ESSnet Big Data

Specific Grant Agreement No 1 (SGA-1)

https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata
https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en

Framework Partnership Agreement Number **11104.2015.006-2015.720**

Specific Grant Agreement Number **11104.2015.007-2016.085**

Work Package 4

AIS Data

Deliverable 4.2

Deriving port visits and linking data from Maritime statistics with AIS-data

Version 2017-02-10

Prepared by:	Tessa de Wit (CBS, Netherlands) Anke Consten (CBS, Netherlands) Marco Puts (CBS, Netherlands) Christina Pierrakou (ELSTAT, Greece) Michal Bis (GUS, Poland) Anna Bilska (GUS, Poland) Olav Grøndal (SD, Denmark) Øyvind Langsrud (SSB, Norway)

ESSnet co-ordinator:

Peter Struijs (CBS, Netherlands)
p.struijs@cbs.nl
Telephone : +31 45 570 7441
Mobile phone : +31 6 5248 7775

Index

1. Introduction

The aim of WP 4 is to investigate whether real-time measurement data of ship positions (measured by the so-called AIS system) can be used 1) to improve the quality and internal comparability of existing statistics and 2) to produce new statistical products relevant to the ESS. Five National Statistical Institutes participate in WP4: the national statistical institutes of The Netherlands (Work package leader), Denmark, Greece, Norway and Poland.

WP4 is subdivided into two phases. SGA-1 focuses on creating a common database, linking AIS-data to maritime statistics and constructing sea traffic analyses (February 2016-July 2017, See Annex 1 for a more detailed description of SGA-1 of WP4.) Methodological, qualitative and technical results, including intermediate findings, will be used as inputs for SGA-2. SGA-2 focuses on the calculation of emissions and future perspectives for AIS data as source data for new statistical output (August 2017 until May 2018). See Annex 2 for a more detailed description of SGA-2.

The current SGA will result in three deliverables:

Deliverable 1:"Creating a database with AIS-data for official statistics: possibilities and Pitfalls" (delivered 21-7-2016)

Deliverable 2: "Deriving port visits and linking data from maritime statistics with AIS data" (the current deliverable)

Deliverable 3: Sea traffic analyses using AIS-data (to be published July 2017)

In this deliverable we report on the activities and the decisions we made in researching the possibility to link AIS data to maritime statistics. As AIS data does not contain any direct information on the goods or passengers transported, the first goal was to link AIS data to maritime statistics on port visits (The F2-table). In order to do this, a reference frame of maritime ships had to be constructed. Then, Poland and the Netherlands tried to link the AIS data to the port visit statistics. Finally, the whole working group investigated other possibilities of improving current maritime statistics using AIS data.

Chapter 2 gives a short introduction into AIS data. In order to link AIS data to maritime statistics a reference frame of maritime ships is needed. Chapter 3 describes how this reference frame of maritime ships in European waters was constructed based on AIS data. Next, this reference frame was used to determine port visits of which the data were linked to the maritime statistics on port visits (Table F2). Chapter 4 describes the method to link AIS data to maritime statistics on port visits and shows some preliminary results from this linking. Chapter 5 describes further possibilities to improve current statistics using AIS data on a European level. Finally, chapter 6 describes conclusions from the first year of WP4.

2. Quality of AIS data

Maritime traffic increased exponentially over the last decades. This asked for different solutions to ensure safety at sea. Based on GPS-technology, the Automatic Identification System (AIS) broadcasts the location and status information of ships over a radio channel, making it possible to detect other ships and exchange other data with nearby ships. AIS data is collected at numerous

places around the world[1]. Within Europe, a.o. EMSA, Kystverket, Hellenic Coastguard, Dirkzwager and Marine Traffic collect the data. The data can be used by a vast number of applications, and some apps are already built on the basis of these data (e.g. SEAiq, iNavx and Boat Beacon).
AIS data can also be of use to create new marine traffic statistics, which could be interesting for estimating emissions or identifying locations at sea with a critical amount of traffic. In the nearer future, AIS data can be used as a backbone for the maritime transport statistics: based on AIS we investigate how a ship visits different ports.

AIS data consists of different messages, containing different types of information each with their own unique ID:

- Static data (information on ship characteristics)
- Dynamic data (information on ship movements)
- Voyage data (information on a current voyage)

In our case, we considered the datasets from Dirkzwager. This dataset contains 6 months of AIS data (8 October 2015- 12 April 2016) and contains AIS data from land based stations only, covering Europe and some non-European countries. Satellite data is not included.

We divided the messages into:

- Location messages– dynamic data transmitted every 2-10 seconds (depending on speed) or every 3 minutes while at anchor (Message-ID: 1, 2, 3 or 21).
- Identity messages – static and voyage data transmitted every 6 minutes (Message-ID: 5). See Annex 3 for a description of the variables in both the location messages and the identity messages.

AIS data can contain different kind of errors: technical- and human errors.

Technical errors - related to dynamic data such as position of ship, speed, course, rotation which comes from AIS device (sensors, cables and antenna).

There are several types of technical errors of AIS device (Bosnjak et al, 2012):

- *t1* - no transmission / AIS device stops transmitting information;
- *t2* - no reception on working frequencies/ AIS device stops transmitting on all channels (n);
- *t3* - general error / AIS device stops transmitting on all channels (n);
- *t4* - no information on position / AIS device continues functioning using internal GPS device;

[1] Normally, an AIS-Receiving station using an external antenna placed 15 meters above sea level, will receive AIS information from AIS-equipped vessels that sail within a range of 15-20 nautical miles around it. Base stations located at a higher altitude may extend the range up to 40-60 NM depending on factors such as: elevation, antenna type, obstacles around the antenna and weather conditions. This can be supplemented by Satellite AIS data, enabling the monitoring of vessels beyond coastal regions, including the oceans. Satellite coverage is not as good as data from base stations.

