**South China Sea Data Initiative, Version 1**

**September 1, 2022**

PIs: Renard Sexton, Nico Ravanilla

Research Assistants: Hanh Linh Tran, Duong Pham, Rocky Intan, Viet Hung Nguyen, Ferth Manaysay, Seoyeon Jenny Kim

## Introduction

This dataset contains geo-located events related to ongoing territorial disputes in the South China Sea, also known as the Biển Đông (East Sea), West Philippine Sea, or (in part) North Natuna Sea.

In this version, reports from 2012 to 2020 are included (9 years).

The data are structured in two levels:

- Events: An event is a discrete occurrence in the basin at a specific location, that fits the criteria for inclusion. These are rows within the geo-located "event" dataset.
- Reports: A report is an individual news item that provides information about the activities that occurred and the location of an event. Most events have more than one report, although some events only have a single report in the dataset.

Because these events are drawn from publicly available news sources, they are by no means a comprehensive accounting of all the clashes that have taken place in the westernmost Pacific Ocean during the decade in question. Indeed, in speaking with journalists in several of the littoral countries, a considerable number of incidents are not reported on, either due to political pressure, lack of capacity or insufficient verifiable details regarding what took place.

The SCSDI project is currently collecting other data, drawn from local fishermen, to help fill in these gaps. But for the purposes of this news-based dataset, the limitations should be clear: relatively high-profile incidents, events that go "viral," and those that involve regional powers tend to get covered more than small fisherfolks that are pushed back from traditional fishing grounds by e.g. the China Coast Guard, or are repulsed from fishing in a disputed EEZ.

## Geo-coding

Some events in the dataset are relatively easy to geo-locate. Clear reporting that describes nearby features, or IMO/MMSI numbers or ship names that can be tracked using AIS historical data allows for quite precise latitude and longitude for the events. Unfortunately, in practice many events lack the features required for easy geo-location.

A few common challenges:

- Ships that are not easily identifiable by IMO/MMSI numbers
- Event took place in location where AIS stations could not pick up the trackers, or trackers were turned off. Small “class B” transponders on fishing vessels are especially prone to not being picked up. Also fishermen sometimes deliberately turn off AIS when in disputed areas. Vessels with “class A” transponders are often detectable, but sometimes choose to turn off AIS.
- In the absence of clear AIS data, news stories, sometimes open-source maritime domain awareness sleuths or fishing communities might upload relevant information on social media, which can help. But these are not always available.

Recognizing the differences in location verifiability, our dataset has a set of “uncertainty levels” that indicate the level of precision that we can code an event to. In the accompanying shapefiles, we draw “uncertainy polygons’ that show the approximate zone where the event likely took place.

- Level 1: precise location, lat/lon (0.08 decimal degree radius)
- Level 2: relatively precise (0.4 decimal degree radius)
- Level 3 & 4: some uncertainty (0.7 decimal degree radius)
- Level 5: some information but quite uncertain (1.8 decimal degree radius)
- Level 6: very low precision, high uncertainty (7.2 decimal degree radius)

**Website Data Fields**

**EVENTID COUNTRY EVENT DATE YEAR MONTH DAY TIME LATITUDE LONGITUDE LEVEL RADIUS NOTE ON LOCATION LOCATION SOURCE NOTES ADDITIONAL LOCATIONS**

- **EVENTID:** unique id for each event
- **COUNTRY:** which countries’ vessels were involved
- **EVENT DATE:** date of event, as reported (local time)
- **YEAR:** which year
- **MONTH:** which month
- **DAY:** which day
- **TIME:**any precise time information (local time)
- **LATITUDE:** geo-located latitude (determined by RA team)
- **LONGITUDE:** geo-located longitude (determined by RA team)
- **LEVEL:** level of precision of location
- **RADIUS:** decimal degree radius size for shapefile (uncertainty polygon)
- **NOTE ON LOCATION:** any notes on how location was ascertained
- **LOCATION:** notes on location itself
- **SOURCE:** list of sources for reports and location
- **NOTES:** descriptive notes

- **ADDITIONAL LOCATIONS:** Does this event have more than a single Lat/Lon associated with it in the geo-coded file? If you want this information please visit our Dataverse: https://doi.org/10.7910/DVN/GCBWA6

**Sources**

| Source | Country | Language | Website |
|---|---|---|---|
| Thanh Nien | Vietnam | Vietnamese | https://thanhnien.vn/tin-tuc/tim-kiem/Ymnhu4NuIMSRw7RuZw==/bien-dong.html |
| ABC News | Australia | English | https://www.abc.net.au/ |
| Reuters | US | English | https://www.reuters.com/news/archive/south-china-sea |
| Bao Phap Luat | Vietnam | Vietnamese | https://baophapluat.vn/ |
| An Ninh Thu Do | Vietnam | Vietnamese | https://anninhthudo.vn/ |
| Tuoi Tre | Vietnam | Vietnamese | https://tuoitre.vn |
| VOA Tieng Viet | Vietnam | Vietnamese | https://www.voatiengviet.com/ |
| BBC Tieng Viet | Vietnam | Vietnamese | https://www.bbc.com/vietnamese |
| VietTimes | Vietnam | Vietnamese | https://viettimes.vn |
| CGTN | Chinese | Chinese | https://www.cgtn.com/ |
| 南海研究论坛 | China | Chinese | http://www.nhjd.net |
| 中华网新闻 | China | Chinese | https://news.china.com |
| 新浪新闻 | China | Chinese | https://news.sina.com.cn |
| 凤凰网军事 | China | Chinese | https://mil.ifeng.com |
| 海外网 | China | Chinese | http://nanhai.haiwainet.cn/ |
| Soha | Vietnam | Vietnamese | Soha.vn |
| Asia Maritime Transparency Initiative | US | English | https://amti.csis.org |
| The Diplomat | US | English | https://thediplomat.com |
| 网易 | China | Chinese | http://dy.163.com |
| 百科 | China | Chinese | https://baike.so.com |

