Science Advances
AAAS

Supplementary Materials for

Tracking elusive and shifting identities of the global fishing fleet

Jaeyoon Park *et al.*

Corresponding author: Jaeyoon Park, jaeyoon@globalfishingwatch.org

Sci. Adv. **9**, eabp8200 (2023)
DOI: 10.1126/sciadv.abp8200

This PDF file includes:

Supplementary Text
Figs. S1 to S16
Tables S1 to S4

Supplementary Text

1. Existing registry sources

Global record of fishing vessels, refrigerated transport vessels and supply vessels

The public version of the Global Record of Fishing Vessels, Refrigerated Transport Vessels and Supply Vessels, a vessel information system launched in 2018 and run by the Food and Agriculture Organization of the United Nations (FAO) (denoted as the Global Record here; https://globalrecord.fao.org/), contains vessel identities with regard to fishing, refrigerated transport, and supply vessels provided by the official State authorities. While the Global Record is an official, global initiative with recent progress on increased participation from its Member States, it has not yet realized its full potential:

- Data availability: A total of 66 States have submitted their vessel data (as of July 2022), with limited vessel identity information.
- Data quality: Member States provide their own data. Many do not have significant fisheries resources to clean, cross-check or maintain official records, meaning the quality and consistency of their records can vary.
- Update cycle: Update dates of each State are not homogeneous, ranging from 2016 to 2022. It is unclear how often States update their data, and when they will update their old data. Again, this is often due to limited resources on the part of the State.
- Historical record: As of July 2022, 3 out of 66 states (<5%) provide historical records, which is an important building block for tracking activities of vessels and changing identities, including reflagging.
- Regional fisheries management organization (RFMO): Member States provide some national license information (~15%, 10 out of 66, as of July 2022), but this is often not complete and does not necessarily include RFMO authorisation information (e.g. no details on authorization range).
- Activity information: Tracking information, such as when the vessel was last active and where they have operated is not provided (e.g. the temporal geo-locational information that the Automatic Identification System (AIS) or national vessel monitoring systems provide).
- International scope: Some vessels operating in international waters may not be assigned an International Maritime Organization (IMO) Number and thus, because the Global Record is limited to vessels with IMO numbers, may not be included in the Global Record as per its current scope.

As of July 2022, the Global Record is within Phase 1 of its programme of work. The programme's Phase 1 means that all vessels of 100 gross tonnage, or 24 meters and above, are being included in the Information System. Subsequently, a Feasibility Study will be carried out to assess the most effective ways to move into Phases 2 and 3, which will respectively mean inputting data for vessels between 50 and 100 gross tonnage or between 18 and 24 meters, and data for vessels between 10 and 50 gross tonnage or between 12 and 18 meters. It will offer a comprehensive database of vessel identity once fishing nations submit their data with complete information, including vessel historical identity information, on a regular basis and in a consistent way.

Our data will continue to complement information included in the Global Record by providing an alternative mechanism for aggregated data shared by flag States to be cross-checked with vessel level data reported via AIS broadcasts. Our data can also support vessel identification needs whilst the Global Record undergoes continued development and until Phases 2 and 3 are complete.

The FAO describes the Global Record as a programme that *aims towards providing a single access point for information on vessels used for fishing and fishing-related activities with the primary objective being to combat illegal, unreported and unregulated (IUU) fishing by enhancing transparency and traceability.* It is our intention to fully support this aim. Table S3 presents a brief comparison between the Global Record and the vessel identity data set used in this paper.

Registry from regional fisheries management organizations

Currently, 13 RFMOs provide vessel registries. These RFMOs provide registries with varying degrees of identity information, often with incomplete data. There is no standard vessel registry format, therefore no consistency is guaranteed across multiple RFMOs. Previously, the consolidated list of authorized vessels, known as CLAV, incorporated data from five Tuna-RFMOs in a more consistent manner. However, the initiative did not continue due to lack of momentum and financial support. Table S1 lists the RFMOs that publicly provide vessel registry data.

Registry from national authorities

The list of registries obtained from national authorities includes 17 country-level vessel registries (the European Union's fleet registry and the Pacific Islands Forum Fisheries Agency's vessel registration are each considered one registry) is available in table S2.

Other sources of vessel information

In addition to the vessel registries provided by RFMOs or national authorities, vessel identity information can be found from other sources. The international seafood sustainability foundation offers lists of vessels owned by companies that commit to fisheries transparency by publishing information about their vessels among others. TM-Tracking, a non-governmental organization that generates fisheries intelligence, maintains a combined illegal, unreported, and unregulated (IUU) fishing vessel list consolidating IUU vessel lists from main RFMOs. A few vessel lists with various types of information are also obtained from scientific research articles, publications from non-governmental organizations, and fisheries experts who manually checked vessel information either derived from vessel tracking data or through desktop research. Another public source of information is the Global Integrated Shipping Information System that offers free access to the identity information of IMO-numbered vessels made public by the International Maritime Organization. Additionally, vessel information is available through private vessel databases such as Sea-Web maintained by IHS Markit Ltd., a private information provider. For this study, we cross-checked a variety of data sources to ensure the best quality of the information in our data used in this study, however, the vessel information in our data is subject to error and omission (see Supplementary Text section S3).

2. <u>Data preparation</u>

Collecting registry data

From over 40 data sources presented in Supplementary Text section S1, we have collected information on a monthly basis since early 2019 to capture record changes in the registry. Monthly updates are particularly helpful to track vessel record changes over time when registries do not publish nor maintain historical records (e.g., IATTC, SIOFA; see table S1 for full names of RFMOs). The acquisition of information was performed by a combination of manual and automatic techniques. Most of the publicly available registries are made ready for download as files in a tabular format whereas others keep their vessel lists displayed only on their web pages. For the latter case, we applied an automated script to acquire text-based information. We used the acquired information uniquely for the non-commercial purpose of aggregating it with other data sets. We do not publicly disclose the raw information that we obtain, and we include source attributes in the data for this study.

Registry-AIS matching

We matched the collected vessel information to tracking data of AIS. For this purpose, we used an approximate string match, also known as fuzzy logic, in multiple identity fields between these two data sources (registry and AIS). The identity fields to be compared include ship name, international radio call sign, IMO number, AIS identifier number (Maritime Mobile Service Identity, MMSI), and flag. The matching method uses a multi-step process and each step compares various combinations of the identity fields to determine that records from both sources are close enough to be considered a match. Each step uses a scoring system with preset thresholds to determine matches. This includes cases where:

i) Both ship name field and call sign field match

ii) At least two fields out of three fields among ship name, call sign, and flag match when the IMO field matches

iii) At least two fields out of four among ship name, call sign, flag, and IMO number match when the MMSI field matches

a) Match based on ship name and callsign

First, records from registry and AIS that share the same ship name are compared and called a match if the call signs from both records are close enough (defined by a threshold "max_distance" using *Levenshtein edit distance*). This value varies from 0 to the maximum length between call signs that are compared. 0 means a perfect match and the larger the value is, the greater the difference is. The score is calculated by the average of the weighted sum of the number of messages from the AIS record. If the score is equal to 1, call signs from both registry and AIS records whose ship names are identical perfectly match. If the score is equal to 0, call signs from both records are completely different. Only records that pass the threshold are accepted:

$$(1 - max\_distance \,/\, length\_of\_callsign) < \frac{\sum_{k=1}^{n}((1 - ratio\_of\_levenshtein\_distance\_k) \times messages\_k)}{\sum_{k=1}^{n}(messages\_k)} \leq 1$$

Records from registry and AIS that share the same call sign are then compared and considered a match if the ship names from registry and AIS are close enough (defined by a threshold "max_ratio" using *Levenshtein edit distance*). This value varies from 0 to 1. 0 represents a perfect match and 1 represents a complete difference. The score is calculated by the average of the weighted sum of the number of AIS messages. If the score is equal to 1, ship names from both registry and AIS records whose call signs are identical perfectly match. If the score is equal to 0, ship names from both records are completely different. Only records that pass the threshold are accepted:

$$(1 - max\_ratio) < \frac{\sum_{k=1}^{n}((1 - ratio\_of\_levenshtein\_distance\_k) \times messages\_k)}{\sum_{k=1}^{n}(messages\_k)} \leq 1$$

Based on both scores calculated above, we mark record pairs as a match when both scores are above thresholds (meaning that ship names and call signs from registry and AIS are close enough to be paired).

b) Match based on IMO number

The second round of matching is carried out by pairing both registry and AIS data sets on the same IMO numbers. If at least two identity fields among three—ship name, call sign and flag—match without conflicting fields, while IMO numbers are identical, the AIS and registry are considered to match. This match could have already been made using the process described in a) or could represent a new record match. If one of the comparing fields is missing, it contributes 0 points, whereas if a field does not pass the respective thresholds, then it contributes -1 point (e.g., if call signs and flags match—gaining 2 points—while ship names are unmatched—losing 1 point—the final match is represented by 1 point) to avoid false-matching. The score is calculated by normalizing the points the pair gains and averaging on the weighted sums of the number of messages, and only those that score above the set threshold are accepted:

$$2/3 \leq \frac{\sum_{k=1}^{n}((imo\_points)/3 \times messages\_k)}{\sum_{k=1}^{n}(messages\_k)} \leq 1$$

c) Match based on MMSI number

The last matching round is carried out by pairing both registry and AIS data sets on the same MMSI numbers. If at least two identity fields among four—ship name, call sign, IMO number, and flag—match without conflicting fields, while the MMSI numbers are identical, then we call it a match. This match could be a duplicate of what has already matched in the processes above, but could also be an addition. If one of the comparing fields is missing, it yields 0 points to be added, whereas if a field does not pass the respective thresholds, then it loses 1 point (e.g., if ship names, call signs and flags match—gaining 3 points—while the IMO numbers are different—losing 1 point—then it finally gets 2 points). However, if ship names do not match, regardless of whether all other fields match, it is considered a non-match because we consider identities that share all the same identity fields except ship names as different vessel identities. The score is calculated by normalizing the points the pair gains and averaging on the weighted sums of AIS messages, and only those that score above the set threshold are accepted:

$$2/4 \leq \frac{\sum_{k=1}^{n}((mmsi\_points)/4 \times messages\_k)}{\sum_{k=1}^{n}(messages\_k)} \leq 1$$

Finally, we aggregate all matches from the processes a) to c) above to identify matched registry-AIS pairs. In addition to this standard matching, we applied an additional rule for vessels from flag States that publicly share their entire vessel list (e.g., Australia, Peru). We are confident that we can determine a match if only ship names and flags match without additional identity fields provided. We exclude, however, the matches that include ship names that are shared by multiple vessels in the registries. For instance, there are more than 200 vessels that share the same name of "FREEDOM" among which we cannot determine which one broadcasts AIS messages with the same name.

Sensitivity analysis

We carried out a sensitivity analysis for each category of matching above to determine the threshold over which a match is accepted. Out of roughly 200,000 AIS-registry pairs that are compared in the latest four months of data collection, about 4,000 (2%) were manually annotated as binary ground truth (true vs. false). We then set a range of thresholds over which a match is accepted for each category of matching (ship name, call sign, IMO number, and MMSI number), and calculated F1-score to determine the thresholds that yield the highest F1 score on each category. Fig. S1 illustrates that the similarity in ship name pairs (calculated based on *Levenshtein edit distance*) returns the highest F1 score at 0.75, and that in call sign pairs at 0.7 while the score is best at 0.67 and 0.5 for matching based on IMO number (calculated based on the number of other identity fields to be close enough) and MMSI number respectively. As a call sign is mostly composed of 4 to 6 characters, we used a threshold of one character difference for call sign pairs.