- *t5* - no speed over ground / AIS device continues functioning but does not show information on speed;
- *t6* - no course over ground / AIS device continues functioning but does not show information on course;
- *t7* - no information on the rate of turn (device showing intensity of alteration of course and side to which the vessel will turn).

As AIS is a radio signal, technical errors can also arise due to meteorological or magnetic factors disturbing the transmission of the radio signal. These errors can affect every part of the message, see page 10.

Human errors – related to static (ship number, ship's name, call sign, type, length) or voyage data (draught, destination) which are manually entered in the AIS device so therefore are a common cause of errors.

These values should be entered during installation of AIS instrument (static) or if voyage information changes. It is worth noting that voyage data must be manually updated after each port visit.

Annex 4 describes the systematic errors. Most of these errors are due to faulty or missing input by the ship crews. Apart from these systematic errors, all of the parameters can be erroneous due to technical issues (e.g. meteorological factors, distance to receiver). These errors can take any form. This can for example result in a wrong ship identifier. AIS quality thus depends on correct installation of AIS device, frequent manual updates of information, and technical devices. Most of the issues we deal with by detecting and removing erroneous data. For example, unreadable messages are already filtered out during preprocessing (see Deliverable 1). Issues concerning the ships' identity we deal with by constructing a reference frame (see next chapter). As the amount of data is huge, there are many errors. However the amount of remaining data is still ample for further analyses. Chapter 4 describes how this problem can be solved.

In our opinion we should build a common database of known errors which can be useful for validation, processing and analyzing data.

To investigate the issue of coverage by AIS receivers, we also plan to compare national data from Denmark, Greece and Poland to the Dirkzwager data. Another issue Poland wants to investigate is the time of arrival. The declaration from the maritime statistics will be compared to the time of arrival measured from AIS. We will describe the research into the quality of AIS data and improvements in a separate document about the quality of AIS data.

Figure 1: Overview of type of errors in AIS

3. Building a reference frame of ships in European waters based on AIS-data

Not only transport ships, such as bulk ships and container ships navigate the seas, but also fishing ships and yachts. All of these ships emit AIS data. Therefore, translating AIS data into maritime statistics means filtering out maritime ships only. However, the Location messages emitted by ships, containing information needed to determine the location of a ship, only contain the MMSI number (Maritime Mobile Service Identity). These MMSI's also come from non-maritime ships that should not be included for maritime statistics. Maritime ships can be identified on the basis of their so called IMO number (International Maritime Organization) [2]. Furthermore, not all statistical offices (can) collect MMSI or IMO numbers, some collect call sign as a ship identifier. Thus, the reference frame of maritime ships should include all three identifiers: MMSI, IMO and call sign. Identity messages also contain information on the flag and the type of ship. Therefore, the following information from AIS is needed for a reference frame of maritime ships. See Table 3.1 for an overview of information needed for a reference frame of maritime ships.
These three identifiers are not included in the Location messages, only in the Identity messages containing the static and voyage related data. Thus, these Identity messages have to be used to build a reference frame and filter out maritime ships on the basis of available MMSI-IMO couples. To build a reference frame, we initially linked all MMSI-IMO couples and filtered out couples with invalid IMO or MMSI numbers, purely based on length and for IMO the check digit (see: https://en.wikipedia.org/wiki/IMO_number).

The code can be found here: https://github.com/mputs/WP4/tree/master/aisframe2

Table 3.1 information from AIS needed for creating a reference frame of maritime ships

Field Name	Type (Format)	Description
IMO	String	An International Maritime Organization number is 7-digit unique number assigned to seagoing ships under the SOLAS regulation. It is assigned to the hull regardless of a name, flag or owner changes. The last digit is a check digit to validate the first 6 digits.
MMSI	integer	A Maritime Mobile Service Identity (MMSI) is a 9-digit unique number assigned to ships that have specific types of equipment such as AIS and (Very) High Frequency radio with Digital Selective Calling. This MMSI-number is sent in a digital form over a radio frequency channel to uniquely identify ship stations, ship earth stations, coast stations, coast earth stations and group calls.
Name	String	Name of a vessel
Callsign	String	Call sign of a vessel is a worldwide unique identifier for ships and boats assigned by their national licensing authorities Each call sign begins with the Call Sign alphanumeric prefix that indicates nationality. The prefix is usually followed by 2 or 3 alphanumeric characters.
Flag	String	2 Digit ISO 3166 Country Code of a vessel flag
Type	String	Type of a vessel

[3] Other ships do not have an IMO-number: ships solely engaged in fishing, ships without mechanical means of propulsion, pleasure yachts, ships engaged on special service, hopper barges, hydrofoils and air cushion vehicles, floating docks and structures classified in a similar manner, ships of war and troopships, wooden ships.

Still, AIS data does not provide sufficient information for a complete reference frame of maritime ships as maritime statistics require more specific information on the identity of vessels. For example, the type of vessel defined in AIS data is less detailed than required by maritime statistics based on Directive 2009/42/EC. (Note that there might be possibilities to infer these variables from AIS: e.g. inferring type of vessel from type of terminal visited). Another variable missing from AIS data is deadweight. Therefore, in order to get more detailed ship information, AIS data has to be linked to existing ship dictionaries, such as:

- Lloyd's Register: a register that has to be paid for and can be bought from parties such as Dirkzwager

- Open data (e.g. http://data.okfn.org/data/warrantgroup/imo-vessel-codes): this data is available freely on the internet, however legal issues surrounding this are not clear, and continuation of data availability cannot be guaranteed.

