2016-10-13 07:06 “@moas_eu: #phoenix rescue of 113 people last night who claim to have been 130 on board. unfortunately no bodies were recovered #safeandlegalroutes . . . ”

Fig. 7. An example of a quantified rescue involving multiple ships. Each point represents an AIS position reading, with point colour illustrating speed in nautical miles per hour (red represents slower speeds). The top track shows the MOAS Phoenix, while the underlying (light blue) track shows Jugend Rettet’s Iuventa.

> At least 17 people are considered to be dead/missing according to the 113 survivors rescued from a single rubber boat . . . The Phoenix crew was alerted of the vessel in distress following a call received at 19:00 from the Maritime Rescue Coordination Centre (MRCC) in Rome last night. Still, it was only at 21:20 that the rubber boat was sighted . . . Thanks to the great effort from the crews of Proactiva Open Arms, Jugend Rettet, the Boat Refugee Foundation and the joint MOAS-CRI mission, the rescue operation was swiftly carried out and all the survivors quickly transferred on board the Phoenix. Following the rescue operation, a search to recover bodies was launched, however this had to be interrupted due to persistent rough weather at sea . . . According to survivors, they departed from Sabratha in Libya at around 14:00 yesterday afternoon. At the time of rescue, many people had already been in the water for several hours.

These examples illustrate the potential of combining geo-tagged data with substantive incident reports, yielding insights into how many rescues ships are conducting, where and how they happen, and how long they take.

B. Considerations and Challenges in Fusing Datasets that Describe Rescues

Ideally, a process for fusing AIS trajectory data with descriptive information from tweets, Broadcast Warnings, WTM reports, and NGO rescue data would involve a local search around individual ship trajectories for nearby incidents. However, this can be difficult given that the geo-coordinates associated with Broadcast Warnings and rescue data can be approximate or rounded, making incidents fall outside of recorded ship trajectories (see e.g., the Broadcast Warning in Fig. 7).

In practice, we have found that timestamps are most helpful for fusing the data: descriptive statements or tweets can be associated with the AIS coordinates that are closest in time. This approach requires that descriptions be filtered by ship before associating them with ship trajectories. However, this is not always possible, especially if the data are vague or incomplete.

The frequency of AIS reporting is also relevant when creating quantified rescues. Since rescues involve sudden movements and frequent changes in positions, a rescue with AIS data reported at 15-minute intervals can look quite different from a rescue with AIS data reported at 2-minute intervals (see Fig. 1 for an example of a rescue with low-frequency reporting, and Figs. 6 and 7 for examples of high-frequency reporting).

Another challenge is that the amount of available information varies substantially by incident. As illustrated above, NGOs such as MSF and MOAS are vocal about their activities and may even share press releases describing their rescues.

However, detailed reports are more likely to result from extreme situations, such as when multiple rescues occur in a short period of time. Other rescue efforts, such as those led by joint EU operations and naval ships, may be less prone to publicity. In addition, observations of rescues as a whole are meaningfully shaped by selection bias: there is no question that many ships disappear at sea without a trace, and are only discovered later as the evidence washes up on the Libyan coast. Rescue organizations have enlisted drones and spotter planes in an effort to prevent this, but there is an open question as to whether satellite data may also be helpful in retrospectively finding unreported vessels.

The last challenge regards information quality. On the one hand, many rescue descriptions lack data–particularly geospatial information regarding where an incident occurred, or practical information on the number of people involved. Therefore, it can also be hard to remove duplicate data. For example, a ship in distress may generate multiple Broadcast Warnings–once when it requests assistance, and again when it begins to sink–but the information in the Broadcast Warning message may not be sufficient to link the records to the same incident. Similarly, news reports of the number of survivors and losses can be inconsistent, because it is hard to accurately count the number of people on board a sinking ship. Incidents are often reported using survivors' estimates, which can vary across different accounts.