Data aggregation

Each matching pair of records between AIS and registry is combined to select the most representative information or generate aggregated data derived from multiple sources of information. The aggregation process depends on the nature of fields, including a) numeric fields (length overall, gross tonnage, and overall engine power), b) categorical fields (vessel classification), c) time range fields (authorization period, AIS active period), and d) text fields (owner name, owner flag State). Through this process, data is cross-checked across multiple registries and outliers are removed to increase the accuracy. Each type of aggregation method is described below.

a) Numeric field

Numeric values from sources for the same vessel identity are averaged excluding the ones that are two standard deviations away from the mean value. Each registry source presents only one value to avoid repetitive values due to regular scrapes. The units used in this data set are meter for length overall, gross ton for tonnage, and kilowatt for engine power.

b) Categorical field

Vessel types are classified into 22 categories based on fishing gear type (for fishing vessels) and fish support functions (for support vessels). Fig. S2 depicts the hierarchy of vessel classifications. We used this classification because it allows us to have records from the registry aligned with the vessel classification inferred by Global Fishing Watch's convolutional neural network model. Based on the classification hierarchy, vessel classes from the source registries are either combined with pipes (which denotes "or", like "trawlers|seiner" meaning that the vessel is either a trawler or a seiner) or rolled down toward leaf nodes (for instance, a vessel is classified to be "purse_seines" and "tuna_purse_seines" by two registry sources, then its final class is considered as "tuna_purse_seines").

c) Time range field

Authorization data and AIS activity data are represented as time ranges. The information about authorization for fishing is provided either by RFMOs or national fishing authorities. As authorization information from various registries for a vessel is recorded at different moments and possibly as different time range segments, we generate minimum, continuous time range blocks that include all the segments. To minimize possible reporting gaps or errors from a registry, any gaps of less than three months between two consecutive authorization periods for the same vessel are accepted as potentially authorized. The same applies to the AIS activity. All segments of AIS activity are grouped into continuous time range blocks and two consecutive time range blocks are combined when the gap between them is less than three months.

d) Text field

Text fields including owner names and owner flags are combined by only distinct values. To avoid redundant information, owner names are standardized using a Python library called *shipdataprocess* which drops most suffixes with regard to types of companies (e.g., Co. Ltd. in English, Kaisha in Japanese, S.A. in French, OOO in Russian). Owner flags are standardized to country code based on ISO 3166-1 alpha-3.

Data cleaning

We conducted multiple rounds of data cleaning to correct records or fill in the missing information in the draft vessel identity data set. This process is done through 1) desktop reviews of vessels drawing upon vessel registries, news articles, and vessel images found on the web, 2) inspection of vessel tracks using Global Fishing Watch's AIS data and public map to determine the correct vessel classification, and 3) personal communication with authorities and fisheries experts with regard to detailed vessel information.

3. Data quality

Comparison with machine learning inference

To analyze the quality and completeness of registry data sources, and of our aggregation process, we compare the registry values with AIS-derived inference values. Kroodsma *et al*. (13) applied a convolutional neural network to infer vessel dimensions (length overall, gross tonnage, and engine power) and vessel classification from movement patterns. Such a model can assist in identifying characteristics of vessels for which there is incomplete, or no information and in highlighting possibly incorrect data in registries.

First, to analyze the completeness of our data, we compare the availability of four categories of information (length overall, gross tonnage, engine power, and vessel classification) from four sources: raw registries (red lines in fig. S3), the aggregated data across registries (orange lines), the model inference (yellow lines), and the combination of the aggregated registries with model inference (blue lines) by RFMO. Fig. S3 illustrates the gain in information starting from the raw registry data to the combination of registry aggregation and model data. The figure shows that only IATTC starts off with nearly complete vessel characteristic information. The other RFMOs are shown to be incomplete, to a varying degree, in one or more categories of information, and most notably, NAFO provides no vessel characteristic information. By comparison, our data collection, its aggregation, and the fusion with the model inference data allows us to obtain a largely complete estimation of the characteristics of vessels registered to RFMOs.

Second, to assess the data quality, we compare how well the reported data about vessel classification from RFMOs and our aggregated data match up with the model inferred values. Some factors may limit the accuracy of the model's prediction (e.g., areas with low AIS coverage, and gear types with less distinctive or inconsistent movement patterns). However, the model has shown overall good performance (13), therefore model inference values can be used as a proxy for ground truth. Furthermore, comparing model predictions with registries can validate the correctness of registry information when these match, or highlight areas for investigation when there is a difference.

For the analysis, matching scores were assigned to records from both RFMOs and our data to indicate the matching quality with the model inference. Exact vessel classification matches with the most granularity (fig. S2) were favored by assigning these the highest score. To reflect when a reported vessel class partially matches but is less specific than the model prediction, these matches were assigned a relatively low score. This lack of specificity may occur 1) when the registered vessel class is in the same hierarchy as the inferred value, but less granular (e.g., purse_seines vs. tuna_purse_seines in fig. S2), or 2) when multiple different classes are provided by different registries for a given vessel. For each RFMO, the matching scores associated with each registered vessel are summed and normalized to indicate the overall quality of vessel class information available for vessels registered to RFMOs (as a proportion of the maximum possible score).

Fig. S4 illustrates scores for pre-aggregated vessel classes averaged over registries and post-aggregation scores for two aggregation approaches. The plot indicates that, while some RFMOs may provide reliable and specific vessel class information, the overall information provided by RFMO registries is potentially ambiguous and inconsistent. The plot also validates that by applying an informed aggregation process, which extracts the higher quality information available from registries, overall more specific and consistent vessel classes are made available for each vessel. For example, the overall pre-aggregation score for vessels registered to NAFO is relatively low, indicating that the vessel classes provided by all the registries to which those vessels are listed (NAFO and other registries) are either unspecific or contradictory. Our aggregation process, indicated by the relatively high post-aggregation scores, provides vessel classes that are more specific and consistent.

4. Number of vessels

Count of vessel identities

We counted the number of vessel identities in three categories: 1) matched identities between records from registries and AIS, 2) AIS records that are matched to no registry records, and 3) registry records that are matched to no AIS records. To compare the total number of vessels across the data set and by flag State, we set the threshold of vessel length to be greater than 24 meters. This length is selected because vessels over 24 meters are more likely to be equipped with AIS devices and registered to public registries (14), which in turn leads to a better comparison. Where vessel length is unavailable, we used gross tonnage over 100 as a proxy threshold which is based on the length-tonnage relationship drawn from our data set. To avoid double counting due to different ways of registering a vessel identity, ship names and call signs are normalized using a public Python library called *shipdataprocess* (developed by the authors for this study) which removes unnecessary characters for comparison, harmonizes vessel name prefix/suffix, and standardizes different types of alphabets into simple roman alphabets (e.g., MV Sea-horse No. II to SEAHORSE2). Flag State names are converted into three-digit country codes defined in ISO 3166-1. IMO numbers are also cleaned drawing on its check character scheme developed by IHS Markit. For AIS records, we removed records with MMSI numbers whose positional messages are too scarce (<50 messages).

Number of vessels by year

To count the number of vessel identities by year, we used the following criteria in addition to the thresholds set above: 1) for matched identities and unmatched AIS records, AIS activity range must include the year of interest and the vessel must be active (measured by the number of AIS messages broadcast for the given year), and 2) for unmatched registry records, the vessel must be authorized to fish for the given year in at least one registry or its registry record must have been extracted in the given year. For instance, fig. S5a shows the number of vessel identities flagged to China and the rest of the world by year. The bars in red indicate unmatched AIS records whereas the bars in blue represent matched records between AIS and registry. The bars in yellow indicate unmatched registry records. The total number of vessel identities over 24 meters in AIS increased about 2.7 times from 2012 to 2020, which is due partly to the increased adoption of AIS devices. Not only has China encouraged vessel operators to use AIS in the past years, but the number of Indonesian fishing vessels broadcasting AIS was also multiplied by 20 times in 2020-2021 compared to the previous years due to policy changes (28 to 570). The total number of matched identities by year also increased two-fold from about 6,200 to 12,500 in 2012-2020.

5. Authorization data

Our data include authorization information explicitly from eight RFMO vessel lists (see table S1 for these RFMOs) and three country registries (Norway, Iceland, and the Faroe Islands).

We formulate authorization time ranges from five additional RFMO vessel lists (GFCM, IATTC, NAFO, NEAFC, and SEAFO) and one country registry (Peru) by assuming that the vessels on the registry at the time of data extraction are currently authorized to operate. We extract these data every month starting from sometime between 2019 and 2020 depending on registries. We then group these monthly data to create proxy authorization time ranges. To avoid potential temporary registry omission or data retrieval errors, a gap less than three months between two consecutive time ranges is considered to be part of an authorized period.

The authorization data can help monitor the compliance of fishing by matching apparent fishing effort in AIS to authorization records published by international or national authorities. Spatiotemporal analyses of fishing compliance provide information about hotspots of potentially unauthorized or unregulated fishing both on the high seas and within EEZs (fig. S6). For the compliance of tuna catch on the high seas, the top 10 flag States that represent the most fishing effort in each of the tuna RFMOs are shown in the following figure (fig. S7). The portion of fishing hours by vessels with unknown authorization in the IOTC Convention Area is due not only to the absence of matched Sri Lankan vessels and some unmatching vessels from the top fishing States, but also to a high level of fishing by vessels with invalid MMSI numbers. The latter suggests that the misuse of AIS, which is discouraged by the International Convention for the Safety of Life at Sea (SOLAS), is more prevalent among the vessels operating in the IOTC Convention Area than other tuna RFMO areas.

6. Reflagging and vessel history

A transfer of flag, or reflagging, takes place between two consecutive vessel identities when a vessel changes its flag, and therefore its identity. We used Vessel Record IDs to identify vessel identities that are associated with the same hulls and establish the temporal order of these identities by drawing on their AIS activity ranges.

Combined with GPS data, the flag information of vessels helps us better understand the flagging history of fleets. To reconstruct the history of the top 15 reflagging States, we extracted AIS data and used a dot to represent AIS activity for a given week regardless of the intensity of activity. The dots are color-coded by a vessel's flag in a given week and are placed in a chronological manner such that one line of dots represents a vessel's flag history from 2012 to 2021 (Fig. 3A and 3B). Fig. S10 illustrates all support vessels from those top 15 reflagging States at the time of January 1, 2022, with the fraction of the fleet that had reflagged at least once in our data in parentheses. For instance, half of the Panama-flagged support vessels as of January 1, 2022 were reflagged from flags other than Panama at least once between 2012 and 2021, and these foreign States are color-coded (the largest one being Liberia). The other half represent Panama-flagged support vessels that have not changed their flags since 2012. The U.S. fleet of support vessels, which exhibits no reflagging practice, shows a recurring pattern of increased AIS activity in the middle of a given year. This coincides with the peak season of salmon being transferred from its cages by U.S. fish tender vessels.

The ability to reconstruct vessel flagging history allows us to track the number of reflagged vessels by flag State in a given year, and consequently, the trend of States that grant reflagging vessels access to their flag. Fig. S11 illustrates yearly trends in the number of vessels with reflagging history in 2012-2021 for the top 10 flag States that have the most reflagging vessels. While Russia and Panama have been the largest recipients of vessels with foreign-flag history in fishing and support fleets, respectively, an increased number of vessels flew flags of a few emerging States recently. These are Namibia, China, Georgia, and Iceland for fishing vessels, and Rep. of Korea and the Bahamas for support vessels.

Flags of convenience

Flags of convenience reported by the International Transport Workers' Federation (ITF) were obtained on October 13, 2021 from https://www.itfseafarers.org/en/focs/current-registries-listed-as-focs. This list is created by the ITF's Fair Practices Committee and is subject to change. The list of flag states reported as flags of convenience on this date are captured in table S4.

7. Caveats of data and analyses

Limit on public identity data

We recognize that the quantity and quality of our collected registry data vary by flag State, thereby introducing inherent bias in information about vessel identity. For instance, we have less complete information about Asian fleets as a smaller number of these flag States make their vessel registries publicly available compared to European nations. While the lack of information is partly complemented by the other types of data sources (self-reported AIS data, machine learning model inference, and vessel monitoring system data published by national authorities), far broader registry data are inevitably required to address the disparity in identity information. Fortunately, a growing number of flag States make their vessel identity data publicly available through global initiatives such as FAO's Global Record of Fishing Vessels, Refrigerated Transport Vessels and Supply Vessels. In addition, we continue to search and incorporate more data sources that are scattered and fragmented in the public domain to provide the public with easy access to information.