- Web scraping methods: data is available on websites such as www.marinetraffic.com or www.vesselfinder.com. Although it cannot be downloaded, there are techniques to extract the data from the website. Like open data, this can be obtained for free, but legal issues and continuation of data availability also render this a bit unsure.

The resulting reference frame of maritime ships in European waters from AIS data, where our reference frame is linked to existing ship dictionaries, could be used as the backbone of maritime statistics. The next chapter describes how the reference frame is further constructed and how aspects such as duplicates are handled.

4. Linking AIS data to maritime statistics: Port visits

When linking AIS data to maritime statistics, the obvious first step is to start with Port visit-statistic as these contain variables that are available in AIS data or in the reference frame of maritime ships.

This statistic of port visits (the so-called F2-table) covers visits to European seaports by vessels (gross tonnage >=100) to load or unload goods or passengers if their voyage was undertaken wholly or partly at sea. This includes ports at which vessels, due to their draught or other reasons, cannot call and which are partially loaded/unloaded outside a port area and whose place of destination is a registration port. The domain also includes cargo dispatched to or collected from offshore installations e.g. offshore drilling units. However, bunkering is not included.

Table 4.1: data set F2

Variables	Coding detail	Nomenclature
Reporting port	Five-character alphanumeric	Selected EEA ports in the port list
Direction	One-character alphanumeric	Inwards, outwards (1, 2)
Type of vessel	Two-character alphanumeric	Type of ship
Size of vessel	Two-character alphanumeric	Gross tonnage size classes
Data: Number of vessels; Gross tonnage of vessels.		

A maritime transport survey is conducted on the basis of Directive 2009/42/EC of the European Parliament and of the Council on statistical returns in respect of carriage of goods and passengers by sea. Countries have different sets of information available to assemble these maritime statistics. The number of ships in AIS data and in maritime statistics will not completely overlap as they have their own underlying population. Maritime data is based on survey data from maritime ships only which can contain omissions and errors, for example wrong ports are sometimes declared (see also Table 5.1: Current problems in maritime statistics). AIS data on the other hand are based on the radio signal that detects all ships with a transmitter. Thus, not only maritime ships transporting goods are detected: but also non-maritime ships such as inland ships or maritime ships, not transporting goods, are detected. Ships are obliged to turn on their AIS system while travelling in inland waterways and ports. It is possible to turn off the AIS system, this only can happen at sea and brings a risk of under coverage of ships in AIS data, but our future algorithms will help solve this as visible parts of the journey can be used to infer missing parts of the journey, so it is still possible to compose the whole journey of a ship.

To generate port visits on the basis of AIS data, the reference frame described in chapter 3 was used to investigate which ships had been in the port on one day for Poland (Świnoujście) and the Netherlands (Amsterdam). The first step was to investigate which MMSI's from the reference frames where present in the location files in a certain area. Statistics Poland did this for a single point for the port of Świnoujście (one latitude/longitude point representing the entry of the port) from the port list provided by Eurostat, for one day. There is a range of points that a ship can pass while entering a port, so using a single point will result in a large underestimation in the number of ship entries. Therefore, external boundaries of the seaport were used (coordinates from the regulation of the Minister of maritime economy), with a spatial resolution of three decimals of latitude and longitude. Figure 2 shows these boundaries for Poland. All ships present in the maritime data were also present in the port in the AIS data.