V. Automated Identification of Rescues

A. Problem Definition and Exploratory Work

As a result of the challenges involved in manually creating quantified rescues, we have begun to explore the potential for automatically characterizing rescue activity. While AIS trajectory classification has been used in a number of contexts, including the automated recognition of fishing behavior [26], to our knowledge there has not been any attempt to automatically discover rescue operation patterns.

We have developed a model of rescue behavior based on the characteristics of trajectory points, including: speed, course over ground, latitude, longitude, day of week, hour of day, and month of year. For this preliminary work, we manually geotagged 77,372 points spanning four search and rescue ships over a period of 100 days. Specifically, we classified points according to two estimated activity types—"Rescue" and "Non-rescue"—based on the shape of the trajectory and the corresponding rescue boat's speed.

By tuning standard binary classification algorithms—AdaBoost, Support Vector Machines (SVM), and logistic regression—using 10-fold cross-validation, we have been able to correctly classify over 96% of points, using a dataset for which approximately 75% of the points are labeled as non-rescues. Although this initial outlook is optimistic, we face a key limitation: classifier performance is location-dependent. Specifically, the classifiers appear to rely on a ship's location near the Libyan coast to identify "rescue" behavior, so that upon removing latitude and longitude coordinates from the set of predictors, accuracy declines all the way down to 69% for logistic regression.

Fig. 8. An example of "rescue signatures" manually tagged in AIS data [49].

To address this problem, we have considered the possibility of clustering points prior to classification. The advantage of clustering is that the characteristics of nearby points can inform individual points' labels. For example, a reading of a fast speed surrounded by slow speeds is likely to be part of a rescue, whereas a reading of a fast speed surrounded by other high speeds is likely to be non-rescue behavior. Other work with trajectory data has used clustering to classify points into "stop" and "move" behavior, and then inferred characteristics about the discovered "stops" and "moves". In our context, we have found the Cluster-Based Stops and Moves of Trajectories (CB-SMoT) algorithm described in [50] to be most suitable. CB-SMoT is a density-based algorithm for trajectory stop-point detection. Density is defined using two parameters: *eps*, a distance parameter, and *mintime*, a time parameter. CB-SMoT uses these parameters to classify points into three categories: (1) *Core points*, from which the object travels less than the distance threshold *eps* in either direction within a period of length *mintime*; (2) *Border points*, falling within *eps* of a core point; and (3) *Other points*, which are neither border nor core. Together, core and border points form clusters that represent periods of slow motion. Although the algorithm is designed to find stop points, we have used it to break trajectories into segments with different motion profiles by allowing sequences of "other" points to form their own clusters. CB-SMoT has proven to be particularly useful for detecting SAR operations because of two key features: its natural incorporation of *both* time *and* distance traveled, and its ability to work with sequences of unevenly-spaced AIS readings.

Initial tests of the algorithm show good performance for a *mintime* of 1 hour and an *eps* of 500 meters (see Fig. 10 for a sample output). With these parameters, the median cluster contains 9 points, and the mean cluster size is approximately 60 points. The median cluster lasts for 33 minutes, and the mean cluster duration is just over 6 hours. However, there

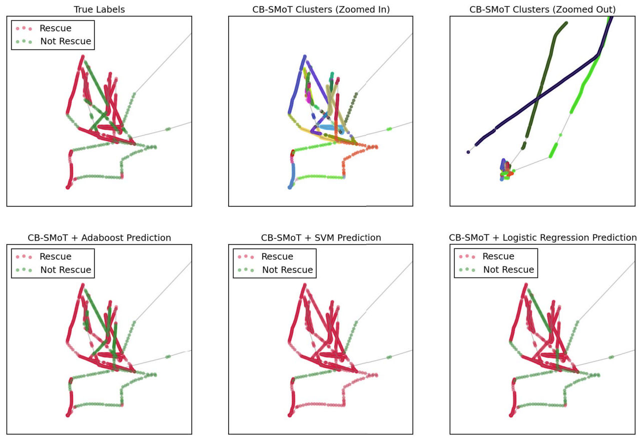


Fig. 10. An example of how CB-SMoT can be used to segment and classify trajectories. The top left cell shows the "true" labels obtained by manually tagging the data. The colored segments on the top right show the clusters formed by CB-SMoT. The second row shows the labels assigned by different classification algorithms on the clustered data.