Additionally, our data has a bias toward larger vessels, in particular those over 24 meters, as they are more likely registered to national or international public registries than smaller vessels. Apart from a few national-level registries that include vessels as small as 15 meters (e.g., European fleet register and U.S. vessel registry), most registries collected provide limited identity data for smaller vessels. Additionally, these smaller vessels are less likely to be equipped with AIS devices because relevant regulation or government support is lacking, limiting our ability to produce AIS-registry matched records. An increasing number of initiatives, including Taiwan Fisheries Agency's financial support in AIS installation (https://www.fa.gov.tw/view.php?theme=Press_release&subtheme=&id=123), encourage broader use of AIS or similar tracking systems (China; http://www.moa.gov.cn/nybgb/2012/dwq/201805/t20180514_6142007.htm) and assist nations in publicly sharing their proprietary vessel monitoring system data to bring greater transparency to large and small-scale fleets (Norway; https://www.fiskeridir.no/Yrkesfiske/Rapportering-paa-havet/AApenhet-om-fangst-og-aktivitetsdata-ERS-og-posisjonsrapportering).

Limit on AIS for vessel tracking

Tracking vessel activity through AIS data is limited by several factors which have been documented in the literature (13, 14). These limitations hinder our ability to tackle IUU fishing as the vessels that engage in IUU fishing are more likely to keep their transponders off. In some cases, these “dark vessels" have perpetrated large-scale illegal fishing (43). Emerging technologies and new approaches to address such gaps have been introduced to tackle IUU fishing. While some of them provide useful information about fleet-level characteristics, they are often unable to offer vessel-level characteristics such as vessel identity, owner, and authorization. These vessels are also unlikely to reveal their identity information in public registries, making it even more difficult to investigate the cases. Recently, flag States have adopted more proactive policies that help tackle IUU fishing and punish illegal operators. Additionally, raw AIS data need to be purchased from companies, and it is not a fully open dataset although some NGOs including Global Fishing Watch publish processed AIS data for download or analysis on web platforms.

Limit on universal standard

Our vessel classification is limited as it does not consider all possible types of vessels around the world. Also, some vessel classes are interpreted differently across states and regions. To ensure the broadest use of our data, however, we are working toward aligning our vessel types, fishing gears, and nomenclature more closely with those adopted by the UN Food and Agriculture Organization.

Unique vessel identifiers are also still lacking. IMO numbers are assigned to a limited number of vessels due to lack of relevant policies encouraging a broader group of fishing vessels to adopt an IMO number. In this study, we developed a Vessel Record ID scheme to fill in such a gap. It is, however, dependent on the availability of public registries, and requires more effort to maintain. The number of vessels having adopted IMO numbers continues to grow in recent years, and more flag States and international bodies encourage vessels to apply for an IMO number for greater transparency. This trend will facilitate more robust tracking of the identity history of vessels.

8. Data representativeness

While the absolute number of fishing vessels >12 meters in our data looks small (around 20% of the world’s fishing vessels >12 meters), we argue that our study reasonably represents a global analysis for the following reasons:

Our data represent a significant portion of the activity of large vessels. In our manuscript, we stated that our data represent about 40% of all vessels >24 meters broadcasting AIS, but these vessels are highly active, accounting for 74% of all fishing by vessels >24 meters broadcasting AIS and 90% of all fishing by vessels >24 meters broadcasting AIS excluding China. Compared to the total global count of vessels regardless of AIS use, our data represents about one-third of fishing vessels greater than 24 meters (15,000 out of 45,000 as of 2020 according to FAO) (1). If vessels without AIS are similar to those with AIS, it suggests that our data would still represent 61% of fishing activity by the global fleet of vessels >24 meters and 75% of fishing by all vessels >24 meters excluding China. Additionally, given that we identified more than half of the Chinese distant water fleet in our data (see fig. S5), our data cover most of the world's industrial fishing (by vessels >24 meters) with the exception of Chinese fishing for which our data still covers the majority of the Chinese distant water fleet. In other words, most of the fishing activity (by vessels >24 meters) that our data miss is from Chinese domestic fishing. For additional reference, AIS can track the majority of the world's fishing vessels above 24 meters (14) and likely covers about 80 percent of vessels fishing in the high seas (16). These findings suggest our data has significant global coverage, because it identifies most of industrial fishing vessels and their fishing activities captured on AIS and there is the potential for our identity data to continue to increase in coverage if more registries become public and matched to AIS.

For vessels of 12-24 meters, which account for a large proportion of the world's fishing vessels, this vessel category likely skews towards vessels that are domestically active only. Only <1% (~150) of the vessels of 12-24 meters in our data (~20,000 vessels) change their flag whereas ~20% of the vessels >24 meters in our data changed flags. Although our data do not cover much of fishing by vessels of this category, our analysis suggests that these vessels are unlikely to engage in frequent identity changes, flag hopping, port visits by foreign flagged vessels, and high seas fishing with unknown authorization, all activities with high IUU risk. Therefore the bias in our global analysis due to the gap in vessels of 12-24 meters would be limited.

Additionally, our data cover most of IMO-Numbered fishing and support vessels in the world. About 18,000 IMO-Numbered vessels exist in our data, making up about 67% of all IMO-Numbered fishing and support vessels found in the database of IHS Markit (which manages the IMO-Number scheme for the International Maritime Organization). The coverage goes up to 80% if one accounts only for vessels greater than 24 meters. This high coverage of IMO-Numbered fishing vessels is important as they are the vessels likely to engage on international voyages (45), therefore highly relevant to our analysis of reflagging, foreign port visits, and high seas fishing. In a simple comparison, the Global Record covers about 45% of all IMO-Numbered fishing vessels as of July 2022.

Regarding the ownership data, we have about 20,000 vessel identities with the owner information (only six out of 17 RFMO websites have online vessel registries with a listed vessel owner, mostly owning company's name and nationality). This coverage represents approximately 70% of the total vessel identities and 65% of fishing hours on AIS.

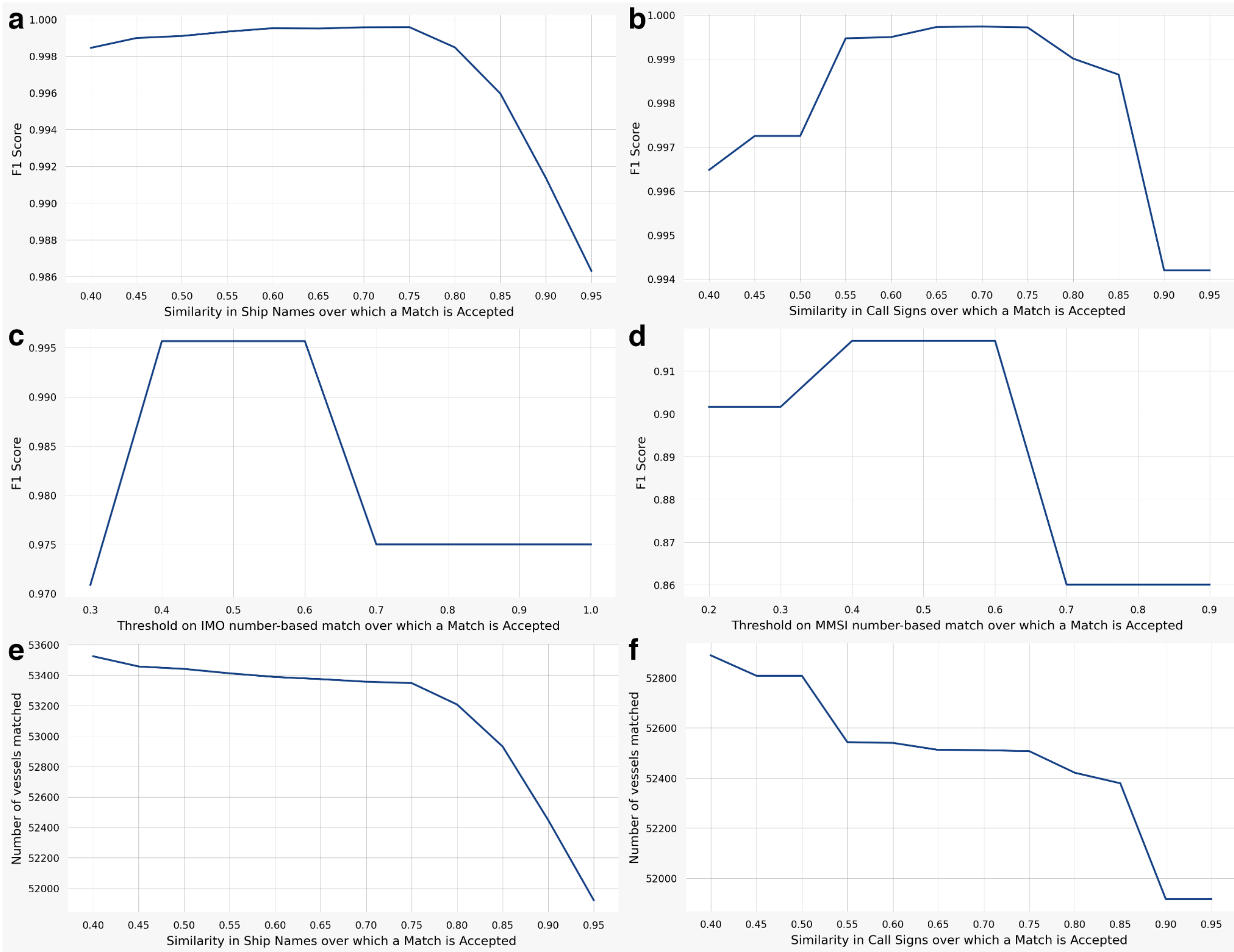

Fig. S1. Optimization and sensitivity analysis of AIS-registry matching.

(a-d) The graphs show the optimum thresholds over which a match on ship name, international radio call sign, IMO number, or MMSI number is accepted. Out of about 3,000 AIS-registry pairs that are manually annotated as binary ground truth, thresholds of 0.75 and 0.7 yield the best F1-score for ship name and call sign pairs respectively. For IMO number- and MMSI number-based matching, thresholds that return the highest F1 scores correspond to 2 out of 3 fields matched and 2 out of 4 fields matched respectively. **(e, f)** The sensitivity analyses show how much the number of matched records change as the thresholds vary. For the threshold of ship name, the matched number declines more rapidly when the threshold value is over 0.75 than when it is under, but the ratio of change relative to the total matched number is less than 3%. For the threshold of call sign, the ratio of change when the threshold is around 0.7 relative to the total matched number is less than 2%.