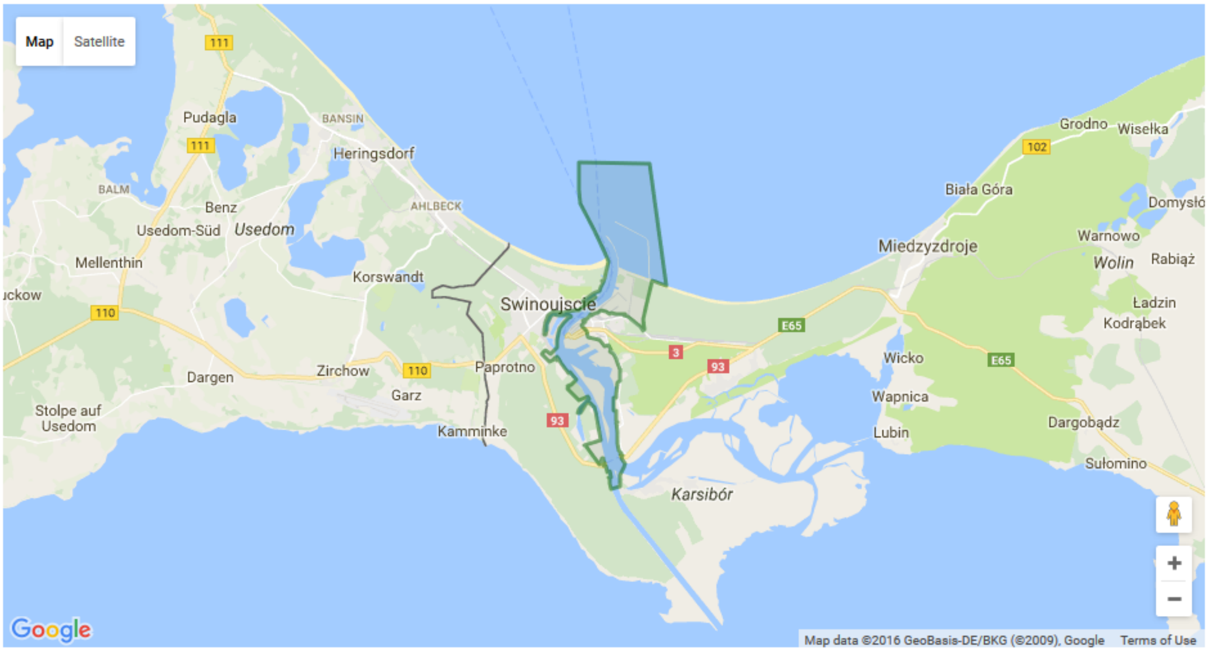

Figure 2: A screenshot of the boundaries of the sea port in Świnoujście

Source: http://www.port.szczecin.pl/granice-portu-morskiego-swinoujscie/

In the investigation performed by Statistics Netherlands for the port of Amsterdam a simple bounding box was used. This bounding box was used to see which MMSI's (out of the correct MMSI-IMO couples) showed up in the bounding box, by comparing locations from the Location files to coordinates of the bounding box. These ships were compared with ships from the port statistics. Firstly, the ships in the maritime statistics data were all present in AIS data. This indicates that AIS data have an excellent coverage of maritime ships in the port. However, using the first rough reference frame of maritime ships resulted in many more visiting ships in the AIS data. Figure 3 shows the location of ships in the Port of Amsterdam area on one typical day in 2015. These ship movements, based on the initial reference frame, are depicted in orange. Closer inspection showed glitches in AIS data elements can cause deformation of all of the elements in a message. As AIS is a radio signal, parts of the messages can get lost or scrambled due to factors such as meteorology or magnetics. Messages are transmitted encrypted. As a result, an error in one transmitted 'byte' can result in an error in one or multiple fields in the decrypted message. Most of the times, these errors are detectable, but sometimes they result in valid fields. For instance, by coincidence the resulting MMSI can be a technically valid, but incorrect MMSI, resulting from an erroneous transmission. These errors can arise for every variable, so this can also result in erroneous latitude and longitude, resulting in faulty locations that are quite far away from the actual location of the ship.

Figure 3. Ship movements in Port of Amsterdam for 1 day

This showed that the reference frame of maritime ships had to be improved. This was done by selecting only correct pairs by filtering out the most frequent MMSI-IMO pair over a 6-month period for each MMSI. This improved reference frame resulted in a smaller number of ship visits. Figure 4 shows all maritime ships based on our reference frame of ships in blue. Again, all ships present in the original maritime statistics were present in our sample. However, we did still find maritime ships that were not actually present in the maritime statistics. Further inspection showed that there were valid reasons for why they should be counted for the port visits. For example, ships that were missing in the maritime statistics had already been counted in the port of Velsen. A particularity of the port of Amsterdam is that ships coming from the North Sea enter the so-called Amsterdam-Rijnkanaal via the port of Velsen. Some of these ships actually already unload here, before going to the port of Amsterdam. Technically, they should be included in the visits for the port of Amsterdam, but due to administrative reasons, they were missing from the original maritime statistics.

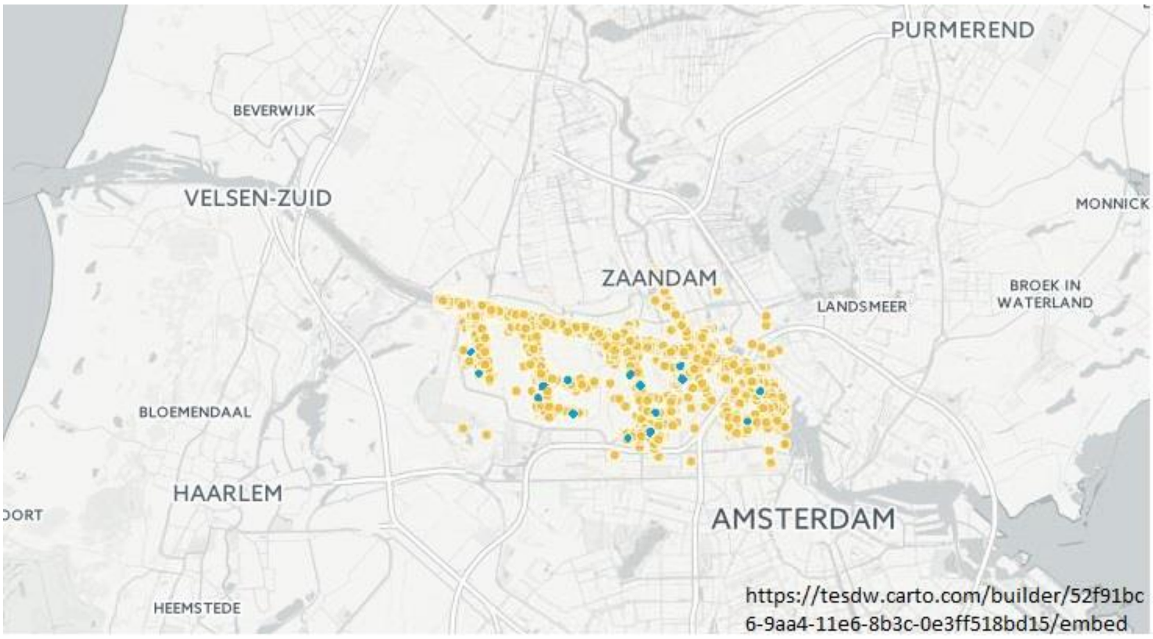

Figure 4. Port visits by maritime ships in Port of Amsterdam for 1 day

Another reason why there were more visiting ships in the AIS data was that this simple algorithm counts ships that were in the port, not their arrival per se. This means that ships that for example arrived the day before but stayed for multiple days, were also counted on those other days. Furthermore, due to the noisy nature of AIS data (characteristic of big data in general and in AIS due to the transmission errors described earlier) some ships appear in the port, although they are not really there. This noise also results in messages of ships in ports appearing outside the port. Looking at the time series of the location (latitude/longitude) of a ship, these outliers can be identified. Thus, an algorithm has to be developed to get rid of these ships that arrived earlier and the noise in the data.