are three main issues with this approach. First, clustering points prior to classification substantially reduces the size of the dataset: rather than working with 77,000+ points, we are left with 1,192 clusters. Second, even after including geographic features like mean cluster latitude and longitude, performance is generally worse than for the point-wise data (see Fig. 9). Third, this approach requires tuning the first-stage clustering algorithm parameters *concurrently with* the second-stage classification algorithm parameters, which complicates the cross-validation search.

B. Future Work

To overcome these issues, our ongoing work departs in two directions from the approach described above. First, we plan to convert our trajectory data into images and test the performance of out-of-the-box deep learning image classification algorithms on these rescue "signatures", in an analogy to the classic MNIST handwriting recognition problem (see Fig. 8) [51]. This will eliminate the need for a two-stage clustering-and-classification approach, since deep learning algorithms essentially implement and tune a multilevel learning strategy automatically. Other architectures to be tested include Convolutional Neural Networks (CNNs), which are useful in recognizing patterns with varying location and scale; and Recurrent Neural Networks (RNNs), which could better leverage the sequential structure of our data.

Second, we are in the process of improving our approach to trajectory labeling. When tagging trajectories by visual inspection, it is difficult to determine when a rescue starts and ends. On the one hand, a ship may spot a migrant vessel and move toward its position, but on the other hand, ships can be called on to assist migrant vessels that are hours away [52]. Similarly, ships seem to move along the coast while patrolling for migrant vessels, creating the appearance that the ship is engaging in an abnormal rescue-like maneuver. Rescues and transfers can also be easily confused, given that they both occur at slow speeds. As a result, it can be hard to determine a ship's intent from trajectory characteristics alone.

To address these problems and augment the labeled dataset, we have been using the ground-truth rescue data reported by MSF [40]. Exploratory work with these data suggests that MSF rescue timestamps correspond to the period of time when people pass from a rescued ship to a rescue ship. Therefore, we will consider extending the MSF rescue labels forward and backward by a fixed distance or time, in order to capture the full change in ship behavior as a result of rescue maneuvers.

VI. Policy and Operational Implications

We see a number of potential policy and operational applications for quantified rescues and the automatic classification of rescue patterns at scale.

First, they can be used to study the geographic evolution of rescue operations. For example, smuggling activities may change locations over time, either to avoid law enforcement, or in tandem with the rise and fall of local power brokers. Fig. 3 shows that between 2016 and 2017, rescues moved closer to the Libyan shore, as the quality of the vessels deployed decreased, and the rescue effort became more proactive.

Second, they can be used to retrace potentially contentious rescue maneuvers. Fig. 2 shows a clear pattern of trips from ports in Italy and Malta to the SAR zone off the Libyan Coast. It also shows that rescue operations generally respect Libyan sovereignty by avoiding waters within 12 nautical miles from the coast. This has been an important matter of conflict between the Libyan authorities and NGOs [53].

Third, they can inform the theory and practice of conducting rescue operations, helping to optimize maneuvers according to e.g., conditions at sea. They can be used to synthesize the practical knowledge of NGO SAR crews, and to provide insights for the design of future life-saving operations.

Fourth, they can be used to facilitate coordination between rescue ships, and to ensure proper spatiotemporal coverage of rescue operations. For example, it appears there are occasionally gaps in rescue ship availability, when no ship is available to assist a vessel in distress [54]. They can also help detect potential departures from coastal regions that are not actively patrolled by NGO rescue boats (e.g., from Tunisia, Eastern Libya, or Egypt), and highlight areas that require more thorough monitoring.