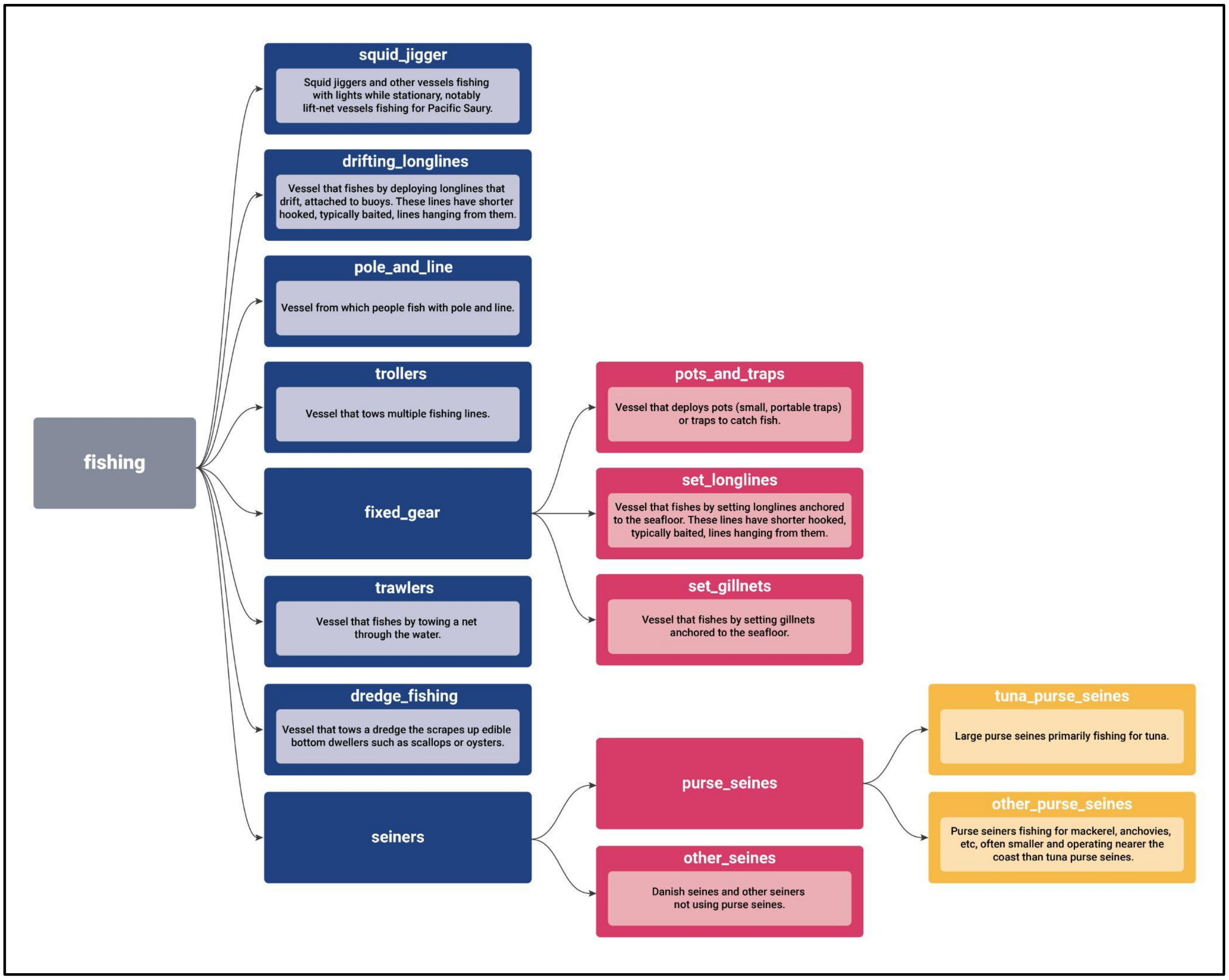

Fig. S2. Hierarchy of vessel classification used in this study.

Some vessel classes are nested under broader classes (e.g., tuna_purse_seines is nested under purse_seines). The "fishing" class is the top node of the hierarchy representing a fishing vessel whose specific vessel class is unidentified.

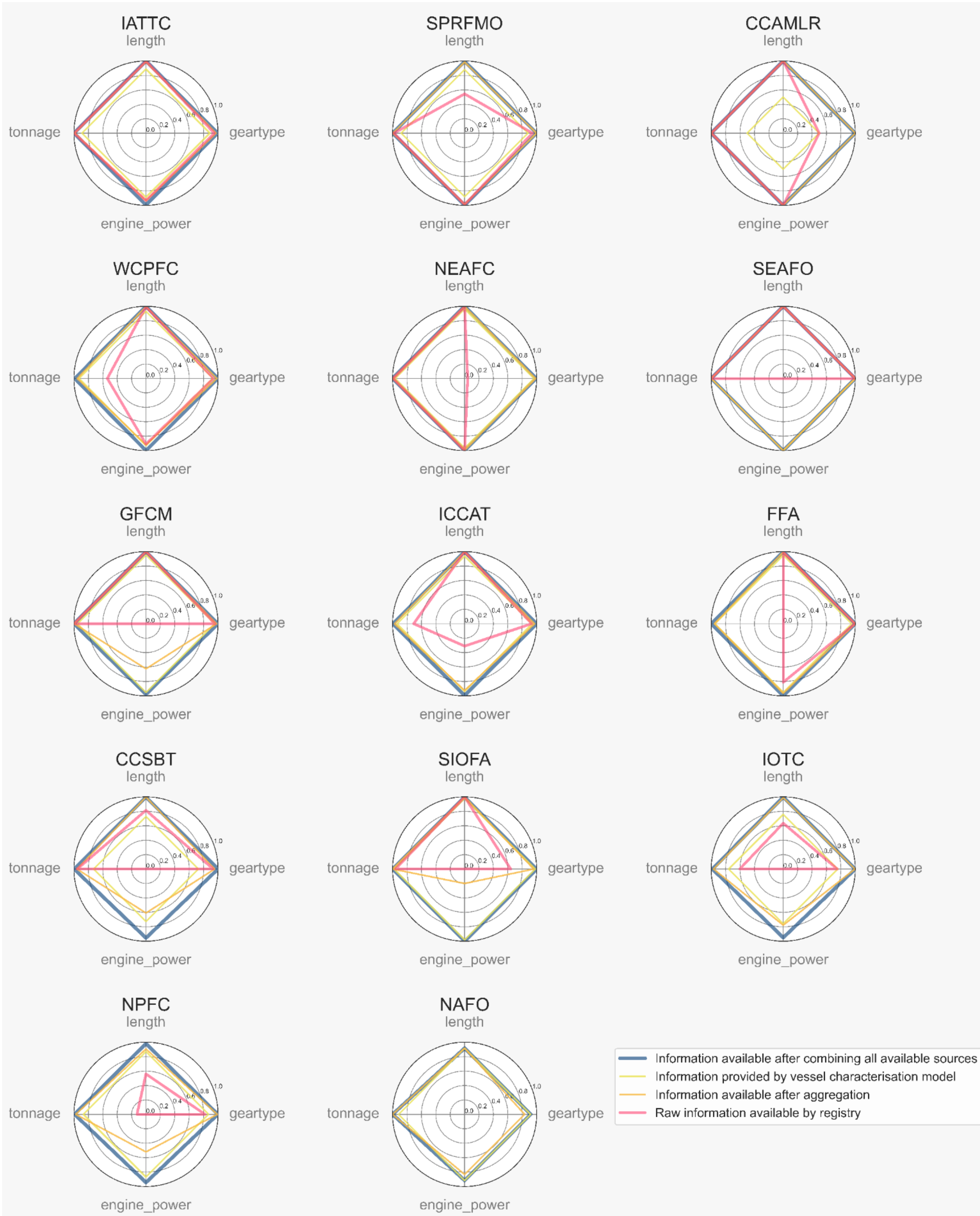

Fig. S3. Information available by registry, before and after aggregating, from the vessel characterization model, and all combined.

The completeness of our data for vessel length, gross tonnage, engine power, and vessel classification. The sources are raw registries (red lines), aggregated data across registries (orange lines), vessel characterization model (yellow lines), and the combination of registries with the model (blue lines).

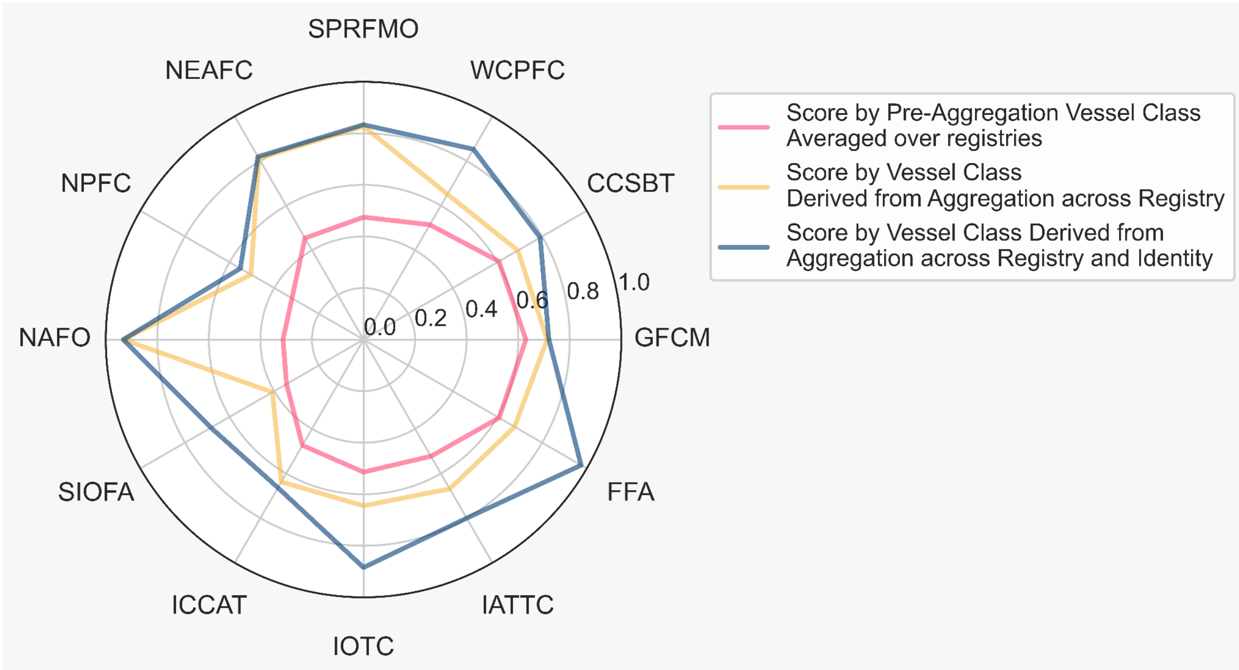

Fig. S4. Comparison of vessel classification reported by RFMO, and by our aggregated data.
This figure compares vessel classification reported by RFMO and the aggregated data used in this study. The scores (0 being disagreement and 1 being the complete agreement) indicate the quality of matching with the vessel characterization model. The highest scores are assigned for specific and consistent vessel classifications that match with the model.

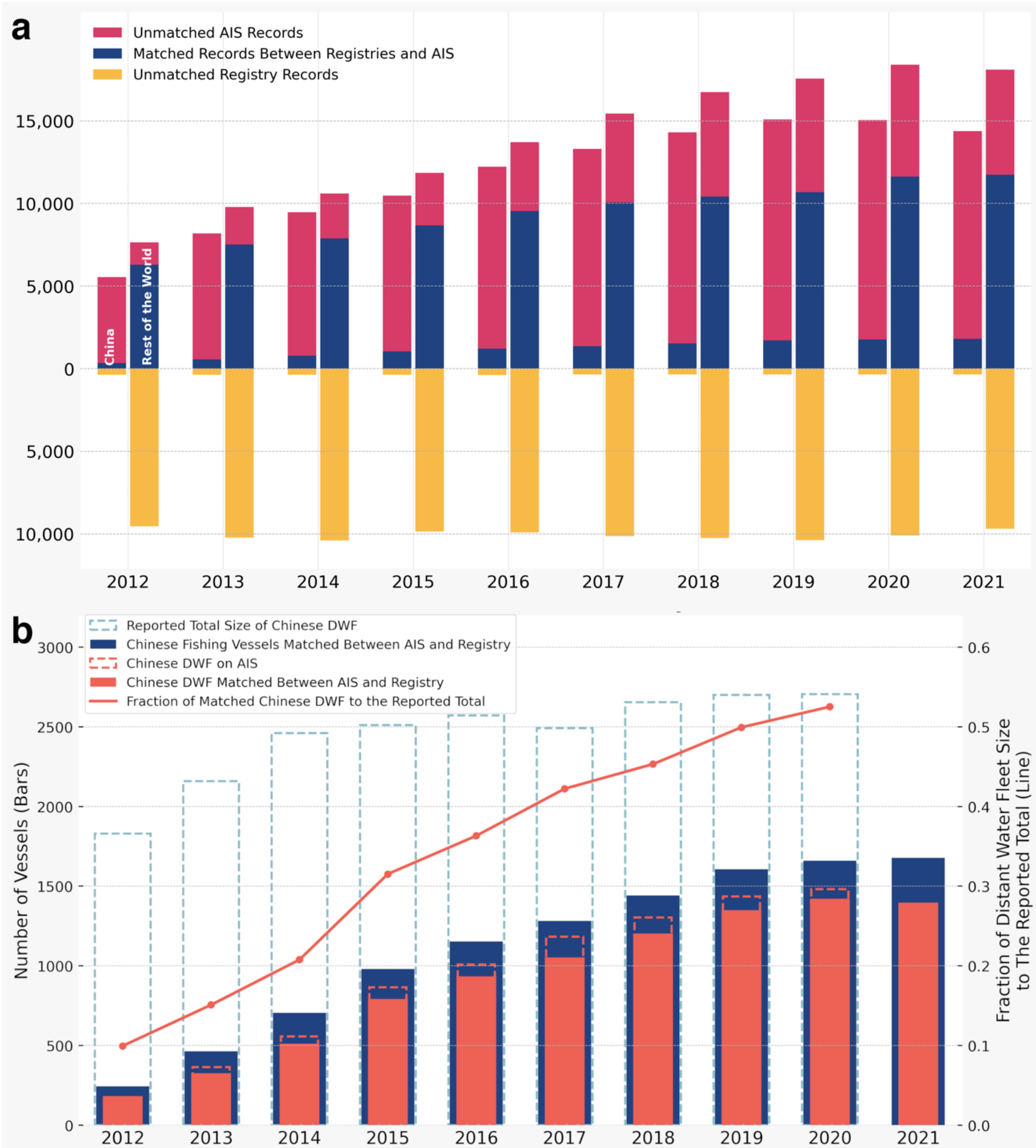

Fig. S5. Number of vessel identities in registry and AIS and Chinese distant water fleet by year.
(a) The graph represents the yearly number of vessel identities (vessels over 24 meters) from 2012 to 2021 between China and the rest of the world. The bars in red represent unmatched AIS records, the bars in blue indicate AIS-registry matched records, and the bars in yellow mean unmatched registry records. **(b)** The graph shows the number of Chinese fishing vessels matched to both AIS and registry by year in 2012-2020. Nearly 90% of the Chinese fishing vessels, over 24 meters, matched to both AIS and registry are detected to have fished on the high seas or within non-neighboring foreign EEZs. Compared to the total distant water fleet reported by the Chinese ministry (dotted bars in light blue), the number of the known Chinese fishing vessels that fished outside China (orange bars) increased over the past years and reached above 50% in 2020. The size of the Chinese distant water fleet for 2021 has yet to be reported.