The algorithm developed only counts ships entering in the port. This was done by splitting up the ships' route into locations outside the port (sea) and inside the port (port). The point where the location changed from sea to port was counted as an entry. The latitude, longitude and speed data for each ship were averaged over 1 minute. The resulting number of entering ships still was too high. Furthermore, ships could have up to 120 arrivals on one day. Averaging over 10 minutes did not improve the data much either. For some locations the latitude or longitude was quite far out of range, moving a ship outside of the port for one measurement, and returning for the next. This resulted in ships entering the port for more than ten times. Thus, these large deviant latitude or longitude has to be filtered out. This was obtained by using a median filter over 10 minutes, resulting in a significant improvement of the data. Without the filter, 45 ships seemed to be in the port on one day. From the maritime statistics and time series of the ships, we knew that not all of these ships actually entered the port on that day. With the filter only 15 ships were left. Of the 8 ships entering multiple times (without the filter), only two ships entered twice with the filter. The link mentioned below shows the results.

https://github.com/mputs/WP4/tree/master/Portvisit

This research resulted in an improved reference frame of ships and an algorithm to select ships that arrived in the port. This was done by linking MMSI-IMO couples to IMO-data from the maritime statistics. Subsequently using an algorithm to follow a ship in a certain pre-defined area, the number of ships visiting a port can be derived from AIS data. This enables us to analyse ships entering ports for a certain time period, such as month, week or time of the day.

5. Improving the quality of current statistics and potentially new statistics on the basis of AIS

On November 28th we had a brainstorm session with all participants of WP4. During this session we discussed how current statistics can be improved by using European AIS data. See Table 5.1 for an overview of problems identified. In next paragraphs we describe possible solutions to these problems. For every solution we also describe if we are able to do a Proof of Concept (PoC) of the specific idea within WP4 and in which deliverable the results of the PoC will be described. If there are good reasons for not doing a PoC this is also described (see also table 5.2).

Table 5.1: Current problems in maritime statistics that could be improved/resolved by European AIS data

Nr.	Problem
1.	Information on the next destination of departing ships is incomplete. This can also be used to construct new tables with to and from traffic matrixes
2.	Not all ports are well-specified, they are sometimes misclassified by port authorities
3.	Distance travelled per ship is now based on an inaccurate average distance matrix for ports
4.	Fluvio-maritime transport is incomplete
5.	Investigate relationship between maritime and inland waterway transport
6.	Intra-port travel distances are unknown
7.	Missing Information on travel routes for goods to estimate unit prices for transit trade statistics
8.	Current statistics on fuel consumption and emissions are not accurate enough.
9.	Small ports experience response burden from the survey
10.	Customers need faster information on maritime statistics
11.	Experimental ideas: now-cast economic time series on the basis of AIS

5.1 Improving statistics by determining ship routes in European AIS data

By using AIS data on a European level we can follow all (maritime) ships in European waters through time. Consequently we can infer the journeys from the data. These ship routes gives us the opportunity to **improve data on departing ships regarding the next destination** that is missing from our current maritime statistics. Note that data on goods unloaded in the next port is also missing in our current maritime statistics, but this would not be possible to derive from AIS data directly (draught and type of terminal visited could provide some information on this). Generating information on previous and next port visits from AIS, allows us to **make new tables with to and from traffic matrices**.

Following ships on European waters makes it possible to **improve the quality of specifying ports/terminals of visits**. For example, port authorities in The Netherlands do not always differentiate all ports near the port of Amsterdam, but sometimes aggregate them into the port of Amsterdam.

Based on European AIS data it is also possible to **calculate the distance travelled by the ships**, allowing us to calculate the transportation volume (TKM). This is similar to **improving the quality of the average distance matrix for all ports.** Poland wants to improve their port Distance Calculation Tool by calculating the transportation volume. However, in the end this average distance matrix does not have to be used anymore, as the travel distance for each ship itself is known. This of course is much more accurate than this average distance matrix, rendering the distance matrix obsolete.

Knowing the routes of ships from European AIS data also makes it possible to **gain insight in fluvio-maritime transport** (transport by ships that travel across both sea and inland waters). This data has to be delivered to Eurostat, but is incomplete in our current data. The extent of the missing data is not clear. For the group members, this idea is only relevant for Netherlands. Though France, Germany, Belgium and the United Kingdom can also improve their current statistics on fluvio-maritime transport. All ideas mentioned improve the quality of maritime statistics requested from Eurostat. That is the reason we would like to execute a PoC for these ideas. The results of these PoC's will be described in deliverable 3: the report about traffic analyses.

More generally, AIS can be used to gain insight in the **relationship between maritime and inland waterway transport** as direct links can be seen. That is, which goods are shipped from maritime ships to inland waterway ships and the other way around?

Calculating intra-port travel distances based on AIS data is under investigation at Statistics Netherlands. The ports of Amsterdam and Rotterdam have showed interest in this information. Calculating intra-port travel distances is a new statistical product that could be generated by using AIS data. Therefore, the results of this experiment will be described in SGA-2 deliverable 4.8 in which potentially new statistical output based on European AIS data will be described.

Statistics Netherlands publishes statistics on transit trade. This statistic is based on an integration of transport and international trade data and provides information on quantity and value of transported goods for different modalities. It also provides important information on re-export and transit trade, that in return can also be used for an improvement of transport and international trade statistics. **Characterizing maritime travel routes for specific types of ships** will help improve this integration of transport and international trade data. Currently, the Netherlands is the only country producing this statistic. Therefore we do not plan to investigate this specifically during this WP4.

All in all, this information on ship routes can contribute to a great number of improvements of Eurostat data. However, at this moment we only have European data, so journey information would be restricted to European waters only. Because ships routes are not restricted to European waters, worldwide coverage is crucial for knowing for example the total amount of TKM of a ship or for knowing the next destination of goods based on AIS data.