Finally, they can be used to identify rescues that might otherwise go unreported, such as those conducted by commercial ships. Under maritime law, any capable ship at sea is required to provide support to a nearby vessel in distress. These kinds of rescues were particularly salient in early 2015, prior to the large-scale involvement of NGO-operated ships. In April 2015, two rescue efforts involving migrant or refugee vessels and ill-equipped commercial ships resulted in an estimated 1,000+ deaths [19]. The role of commercial ships is believed to have declined over time, and there are anecdotal reports of ships shifting their routes to avoid encountering migrant boats. However, the Coast Guard data presented in Fig. 5 suggests

	Accuracy AdaBoost	 Logistic Regression	 SVM	Precision Adaboost	 Logistic Regression	 SVM
Point-wise dataset	0.966	0.976	0.960	0.937	0.968	0.907
Point-wise dataset - no location	0.936	0.689	0.778	0.884	0.345	0.572
Clustered dataset	0.858	0.881	0.716	0.890	0.959	1.000

Fig. 9. Preliminary results from applying classification algorithms to manually-tagged AIS data. Accuracy and precision are shown for a dataset of trajectory points; a dataset of trajectory points with latitude and longitude features removed; and a dataset of clusters obtained with the CB-SMoT algorithm. Ten-fold cross-validation was used to select: the number of estimators for Adaboost, the regularization penalty for logistic regression, and the degree of the polynomial kernel and the misclassification penalty for the Support Vector Machine (SVM) classifier. The data show that accuracy drops when location information is removed from the point-wise dataset, and that point-wise classification currently performs better than classification at the cluster level.

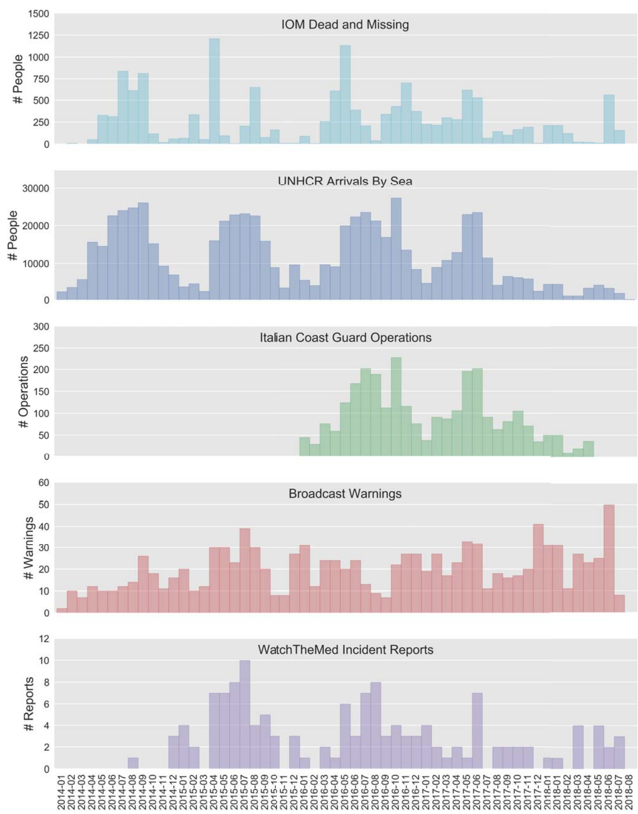


Fig. 11. Monthly counts of deaths, arrivals, rescue operations, Broadcast Warnings, and incident reports in the Central Mediterranean, compiled from various sources. Note that the y-axis scale varies by actor.

that merchant ships still play a meaningful part in the rescue effort. Better documentation of these efforts can be used to quantify their cost to the merchant shipping industry, and can support claims to compensate ships for their involvement.