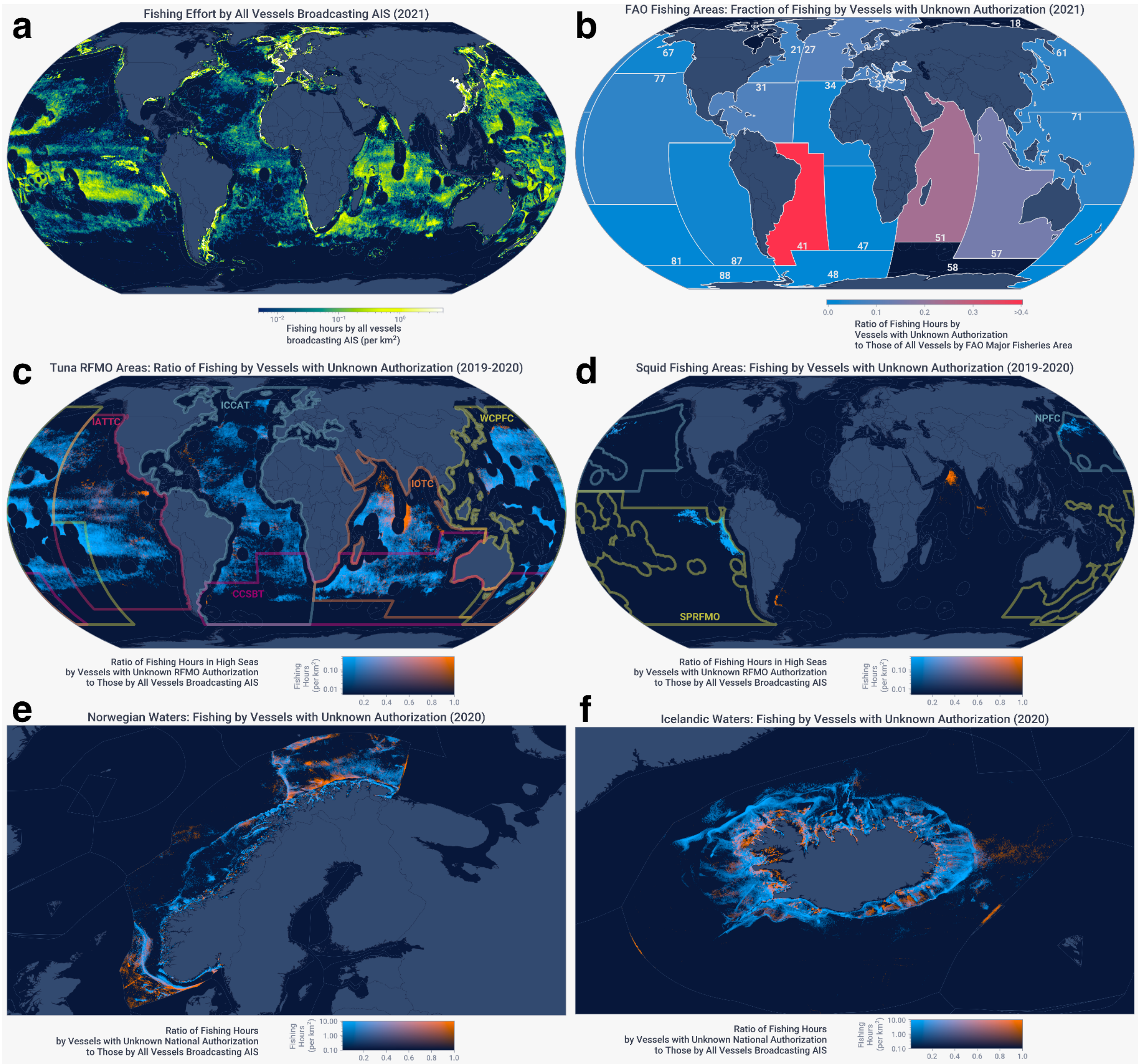

Fig. S6. Fishing effort by authorization known vs. unknown vessels.

(a) All fishing effort by vessels active in AIS in 2021 is binned into 0.2° by 0.2° gridded cells of the world's ocean. **(b)** The map represents the fraction of fishing effort by vessels with unknown authorization from RFMOs by FAO major fishing areas. It only includes fishing effort on the high seas. **(c)** This map shows fishing effort by vessels with authorization known vs. unknown from tuna RFMOs. **(d)** This map shows fishing effort by vessel with authorization known vs. unknown from RFMOs that regulate squid fishing. **(e, f)** The maps represent fishing effort by vessel with authorization known vs. unknown in the EEZs of Norway and Iceland, respectively, based on public registry information (and matched only with AIS, not VMS).

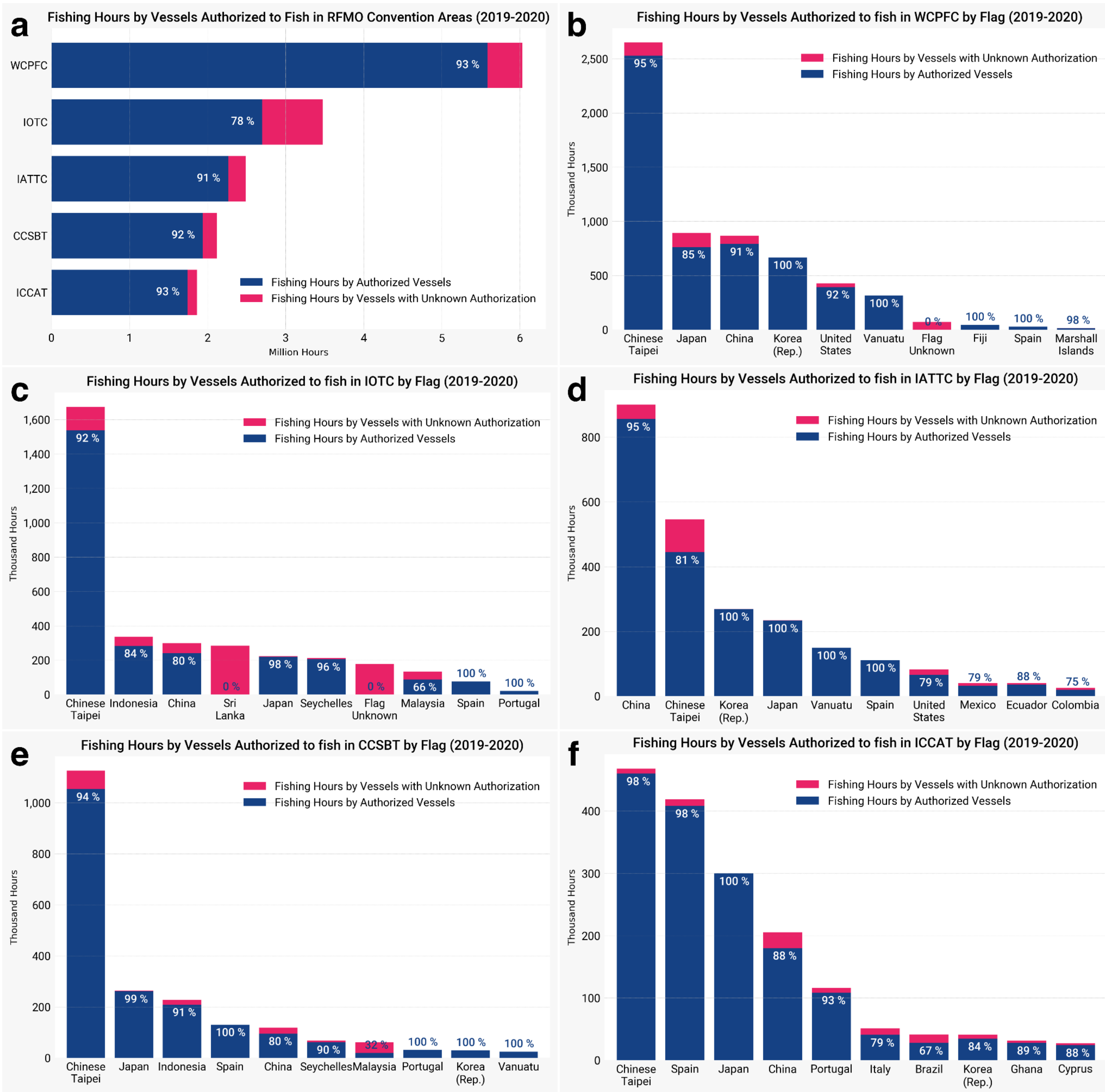

Fig. S7. Fishing effort by vessels with known authorization information by flag State in RFMO convention areas.

(a) Fraction of fishing hours by vessels authorized to fish in each of tuna RFMOs in 2019-2020. The bars in blue represent the total fishing hours carried out by vessels whose fishing activity in AIS match to their authorization time ranges recorded in each registry. The bars in red indicate the total fishing hours by vessels whose authorization is unknown based on public registries. **(b-f)** Fraction of fishing hours by fleets from the top 10 flag States fishing in each of tuna RFMOs in 2019-2020. Unknown flag is assigned to a vessel using invalid MMSIs in AIS unless its identity matches to the registry and consequently its flag can be identified. Due to the unavailability of historical records, authorization information of IATTC and ICCAT is analyzed from May 2019 to the end of 2020.

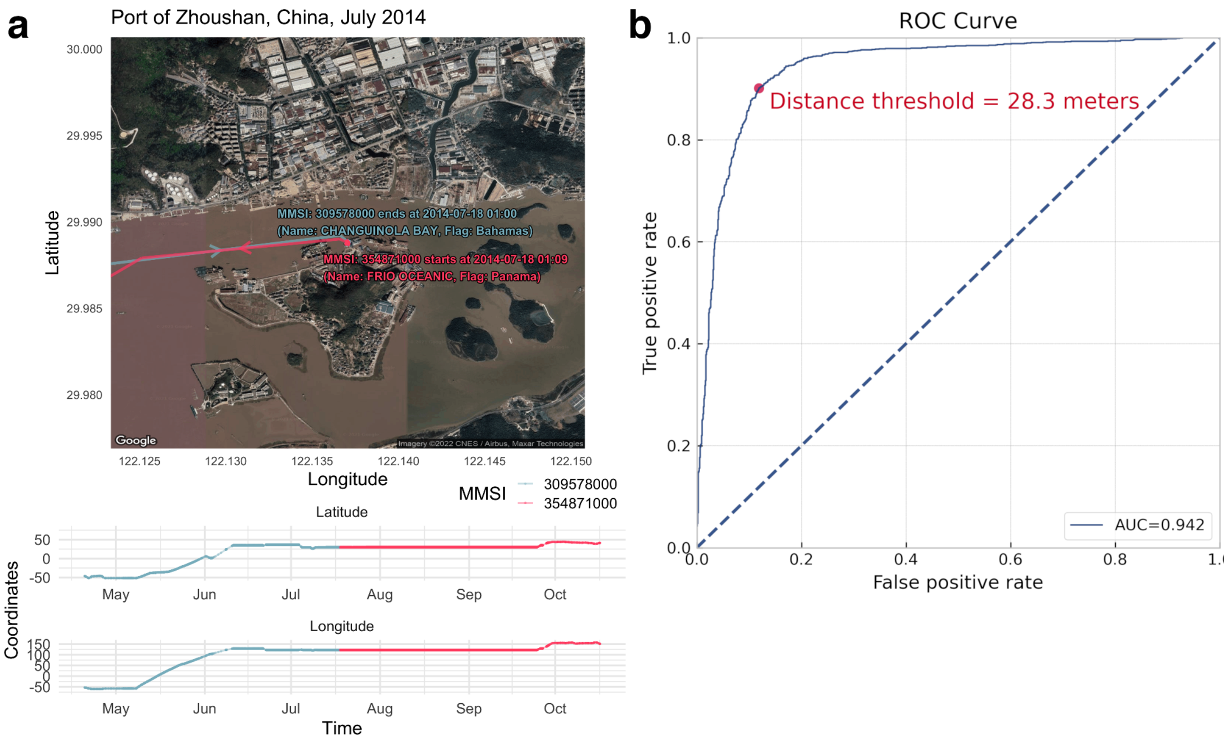

Fig. S8. Development of Vessel Record ID.