5.2 Improve emission statistics

Deriving ship journeys from AIS data can also be used to improve **existing statistics on fuel consumption and emissions**. By combining ship routes with a model to estimate the emission of vessels (which depends on travel distance, speed, draught, weather conditions and characteristics of the vessel itself), emissions of e.g. CO_2 and NO_x can be estimated per ship and per national territory. For example in Norway it would be possible to improve these statistics related to domestic (between two Norwegian harbors) traffic. Estimating emissions is planned for SGA-2 deliverable 4.6.

5.3 Reduce the response burden

Smaller ports have to fill out questionnaires consisting of variables that are available in AIS data. Thus, for some statistical offices it is possible to **reduce the response burden for smaller ports**, by asking fewer questions to the ports and by prefilling estimates on the relevant questionnaires. For example, some small ports in Denmark have to fill out questionnaires that could also be based on AIS data. These ports do not have to be approached anymore. Possibilities of reducing the response burden will be part of the cost/benefit analyses of using AIS data for official statistics (deliverable 4.9 of SGA-2).

5.4 Timeliness of publications on selected maritime statistics

If AIS data are going to be a used as a basis for official statistics it is possible to **improve the timeliness of publications of some maritime statistics** (particularly the F2-table). The speed will depend on the arrangements for getting AIS data on a regular basis. Investigating the possibilities and the needs of improving the publishing speed will also be part of deliverable 4.9 of SGA-2.

5.5 Experimental issues

AIS data has the potential to be used for new statistics. WP4 will go into this in deliverable 4.8 of SGA-2 (planned for April 2018). An idea is to generate time series of activity for ships in ports and then check for correlations with economic time series, like trade volume. In that way, AIS data could be used to **now-cast some economic time series** if we receive real-time AIS data in the future. If there is some time left we will do this experiment for example with national AIS data from Denmark, because they get historical AIS data, enabling time series analysis.

Table 5.2: summary of ideas for improving current statistics by using European AIS data

Field	**Idea**	**Addressing problems**	**Reason(s) for (not) execute PoC**	**Results PoC in deliverable**
A.	Determining ship routes	1,2,3,4,5,6,7	Problem 1, 2, 3 and 4 are relevant for (nearly) all countries and improve required statistics by Eurostat. That is why we will execute a PoC on this. Although important for quality overall, problems 5 and 7 do not directly apply to required Eurostat statistics. That is why we will not execute a PoC on this in WP4. Problem 6 also seems relevant only for the Netherlands, but, Statistics Netherlands is currently running a pilot on Intra-port travel distances for another project, that's why we decided to describe the results also in WP4.	4.3(July 2017)
B.	Improve existing statistics on fuel consumption and emissions.	7	Part of SGA-2	4.6 (January 2018)
C.	Reduce response burden for some ports	8	Part of cost-benefit analyses in SGA-2	4.9 (May 2018)
D.	Accelerate publishing speed for some maritime statistics	9	Investigating the possibilities and the needs of improving the publishing speed is part of SGA-2	4.9 (May 2018)
E.	Experimental: Now-cast economic time series	10: Only if there is time left	Interesting experiment: would need larger sample of AIS data	*4.9 (May 2018)*

6. Conclusion

First results show that AIS data can be used as a backbone for maritime statistics. We have developed a method to build a reference frame of maritime ships. The experiments done at Statistics Poland and Statistics Netherlands show that all ships in the maritime data were also present in the port in the AIS data. In the AIS data for the port of Amsterdam we even found more ships in the port than in the maritime statistics data. Further investigation showed that there were valid reasons for the higher amount of ships in this port in the AIS data. From this, we can conclude that the number of port visits is more accurate in the AIS data than maritime statistics. However, AIS data do not contain the level of detail needed for the type and gross tonnage of the ships to be able to generate port visit statistics. Therefore, in order to get more detailed ship information, AIS data has to be linked to existing ship dictionaries, such as Lloyd's register, open data and web scraping methods.

European AIS data can improve current statistics. By using European AIS data it is possible to determine ship routes in European waters. This can improve the data on departing ships regarding the next European destination that is missing from our current maritime statistics. AIS data can also be used to construct new tables with to and from traffic matrixes by generating information on previous and next port visits. Determining ship routes also improves the quality of specifying ports/terminals of visits and European AIS can improve the average distance matrix for ports. Or even replacing the average distance matrix by actual distance between ports. European AIS data can also help by gaining insight in fluvio-maritime transport.

More generally, AIS can be used to gain insight in the relationship between maritime and inland waterway transport. Which goods are shipped from maritime ships to inland waterway ships and the other way around?

Also, intra-port visits can be based on AIS data and AIS data can improve transit trade statistics bycharacterizing maritime travel routes for specific types of ships.

All in all, this information on ship routes can contribute to a great number of improvements of Eurostat data. However, at this moment we only have European data, so journey information would be restricted to European waters only. Because ships routes are not restricted to European waters, worldwide coverage is crucial for knowing for example the total amount of TKM of a ship or for knowing the next destination of goods based on AIS data.

AIS data can also improve statistics on emissions and fuel consumptions. In SGA-2 we aim to investigate the use of AIS data to calculate emissions. During the remaining of WP4 we will execute a couple of PoC's to investigate to what extend European AIS data can solve current problems on:

1. incomplete information on the next destination of departing ships in maritime statistics;
2. specifying ports/terminals of visits;
3. determining TKM on an inaccurate average distance matrix between ports;
4. incomplete information on fluvio-maritime transport in maritime statistics;
5. unknown intra-port distances;
6. inaccurate statistics on fuel consumption and emissions.