VII. Conclusion

In this article, we have provided detailed and comprehensive documentation of available data related to migrant and refugee flows in the Central Mediterranean. We intend for this to serve as a resource for others studying the region. We have highlighted "ready-made" data sources that offer real-time, always-on monitoring capability, and "custom-made" data sources that are designed specifically to track an evolving situation. We have discussed the practical aspects of fusing these data, as well as their individual strengths and weaknesses. We have illustrated two potential applications: (1) the creation of *quantified rescues* that provide a holistic perspective on rescue operations; and (2) the use of machine learning to automate the detection of rescue maneuvers in ship trajectories.

Our main contribution is twofold. First, we have introduced AIS data and Broadcast Warnings for the large-scale quantitative analysis of SAR operations. Second, we have provided a case study of the use of data fusion to describe and analyze a rapidly evolving humanitarian crisis, offering a blueprint for future events. This fusion can offer faithful information on a humanitarian emergency that requires immediate action. Such transparency can help advocates and rescue organizations intervene to protect human rights and individual dignity. It can also ensure that citizens and policymakers are correctly informed on emerging crises. While the flow of migrants from Libya to Italy has declined in recent months, it seems unlikely that it will abate altogether, and the number of dead and missing is still high (see Fig. 11). In addition, the emergence of new maritime crises—such as the departure of boats from Venezuela and Myanmar [55], [56]—indicates that migration by sea will continue to be a pressing humanitarian issue.

That said, transparency can also pose a risk. Adversarial actors can use information on ship movements to intervene in or obstruct rescue operations, as the Libyan Coast Guard has already done. When combining these data, it is necessary to conduct a thorough risks and harms assessment and to implement privacy protection measures, as the migrants and refugees captured in e.g., photos or tweets are particularly vulnerable people. Consequently, careful attention should be paid to anonymizing and appropriately aggregating observations, including the removal of personally identifiable information and other measures to preserve individual and group privacy.

Ultimately, the value of this type of data fusion depends on its use by government officials, rescue organizations, and humanitarian agencies. We hope that the work presented here leaves them better equipped to act on the available information, and highlights the potential of data science and machine learning in a real-world policy setting.

ACKNOWLEDGMENTS

We thank Amanda Zerbe, Gijs Brouwer, Felicia Vacarelu, James-Léon Dufour, Duncan Breen, Rebeca Moreno Jiménez and Christopher Earney who advised and supported this research. We thank UNHCR, IOM, MarineTraffic, Guardia Costiera, and MSF for helpful conversations on the project and/or data. Global Pulse thanks the Government of the Netherlands for its support.

REFERENCES

[1] UNHCR, "Mediterranean situation," Jul. 2018. [Online]. Available: https://data2.unhcr.org/en/situations/mediterranean

[2] ——, "Central Mediterranean route situation: Supplementary appeal," 2018. [Online]. Available: http://www.unhcr.org/5aa78775c.pdf

[3] M. Specia, "Italy and Malta block another rescue ship carrying migrants," *The New York Times*, Jun. 2018.

[4] Associated Press and AFP, "Libyan coastguard accused of abandoning three migrants in sea," *The Guardian*, Jul. 2018.

[5] S. Osborne, "Libya's coastguard says it has intercepted more than 570 refugees bound for Europe," Aug. 2018. [Online]. Available: https://goo.gl/4xLTGu

[6] UNHCR, "UNHCR sounds alarm as Mediterranean Sea deaths pass 1,500 mark," Aug. 2018. [Online]. Available: https://goo.gl/XVvbHs

[7] IOM, "Missing Migrants Project: Mediterranean," Aug. 2018. [Online]. Available: https://missingmigrants.iom.int/region/mediterranean

[8] M. J. Salganik, *Bit by Bit: Social Research in the Digital Age*. Princeton University Press, 2017.