(a) An instance of permanent identity switching on AIS plotted on a Google map using spatiotemporal information of two subsequent identities on AIS (using different name, flag, and MMSI numbers). The switching takes place at a port in Zhoushan, China with a time gap of about 10 minutes. **(b)** A receiver operating characteristic, or ROC, curve is used to determine the threshold of the distance between the end point of the previous identity and the start point of the next identity at the port.

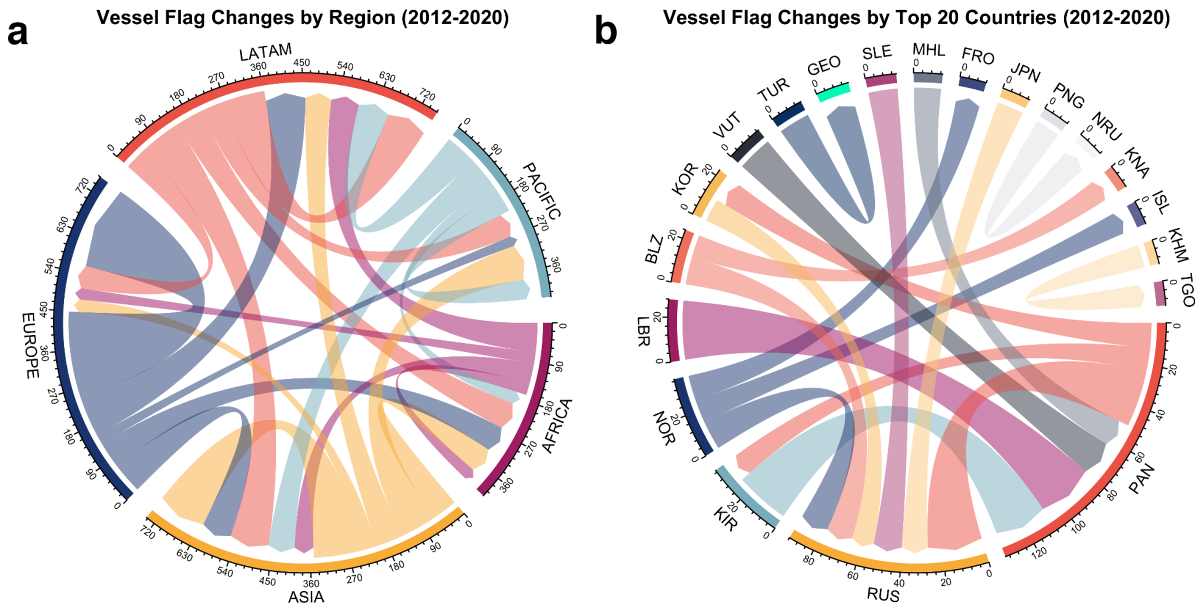

Fig. S9. Vessel flag changes by region and by top reflagging states.

(a) Relationship among regions with regard to reflagging of vessels in 2012-2020 with arrows indicate the flow of flag transfer. Each region is color-coded, and the numbers at the outer axis represent the number of reflagging events. **(b)** Relationship among the top 20 flag States (referred to under three-letter ISO-3166 country code) that are responsible for the most reflagging events in 2012-2020. For instance, Panama (PAN) flagged out to various states in 50 cases while foreign vessels reflagged to Panama in about 70 cases.

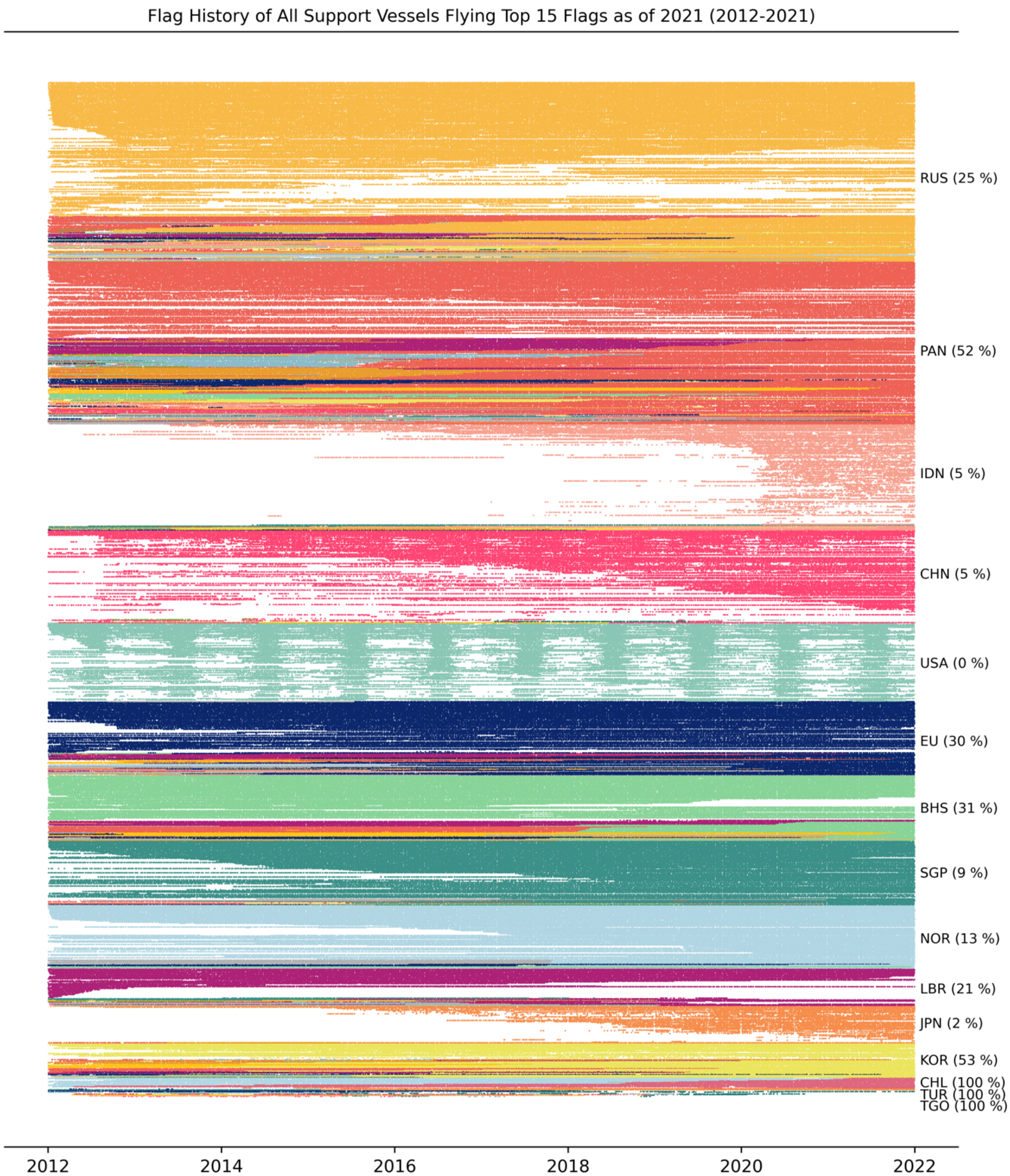

Fig. S10. Flagging history of all supporting vessels flying the top 15 flags by number of support vessels.

Each horizontal line with dots corresponds to a vessel and each dot is color-coded by flag it flies at a given time between 2012 and 2021. On the right are ISO-3166 country codes of flag indicating the flag a vessel is flying on January 1, 2022. Percentages in the parentheses indicate the ratio of vessels that have reflagged at least once between 2012 and 2021 to the total number of vessels flying the flag on January 1, 2022. See Fig. 3 for ISO-3166 country codes.

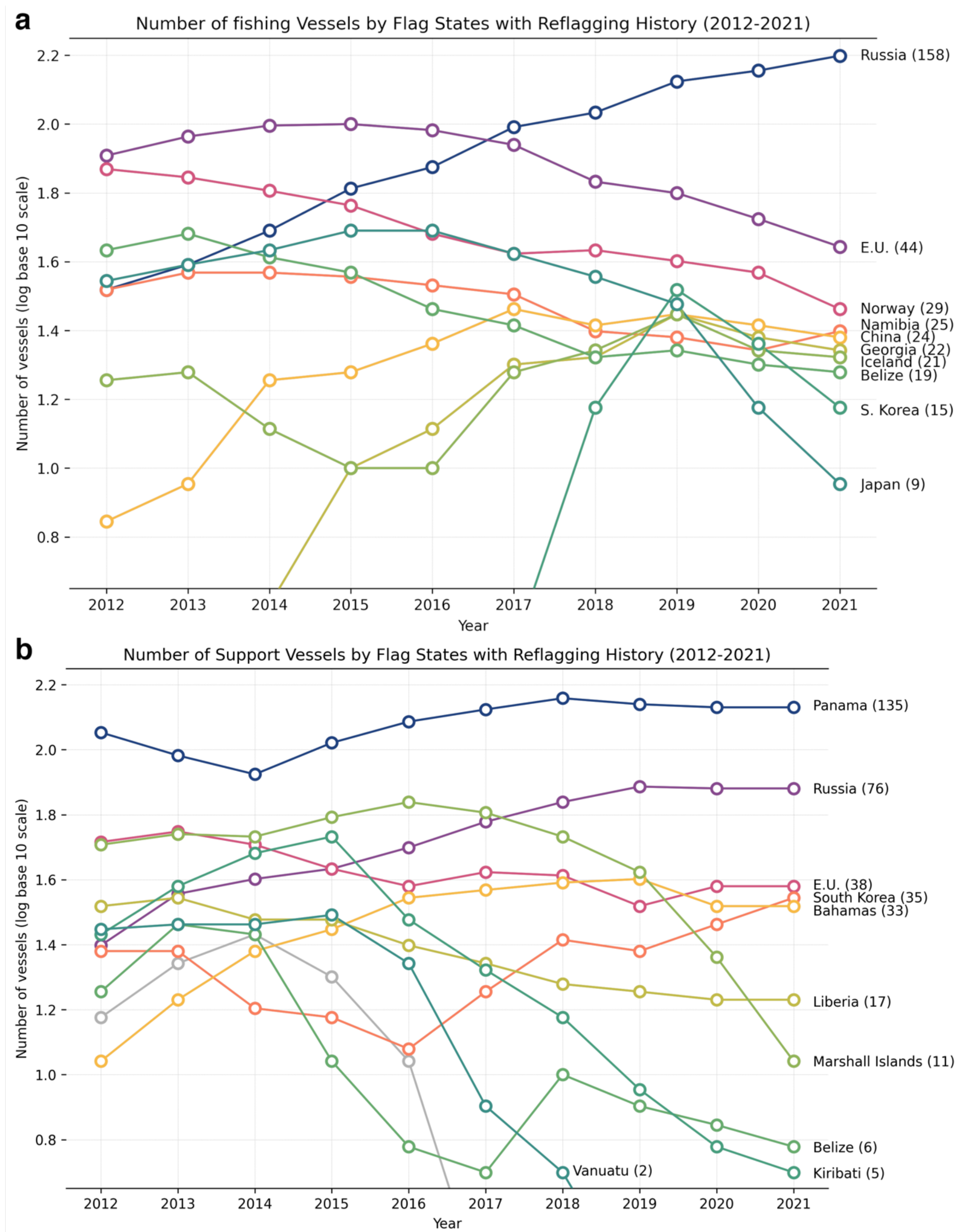

Fig. S11. Number of fishing and support vessels over time with reflagging history by the top 10 flags.