The results on PoC 1,2,3, 4 and 5 will be described in deliverable 4.3 (July 2017). The results on improving existing statistics on fuel consumption and emissions by using AIS data (PoC 6) will be described in deliverable 4.6 (January 2018)

European AIS data also could reduce the experienced response burden by small ports and it could help to improve timeliness of publication for some statistics. We would like to set up a questionnaire for all partners of the ESSnet Big Data programme to gain more insight in how other countries experience response burden and if there is a wish for speeding up publication of maritime statistics. Furthermore, other members will also be asked if they have any other ideas to improve current and possibly new statistics by using European AIS data. The results of the questionnaire will be used as input for our final report (deliverable 4.9 May 2018).

7. References

1. Dave Winkler, AIS Data Quality and the Authoritative Vessel Identification Service (AVIS), 10 January 2012: *http://www.navcen.uscg.gov/pdf/AIS/FAQ/16/Winkler@GMDSS_TF_(2012-01-11)_AIS_Data_Quality.pdf*
2. Clément Iphar et al, Data Quality Assessment for Maritime Situation Awareness, 5 February 2016: *https://hal-mines-paristech.archives-ouvertes.fr/hal-01269684/file/Iphar_ISSDQ.pdf*
3. Rino Bosnjak et al, Automatic Identification System in Maritime Traffic and Error Analysis, February 2012: *http://www.toms.com.hr/archive/vol1/no2/toms_vol1no2_doi002.pdf*
4. Royal Belgian Institute of Marine Engineers, Ensuring the reliability of AIS data, January/February 2015: *http://www.gallois.be/ggmagazine_2015/gg_03_03_2015_106.pdf*
5. N. Bailey et al, Training and Technology Onboard Ship: how seafarers learned to use the shipboard Automatic Identification System (AIS), July 2008: *http://www.sirc.cf.ac.uk/uploads/publications/Training%20&%20Technology%20AIS.pdf*
6. Adri Fluit, AIS information quality report of static AIS messages: "AIS Information Quality Report" Region: HELCOM, 4 November 2011: *http://efficiensea.org/files/mainoutputs/wp4/efficiensea_wp4_13.pdf*
7. Xilei Lu, Study on human elements in the application of Automatic Identification System (AIS), 2006: *http://commons.wmu.se/cgi/viewcontent.cgi?article=1109&context=all_dissertations*

Annex 1: SGA-1 of WP4 in more detail

Work package number	4	Start date: End date:				1.2.2016 31.7.2017		
Title	**AIS Data**							
Partner/co-beneficiary	**NL** **120**	DK 60	EL 40	NO 67	PL 40			

Aim of this work package is to investigate whether real-time measurement data of ship positions (measured by the so-called AIS-system) can be used 1) to improve the quality and internal comparability of existing statistics and 2) for new statistical products relevant for the ESS. Improvement of quality and internal comparability can be obtained e.g. by developing a reference frame of ships and their travels in European waters and then linking this reference frame, by ship number, to register-based data about marine transport from port authorities. These linked data can then be used for emission calculations. New products can be developed for e.g. traffic analyses. The added value of running a pilot with AIS-data at European level is that the source data are generic word wide and data can be obtained at European level. Challenges ahead with this dataset are: obtaining the data at European level, processing and collecting the data in such way that they can be used for multiple purposes, and visualising the results. A part of this work package is also to look into AIS analyses done by others and to investigate the possibility of obtaining already processed data as input for creating comparable official statistics. Especially it is important to make contact with other public authorities. This work package may require data acquisition in collaboration with Eurostat.

Methodological, quality and technical results of the work package, including intermediate findings, will be used as inputs for the envisaged WP 8 of SGA-2, in case SGA-2 will be realised. When carrying out the tasks listed below, care will be taken that these results will be stored for later use, by using the facilities described at WP 9.

Task 1 – Data access.

AIS-data are available for national territories and the entire European territory. For example, AIS-data in the Netherlands are provided by RijksWaterStaat (a government agency which is part of the Ministry of Infrastructure and the Environment). It is expected that similar agencies in other countries have the AIS-data for their national territories. At the European level, a dataset of AIS data is available at the European Maritime Safety Agency (EMSA). The advantage of using one AIS-dataset for the entire European territory is a) a better comparison of international traffic between the countries and b) more synergy as all participating countries work on the same dataset. A disadvantage is that these data are stored by private companies and handling fees have to be paid. Aim of this task is to decide how European data could be used for this project, to investigate the possibilities of acquiring data from EMSA (to be coordinated with Eurostat) and, if European data are too costly or too hard to obtain, how national datasets can be obtained.

This task will involve:

- Exploration of the possibilities to collect the data at a European (or worldwide) level.

Participants: NL, DK, EL, NO, PL

Task 2 – Data handling

Aim of this task is to process and store the data in such a way that they can be used for consistent multiple outputs, like

- linking AIS-data with data from port authorities
- traffic analyses
- Inference of journeys from AIS data.

Key elements of this task are:

- which programming language and environment should be used for transformation
- where will the data be processed (in each NSI, by NSIs at European level, by data holders)
- how can we create an environment which is easily accessible for all partners

Participants: NL, NO, PL, and possibly DK (if national sources are used)

Task 3 – Methodology and Techniques

Develop traffic statistics: Linking with data from port authorities.

AIS-data may be linked to data from port authorities. Added value of linking AIS-data to data from port authorities is that the same reference population (= ship number) is used in all harbours. As the journeys and harbour visits of ships can be derived from AIS this linking provides the ESS information about the origin/destination of the cargo, too.

Aims of this task are:

- to build a reference frame of ships in European water (based on AIS-data)
- to find out how data from port authorities can be linked to AIS-data
- to check whether information improves the quality of current statistical outputs and provides more information about the origin/destination of the cargo.