[9] IOM, "Fatal Journeys Volume 3 Part 1: Improving data on missing migrants," Tech. Rep., 2017. [Online]. Available: https://goo.gl/FfZWyq

[10] IMO, "AIS transponders," 2018. [Online]. Available: http://www.imo.org/en/ourwork/safety/navigation/pages/ais.aspx

[11] Raymarine, "AIS - The utimate guide to Automatic Identification Systems," Sep. 2016. [Online]. Available: https://goo.gl/A15wSe

[12] MarineTraffic, "Global ship tracking intelligence," 2018. [Online]. Available: https://www.marinetraffic.com/

[13] exactEarth, "AIS vessel tracking | Maritime ship monitoring," 2018. [Online]. Available: http://www.exactearth.com/

[14] ORBCOMM, "IoT and M2M tracking, monitoring and control solutions - Satellite, cellular and dual-mode communications," 2018. [Online]. Available: https://www.orbcomm.com/

[15] FleetMon, "Professional live vessel tracking for the maritime industry," 2018. [Online]. Available: https://www.fleetmon.com/

[16] VesselFinder, "Free AIS ship tracking of marine traffic," 2018. [Online]. Available: https://www.vesselfinder.com

[17] VesselTracker, "Terrestrial & satellite AIS tracking service in realtime," 2018. [Online]. Available: https://www.vesseltracker.com/en/Home.html

[18] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting AIS data for intelligent maritime navigation - A comprehensive survey," *arXiv:1606.00981 [cs]*, Jun. 2016. [Online]. Available: http://arxiv.org/abs/1606.00981

[19] C. Heller and L. Pezzani, "Death by rescue." [Online]. Available: https://deathbyrescue.org

[20] "FAQs." [Online]. Available: https://goo.gl/rojGjr

[21] Wikipedia, "Automatic Identification System," Jul. 2018. [Online]. Available: https://goo.gl/TFXPtQ

[22] K. Schwehr, "libais: C++ decoder for Automatic Identification System for tracking ships and decoding maritime information," Jul. 2018. [Online]. Available: https://github.com/schwehr/libais

[23] M. Robards, G. Silber, J. Adams, J. Arroyo, D. Lorenzini, K. Schwehr, and J. Amos, "Conservation science and policy applications of the marine vessel Automatic Identification System (AIS): A review," *Bulletin of Marine Science*, vol. 92, no. 1, pp. 75–103, Jan. 2016.

[24] Windward, "AIS data on the high seas: An analysis of the magnitude and implications of growing data manipulation at sea," Windward, Tech. Rep., Feb. 2015. [Online]. Available: https://goo.gl/XBhGLo

[25] M. Robinson, "Ghost ships turning off GPS when entering European waters," *Daily Mail*, Mar. 2017. [Online]. Available: https://goo.gl/YxvcSe

[26] F. Natale, M. Gibin, A. Alessandrini, M. Vespe, and A. Paulrud, "Mapping fishing effort through AIS data," *PLOS ONE*, vol. 10, no. 6, p. e0130746, Jun. 2015.

[27] B. J. Rhodes, N. A. Bomberger, M. Seibert, and A. M. Waxman, "Maritime situation monitoring and awareness using learning mechanisms," in *MILCOM 2005 - 2005 IEEE Military Communications Conference*, Oct. 2005, pp. 646–652 Vol. 1.

[28] M. Vespe, I. Visentini, K. Bryan, and P. Braca, "Unsupervised learning of maritime traffic patterns for anomaly detection," in *9th IET Data Fusion Target Tracking Conference: Algorithms Applications*, May 2012, pp. 1–5.

[29] Pulse Lab Jakarta, "The fourth research dive on trade and competitiveness," UN Global Pulse, Technical Report, Sep. 2017. [Online]. Available: https://goo.gl/obYMA2

[30] C. Langford, T. Cheng, and M. Vespe, "Data driven identification of migrant vessels at sea," in *Maritime Knowledge Discovery and Anomaly Detection Workshop Proceedings*, Ispra, Jul. 2016, pp. 69–73.