| | | | |
|---|---|---|---|
| 新浪军事 | China | Chinese | http://mil.news.sina.com.cn |
| South China Morning Post | Hong Kong | English | https://www.scmp.com |
| The Defense Post | US | English | https://thedefensepost.com |
| Military Times | US | English | https://www.militarytimes.com |
| 快资讯 | China | Chinese | https://www.360kuai.com |
| CGTN | China | Chinese | https://news.cgtn.com |
| Zing News | Vietnam | Vietnamese | https://news.zing.vn |
| 西陆网 | China | Chinese | http://junshi.xilu.com |
| 超级大本营 | China | Chinese | https://lt.cjdby.net |
| ChinaDaily | China | English | http://www.chinadaily.com.cn/ |
| Gatra | Indonesia | Indonesian | https://www.gatra.com/ |
| Jakarta Post | Indonesia | English | https://www.thejakartapost.com |
| Kompas | Indonesia | Indonesian | https://kompas.com |
| CNN Indonesia | US | Indonesian | https://www.cnnindonesia.com/ |
| Kumparan | Indonesia | Indonesian | https://kumparan.com/ |
| BBC Indonesia | UK | Indonesian | https://www.bbc.com/indonesia |
| Antara | Indonesia | Indonesian | https://www.antaranews.com/ |
| Mongabay | US | Indonesian | https://www.mongabay.co.id/ |
| Detik | Indonesia | Indonesian | https://www.detik.com |
| Tribun News | Indonesia | Indonesian | https://wartakota.tribunnews.com/ |
| Okezone | Indonesia | Indonesian | https://news.okezone.com/ |
| Medcom.id | Indonesia | Indonesian | https://www.medcom.id/ |
| KKP News | Indonesia | Indonesian | https://news.kkp.go.id/ |
| TNI News | Indonesia | Indonesian | https://www.tni.mil.id/ |
| Merdeka | Indonesia | Indonesian | https://www.merdeka.com/ |
| Bisnis | Indonesia | Indonesian | https://ekonomi.bisnis.com |
| Philippine Daily Inquirer | Philippines | English | https://www.inquirer.net/ |
| Rappler | Philippines | English | https://www.rappler.com/ |

| | | | |
|---|---|---|---|
| Manila Bulletin | Philippines | English | https://mb.com.ph/ |
| Philippine Star | Philippines | English | https://www.philstar.com/ |
| Nikkei | Japan | Japanese | https://www.nikkei.com/ |
| Asahi | Japan | Japanese | https://www.asahi.com/ |
| NHK | Japan | English/Japanese | https://www.nhk.or.jp/ |
| Vietnamnet | Vietnam | Vietnamese | https://vietnamnet.vn/ |
| Radio France International | France | Vietnamese | http://www.rfi.fr/vi/ |
| VTC News | Vietnam | Vietnamese | https://vtc.vn/ |
| Báo Quảng Ngãi | Vietnam | Vietnamese | http://baoquangngai.vn/ |
| Báo Bình Định | Vietnam | Vietnamese | http://www.baobinhdinh.com.vn/ |

**Inclusion criteria**

| **Methodology for searching and coding news events** | | |
|---|---|---|
| 1 | Pick a list of articles to focus on; coordinate with other RAs so there is no overlap in assigned articles | |
| 2 | On the "Source" sheet, fill in the details for each of the source article | |
| 3 | For each of the source articles in your set list | |
| | a. | Use search function to start searching for a set of broad-SCS related keywords (e.g. South China Sea, EEZ, 9-dash-line, etc.) |
| | b. | As relevant news articles are coded, add new keywords on the list of searched keywords (e.g. names of islands and reefs, names of ships, etc.) |

| | | |
|---|---|---|
| | c. | If an event is deemed relevant for coding, use information from the news article to fill in the details in your respective coding sheet |
| | d. | For an event to be useful information, we need clear date (in UTC timezone) and IMO and/or MMSI, but code events even if these information are not currently available |
| | e. | If some information are not included in the news article at hand, use Google or other source articles in the Sources list to find the information (e.g. if you have the name of the ship and not the MMSI, google the name of the ship to find the MMSI) |
| | f. | As ships are identified, add the ship details in the "Vessels" sheet, so effort is not duplicated |
| | g. | Repeat process with new keywords on the list |
| 4 | Rules of inclusion: include if answer is "yes" to all of the questions below | |
| | a. | Did the event occur in SCS? |
| | b. | Did the even involve ships from littoral states? (change to a column about whether the incident included a littoral vessel or not) |
| | c. | Does the news article have enough information to code an event? (Note: not enough information means you can't even code the event date, country, and some rough location. If you can, then event should be coded. |
| | d. | Is the article factual and not opinion-based? |