The vertical axis represents the number of vessels in log scale for an easier comparison among flag States, and the numbers in parentheses indicate the number of vessels with reflagging history flagged to each flag States in 2021.

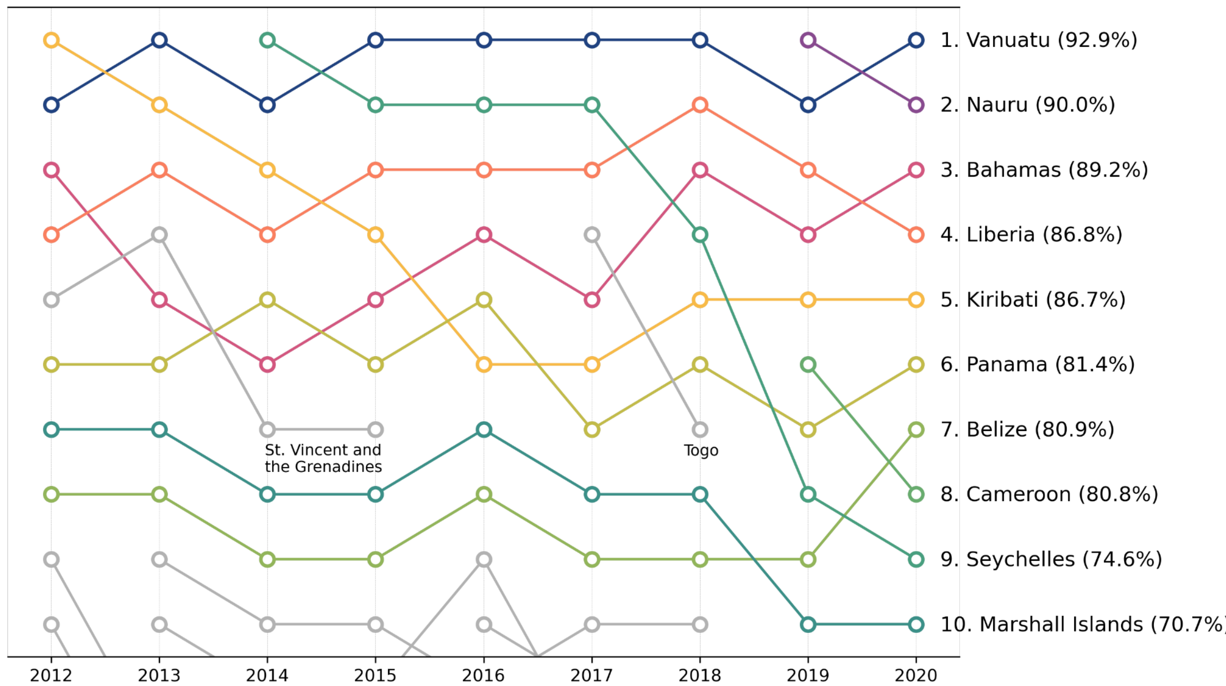

Fig. S12. Ranking of flags with the highest proportion of foreign owned vessels (2012-2020).
Top 10 flag States that represent the highest fraction of foreign owned vessels between 2012 and 2020, including only flag States that have more than 10 foreign owned vessels each year. The percentage numbers in parentheses indicate the proportion of foreign owned vessels to the number of total flagged vessels in 2020.

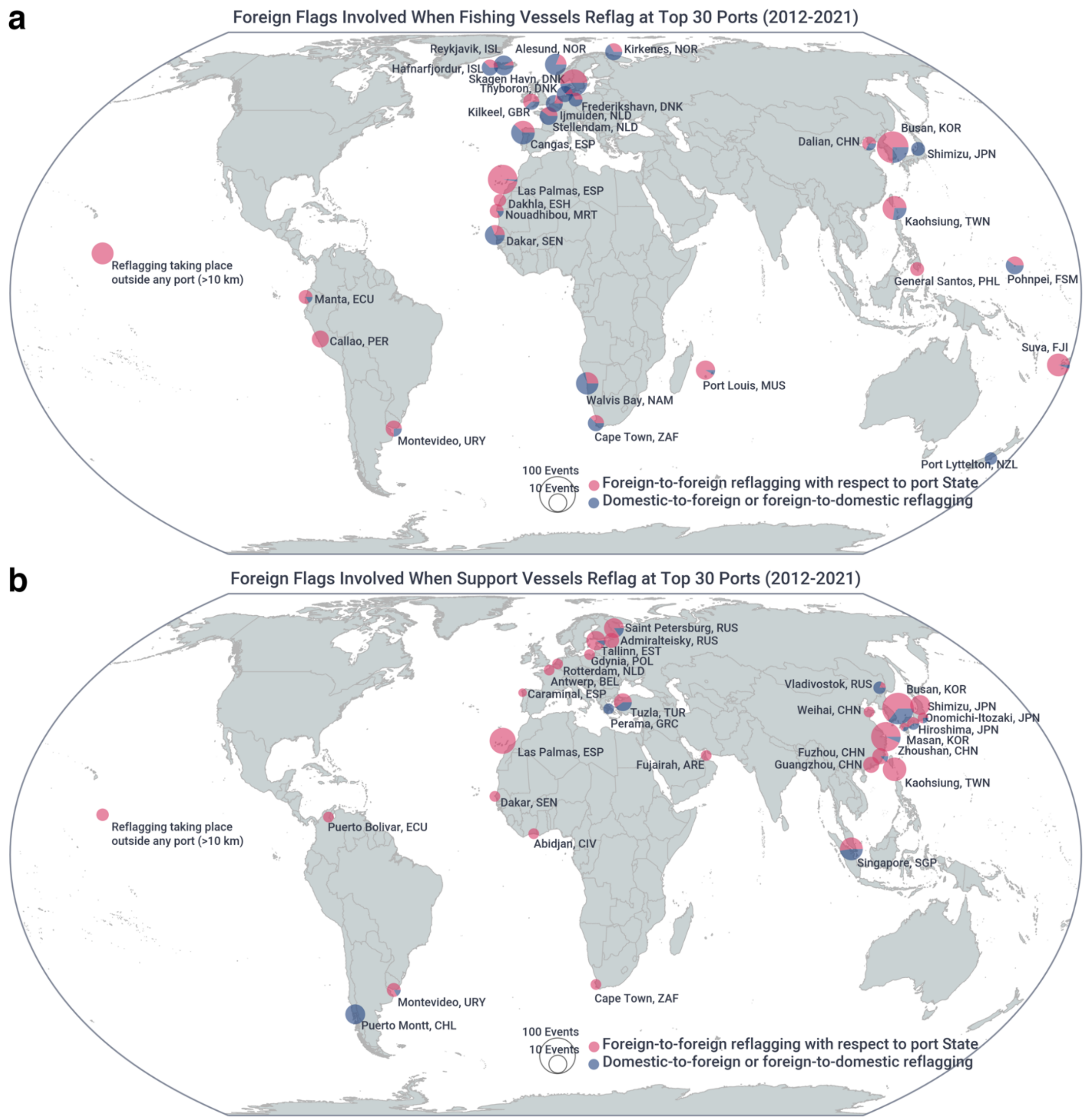

Fig. S13. Top 30 ports where most reflagging events for fishing vessels and support vessels occur. The port for reflagging for **(a)** fishing vessels and for **(b)** support vessels are identified respectively. The size of the circle indicates the number of instances. Blue represents reflagging instances involving one foreign flag and one national flag (either flag-in or flag-out) whereas red indicates reflagging between foreign flags with respect to the flag of the port.

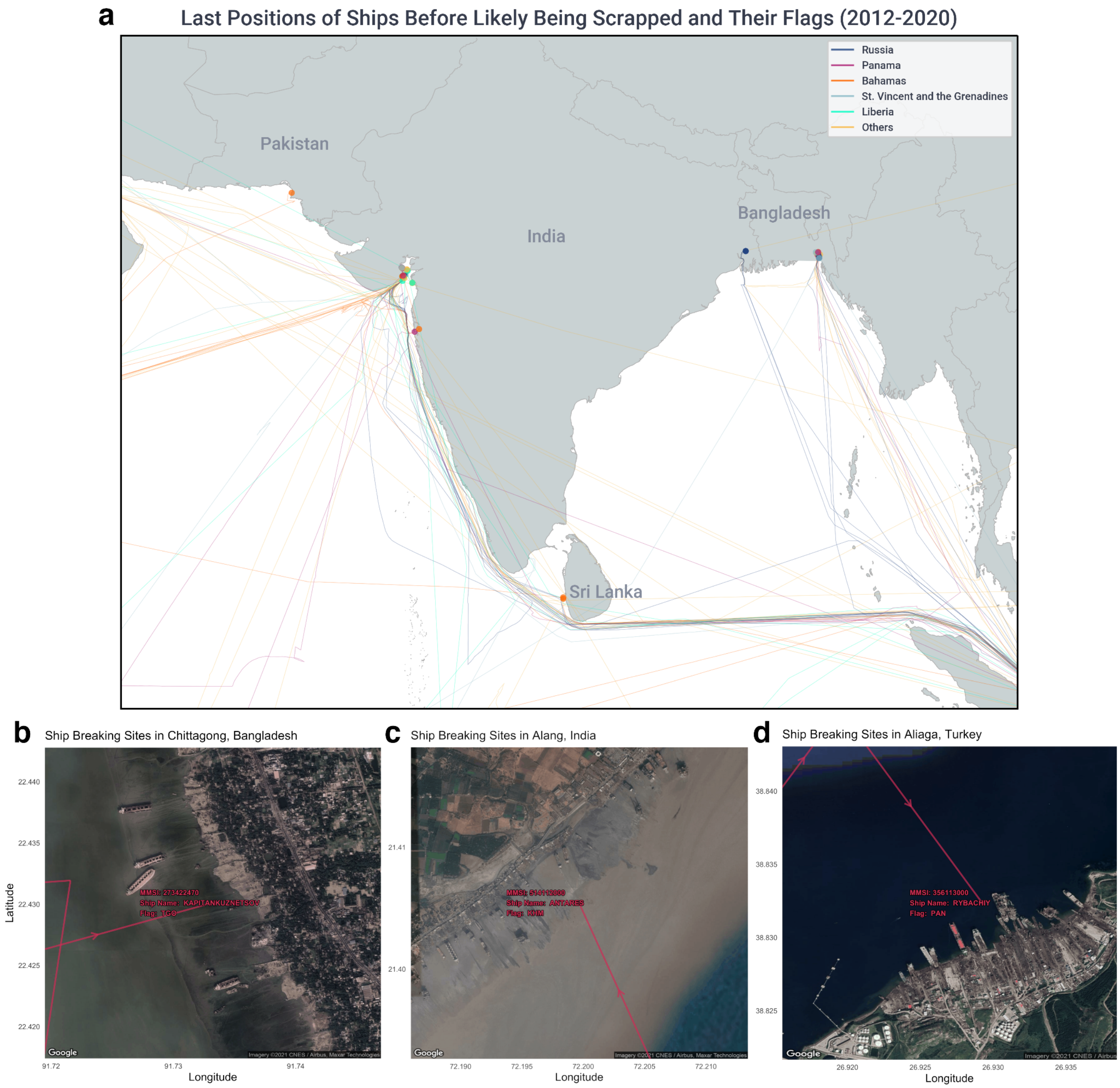

Fig. S14. Last GPS positions of the vessels likely scrapped in South Asia.

(a) AIS tracks of the vessels that arrived at different scrapping sites in Bangladesh, India, Pakistan, and Sri Lanka. The tracks are created by connecting consecutive AIS positions with straight lines, which results in unrealistic lines over the land in some cases. **(b-d)** Top three destinations of these vessels on Google Maps with their ultimate positions.