[31] National Geospatial-Intelligence Agency, "Broadcast warnings," 2018. [Online]. Available: https://goo.gl/PsYUQJ

[32] ——, "Radio navigational warning systems," 2018. [Online]. Available: https://msi.nga.mil/MSISiteContent/StaticFiles/Images/navwarnings.jpg

[33] ——, "Geographic locator," 2018. [Online]. Available: https://msi.nga.mil/MSISiteContent/StaticFiles/Images/subr_1.jpe

[34] Guardia Costiera, "Attivita SAR immigrazione," 2018. [Online]. Available: http://www.guardiacostiera.gov.it/attivita/Pages/Ricerca.aspx

[35] IOM, "DTM - Understanding displacement," 2018. [Online]. Available: https://displacement.iom.int/

[36] ——, "Migration flows - Europe," Jul. 2018. [Online]. Available: http://migration.iom.int/europe/

[37] M. Schwartz, "The Anchor," *The New Yorker*, Apr. 2014.

[38] PRO ASYL, "PRO ASYLs human rights award bestowed upon Father Mussie Zerai," Sep. 2016. [Online]. Available: https://goo.gl/zQ845t

[39] Watch the Med, "Reports," 2018. [Online]. Available: http://watchthemed.net/index.php/reports

[40] MSF, "Search and rescue - Interactive map," 2018. [Online]. Available: http://searchandrescue.msf.org/map.html

[41] Xchange, "Central Mediterranean survey: Mapping migration routes & incidents," Jun. 2018. [Online]. Available: https://goo.gl/sYULoq

[42] P. G. Cadabon, "Migration geo-portal," Aug. 2017. [Online]. Available: http://xchange.org/map/GeoPortal.html

[43] The Migrants' Files, "The human and financial cost of 15 years of fortress Europe," Jun. 2016. [Online]. Available: http://www.themigrantsfiles.com/

[44] United for Intercultural Action, "Liste von 33.293 registrierten Asylsuchenden, Gefluechteten und Migrant*innen, die aufgrund der restriktiven Politik der Festung Europas zu Tode kamen," *Der Tagesspiegel*, Jun. 2017. [Online]. Available: https://goo.gl/fsgUoB

[45] C. Heller and L. Pezzani, "The Left-to-Die Boat." [Online]. Available: https://www.forensic-architecture.org/case/left-die-boat/

[46] ——, "Blaming the rescuers," 2017. [Online]. Available: https://blamingtherescuers.org

[47] MSF, "Mediterranean: MSF rescued nearly 2,000 people in less than seven hours." [Online]. Available: https://goo.gl/VcT31u

[48] MOAS, "MOAS-Italian Red Cross late night rescue: 113 people rescued; at least 17 missing, among which a 3-year-old baby," Oct. 2016. [Online]. Available: https://goo.gl/5UVefk

[49] UN Global Pulse, "Rescue signatures in the Mediterranean," Sep. 2017. [Online]. Available: http://rescuesignatures.unglobalpulse.net/mediterranean/

[50] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, "A clustering-based approach for discovering interesting places in trajectories," in *Proceedings of the 2008 ACM Symposium on Applied Computing*. New York, NY, USA: ACM, 2008, pp. 863–868.

[51] Y. LeCun, C. Cortes, and C. J. C. Burges, "The MNIST database of handwritten digits," 1998. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[52] Watch the Med, "05/08: 98 travellers from Libya rescued to Italy," Aug. 2016. [Online]. Available: http://watchthemed.net/reports/view/552

[53] S. Cullen, M. Veselinovic, and L. Smith-Spark, "NGOs halt migrant rescue operations, citing Libya 'threat'," Aug. 2017. [Online]. Available: https://goo.gl/KFXkV7

[54] M. Sheena and L. Borghese, "Hundreds of migrants die crossing the Mediterranean," Jul. 2018. [Online]. Available: https://goo.gl/cfSKJz

[55] N. Casey, "Hungry Venezuelans flee in boats to escape economic collapse," *The New York Times*, Nov. 2016.

[56] J. Pearlman, "Who are the Rohingya boat people?" *The Telegraph*, May 2015.