Fig. S15. *Ocean Star 96* at a scrapping site.

The vessel, *Ocean Star 96* flagged to Nauru, was photographed at a beach in Chattogram, Bangladesh on June 27, 2019 waiting to be scrapped. © Mohammad Islam

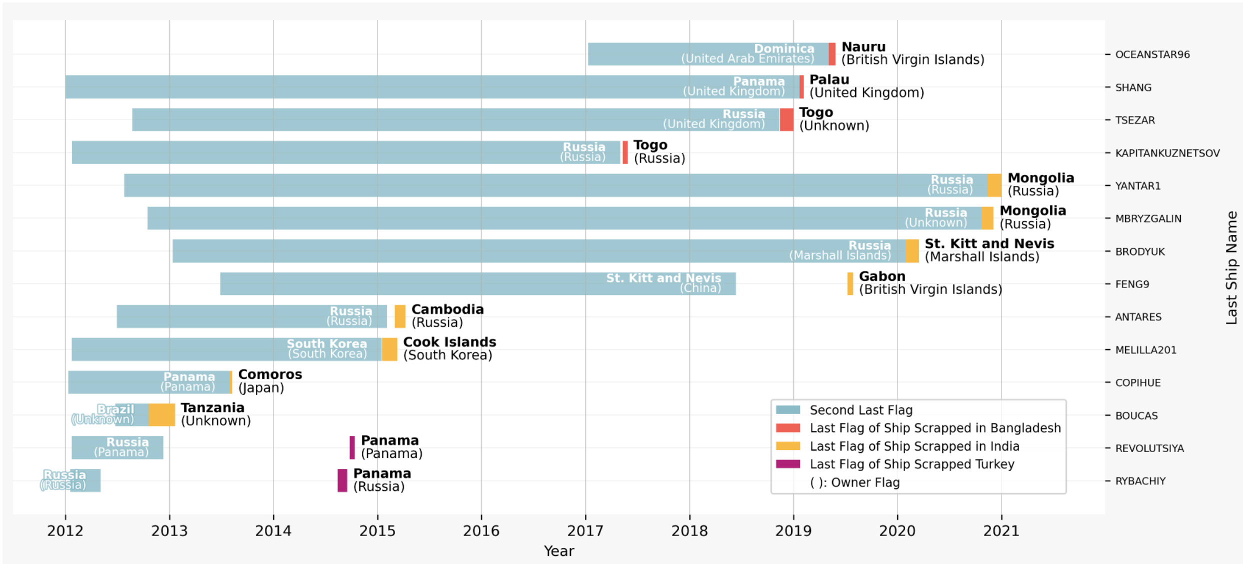

Fig. S16. Ships reflagging before being scrapped.

Selected cases of ships that reflagged and stayed less than six months before being scrapped in India, Bangladesh, or Turkey.

RFMO Code	Full name (website)	Core identity	Characteristic	Vessel class	Authorization	Registered owner info
CCAMLR	Commission for the Conservation of Antarctic Marine Living Resources (https://www.ccamlr.org/en)	Yes	Yes	Yes	Yes	Owner name but no owner flag provided
CCSBT	Commission for the Conservation of Southern Bluefin Tuna (https://www.ccsbt.org/)	Yes	Engine power not provided	Yes	Yes	Yes
GFCM	General Fisheries Commission for the Mediterranean (http://www.fao.org/gfcm/en/)	Call sign and/or IMO number often missing by States	Tonnage and/or engine power not provided for some States	Yes	Only info about vessels being currently authorized no historical data	No
IATTC	Inter-American Tropical Tuna Commission (https://www.iattc.org/)	Yes	Yes	Yes	Only info about vessels being currently authorized no historical data	Yes
ICCAT	International Commission for the Conservation of Atlantic Tunas (https://www.iccat.int/en/)	Call sign and/or IMO number often missing	Tonnage and/or engine power often missing	Yes	Yes	Yes
IOTC	Indian Ocean Tuna Commission (https://www.iotc.org/)	Yes	Engine power unavailable	Yes	Yes	Yes
NAFO	Northwest Atlantic Fisheries Organization (https://www.nafo.int/)	IMO number unavailable	No	No	Only info about vessels being currently authorized no historical data	No
NEAFC	North East Atlantic Fisheries Commission (https://www.neafc.org/)	Yes	Yes	Vessel class often missing for some States	Intermittent update with no historical data provided	No
NPFC	North Pacific Fisheries Commission (https://www.npfc.int/)	Yes	Yes	Yes	Yes	No
SEAFO	South East Atlantic Fisheries Organisation (http://www.seafo.org/)	Yes	Engine power unavailable	Yes	Intermittent update with no historical data provided	No
SIOFA	Southern Indian Ocean Fisheries Agreement (https://www.apsoi.org/)	Yes	Engine power unavailable	Vessel class often missing for some States	Yes	No
SPRFMO	South Pacific Regional Fisheries Management Organisation (https://www.sprfmo.int/)	Yes	Length sometimes missing for some States	Yes	Yes	No
WCPFC	Western and Central Pacific Fisheries Commission (https://www.wcpfc.int/home)	Yes	Tonnage sometimes missing for some States	Yes	Yes	Yes

Table S1. Assessment of public vessel registries from 13 regional fisheries management organizations. This table assesses and compares public vessel registries from 13 regional fisheries management organizations and their vessel identity information. The five categories to be assessed include: 1) core identity (ship name, call sign, IMO number, and flag), 2) characteristic (length overall, gross tonnage, and engine power), 3) vessel classification, 4) authorization for fishing, and 5) ownership information. Cells in green represent sufficient data availability; cells in yellow indicate partial data availability; and cells in red mean no data available.

Country	Data source	Country	Data source
Australia	https://www.amsa.gov.au/vessels-operators/ship-registration/list-registered-ships	Canada	http://wwwapps.tc.gc.ca/Saf-Sec-Sur/4/vrqs-srib/eng/vessel-registrations/advanced-search http://sd.ic.gc.ca/pls/engdoc_anon/mmsi_search.ship
Chile	http://www.sernapesca.cl/area-trabajo/pesca-industrial	Chinese Taipei	https://en.fa.gov.tw/list.php?theme=VR_of_RFMO&subtheme= (Fisheries Agency of Taiwan provides vessels registered to IOTC separately through its website)
Costa Rica	https://www.incopesca.go.cr/acerca_incopesca/transparencia_institucional/datos_abiertos.aspx#HERMES_VTABS_5_1	European Union	https://webgate.ec.europa.eu/fleet-europa/search_en
FFA (Pacific Islands Forum Fisheries Agency)	https://rimf.ffa.int/public/goodstanding/list	Faroe Islands	https://www.teyggjan.fo/Pages/FishingPermit/SearchResult.aspx
Iceland	https://www.mbl.is/200milur/skipaskra/	Korea (Rep. of)	https://www.krs.co.kr/Eng/Exclusive/Ship_Search.aspx
Malaysia	https://www.marine.gov.my/jlm/page/ship-5	Norway	https://www.fiskeridir.no/Yrkesfiske/Registre-og-skjema/Fartoeyregisteret/fartoyregisteret
Panama	The list of vessels is shared by Global Fishing Watch for public vessel tracking on its website.	Peru	The list of vessels is shared by Global Fishing Watch for public vessel tracking on its website.
Russia	https://lk.rs-class.org/regbook/regbookVessel	Seychelles	The list of vessels was shared by Global Fishing Watch through a research partnership.
United States	https://www.dco.uscg.mil/Our-Organization/Assistant-Commandant-for-Prevention-Policy-CG-5P/Inspections-Compliance-CG-5PC-/Office-of-Investigations-Casualty-Analysis/Merchant-Vessels-of-the-United-States/		

Table S2. Source list of country-based vessel registry.

The table indicates the sources of the vessel registries provided by states that this study incorporates into the dataset. The European Union's fleet register provides a consolidated list of vessels from 27 Member States (including Great Britain before 2021). The Pacific Islands Forum Fisheries Agency provides a list of vessels that are authorized to fish within the Exclusive Economic Zones of its 17 Pacific Island Member States.

	FAO Global Record	Vessel identity data used in this study
Scope	• All fishing vessels, carriers, and supply vessels that are assigned an IMO number (~12,000 as of July 2022) submitted by each State's authorities.	• All fishing and fishing support vessels using AIS between 2012 and 2021 (~35,000 as of July 2022).
Access	• Searchable user interface provided on its own website. Download is only available for Member States https://globalrecord.fao.org/	• Data in a CSV tabular format will be made available in the GitHub repository for this paper.
Quality assurance	• 66 States (as of July 2022) submitted their vessel data to Global Record. The data fields provided vary by State. Many States leave some data fields blank. • The Global Record indicates that FAO does not guarantee the accuracy, completeness, or authenticity of the data.	• Authors have developed algorithms to compare multiple public sources to cross-check information and minimize discrepancies and errors that each source may present. A subset of data is also manually reviewed by the authors and other experts to assure the quality of data. • Machine learning inferred data plays an important role in cross-check information from registry sources.
Activity data	• The Global Record is designed to include static information including vessel identity, authorization, event-related records (e.g., inspections, port denials, IUU cases), but not dynamic information (vessel's activity).	• The vessel identity data used in this study contain dynamic information about vessel activity on AIS (active time range, fishing hours per year, and other activity metrics), and is designed to be used by plugging it to other public fishing effort/tracking data.
Update cycle	• The data submission by each State is on a voluntary basis. Consequently, the update dates and frequency appear to vary among States. Some records have not been updated since 2016, and there is no indication when they will be updated, although regular submission is encouraged by FAO.	• The vessel identity data used in this study will be updated every six months, and more frequently later with up-to-date information and incorporating feedback from users.
Historical data	• The Global Record provides limited information about vessel history. Although relative fields are available for States to provide, few States submit this information. Moreover, it is possible that the field does not capture the history of vessels changing flags among different States.	• The vessel identity data used in this study provide the flagging history of vessels with corresponding time ranges.
National authority information	• The Global Record contains fields related to national authorization, inspection data, ownership and port denial history.	• No data on these categories.
RFMO related data	• The Global Record is currently focused on information about the flag State's registry, therefore no data is captured on inter-governmental organizations like RFMOs, although data should be the same across RFMO registries and the Global Record, as all is provided by the flag State.	• The vessel identity data used in this study synthesize information from major RFMOs' public and provides curated data about vessels operating on the high seas and within RFMO Convention Areas.

Table S3. Comparison of characteristics between FAO's Global Record and the vessel identity data of this study.

This table compares FAO's Global Record of Fishing Vessels, Refrigerated Transport Vessels and Supply Vessels with the vessel identity data set used for the analyses in this study.

Flags of convenience					
Antigua and Barbuda (ATG)	Bahamas (BHS)	Barbados (BRB)	Belize (BLZ)	Bermuda (BMU)	Bolivia (BOL)
Cambodia (KHM)	Cameroon (CMR)	Cayman Islands (CYM)	Comoros (COM)	Cook Islands (COK)	Curacao (CUW)
Cyprus (CYP)	Equatorial Guinea (GNQ)	Faroe Islands (FRO)	French International Ship Registry	German International Ship Registry	Georgia (GEO)
Gibraltar (GIB)	Honduras (HND)	Jamaica (JAM)	Lebanon (LBN)	Liberia (LBR)	Madeira (MDI)
Malta (MLT)	Marshall Islands (MHL)	Mauritius (MUS)	Moldova (MDA)	Mongolia (MNG)	Myanmar (MMR)
North Korea (PRK)	Palau (PLW)	Panama (PAN)	Sao Tome and Príncipe (STP)	Sierra Leone (SLE)	Sri Lanka (LKA)
St Kitts and Nevis (KNA)	St Vincent (VCT)	Tanzania (TZA)	Togo (TGO)	Tonga (TON)	Vanuatu (VUT)

Table S4. Flags of convenience reported by the International Transport Workers' Federation. Flags of convenience reported by the International Transport Workers' Federation (ITF) in October 2021 with ISO-3177 country codes in parentheses, where applicable. The French International Ship Registry and German International Ship Registry were not included as they are not explicitly tracked in our dataset as separate from the French and German flags.