Participants: NL, NO, PL

Traffic analyses

The number of ships during a certain time interval at certain coordinates (like inland waterways or at certain points at sea) can be calculated by AIS-data. This possibility will be explored because this information could be interesting for traffic analyses and economic analyses.

Aims of this task are:

- calculate the number of ships at certain coordinates

- visualise the results to analyse variations in time

Participants: NL, NO, PL, EL

Estimate emissions (envisaged under SGA-2)

This task involves following individual vessels through time. Consequently we can infer the journeys from the data. Combined with a model to estimate the emission of vessels (which depends on travel distance, speed, draught, weather conditions and characteristics of the vessel itself), emissions of e.g. CO_2 and NO_x can be estimated per ship and per national territory.

An advantage of doing these analyses on a European scale– instead of the national level – is that more precise estimates for emissions at national territories can be made.

Aim of this task is 1) to infer journeys from AIS-data, 2) visualise the results, 3) combine these journeys with a model to calculate emissions and 4) estimate the impact of carrying out these calculations at the European level on the quality of emissions calculations.

Task 4 – Future perspectives (envisaged under SGA-2)

Aim of this task is to summarise the project results and perform a qualitative cost-benefit analysis of using AIS-data for official statistics. These analyses should include aspects like sustainability of the data source, possibilities of improving international comparability, possibilities of data sharing (at micro- or aggregated level), quality improvement of current statistics and a sketch of a possible statistical process and needed infrastructure.

Deliverables (SGA-1 only):

4.1	Report on creating a database with AIS-data for official statistics: possibilities and pitfalls	month 6
4.2	Report about deriving harbour visits and linking data from port authorities with AIS-data	month 12
4.3	Report about sea traffic analyses using AIS-data	month 18

Milestones (SGA-1 only):

4.4	Progress and technical report of first internal WP-meeting	month 4
4.5	Progress and technical report of second internal WP-meeting	month 9

Annex 2: SGA-2 of WP4 in more detail

Work package number	4	Start date: End date:		1.8.2017 31.5.2018	
Title	**AIS Data**				
Partner/co-beneficiary **(person days)**	**NL** **88**	DK 53	EL 66	NO 21	PL 35

Description of the work package

Aim of this work package is to investigate whether real-time measurement data of ship positions (measured by the so-called AIS-system) can be used 1) to improve the quality and internal comparability of existing statistics and 2) for new statistical products relevant for the ESS. Improvement of quality and internal comparability can be obtained e.g. by developing a reference frame of ships and their travels in European waters and then linking this reference frame, by ship number, to register-based data about marine transport from port authorities. These linked data can then be used for emission calculations. New products can be developed for e.g. traffic analyses. The added value of running a pilot with AIS-data at European level is that the source data are generic worldwide and data can be obtained at European level.

Challenges ahead with this dataset are: obtaining the data at European level, processing and collecting the data in such way that they can be used for multiple purposes, and visualising the results. A part of this work package is also to look into AIS analyses done by others and to investigate the possibility of obtaining already processed data as input for creating comparable official statistics. Especially it is important to make contact with other public authorities. This work package may require data acquisition in collaboration with Eurostat.

Methodological, quality and technical results of the work package, including intermediate findings, will be used as inputs for WP 8 of SGA-2. When carrying out the tasks listed below, care will be taken that these results will be stored for later use, by using the facilities described at WP 9.

SGA-2 of WP 4 will deliver the products below:

Task 3 – Methodology and Techniques

Estimate emissions

This task involves following individual vessels through time. Consequently we can infer the journeys from the data. Combined with a model to estimate the emission of vessels (which depends on travel distance, speed, draught, weather conditions and characteristics of the vessel itself), emissions of e.g. CO_2 and NO_x can be estimated per ship and per national territory.

Estimation of emissions based on only AIS data is impossible. At least, we need to know what the emission should be given, a.o., draught, speed and weather conditions. There are several sources that can be used to get information about the emission of vessels. For example Lloyds Register of Shipping.

Also, all ships have to register their emissions, so there is data available at the freight ship companies. A dataset for the same period as the AIS dataset could be used to model the emissions based on the variables mentioned above. We assume getting these needed data for free for this pilot.

Furthermore, satellite data is available for emissions (see, for instance http://www.globemission.eu), which shows very clearly the maritime routes. This satellite data could be used as a more direct source for measuring emissions and maybe for testing our model. Another possibility to test the model is to compare the estimated emissions by the model with the real fuel purchase or information coming from vessel's oil record book.

An advantage of doing these analyses on a European scale– instead of the national level – is that more precise estimates for emissions at national territories can be made.

Aim of this task is to develop and test a methodology for estimating vessel emissions based on AIS data by:

1) inferring journeys from AIS-data
2) visualising the results
3) investigating methodology for calculating emissions
4) combining these journeys with a model to calculate emissions
5) Testing the model and
6) estimating the impact of carrying out these calculations at the European level on the quality of emissions calculations.

Participants: Norway, Netherlands , Denmark, Greece, Poland

Task 4 – Access to and analysing AIS data from EMSA

AIS-data are available for national territories and the entire European territory. In this work package we use European AIS data from Dirkzwager. But for the future we would like to get free European AIS data. That is why we try to get AIS data at the European level from the European Maritime Safety Agency (EMSA).

In SGA-2 we focus on getting the European AIS data from EMSA and compare these two sources (AIS data from EMSA and Dirkzwager) on their coverage and quality. We will assess the quality of both sources by applying the quality framework. We also describe the strengths and weaknesses of both sources compared with existing maritime data.

This work package may require data acquisition in collaboration with Eurostat, because we would like Eurostat to apply for the data at EMSA's. Eurostat already received AIS data from EMSA for other projects and Eurostat can apply EMSA data for all the partner countries once. This will be better than applying for each country separated. If we do not get access to the EMSA data we cannot process this task and we cannot deliver deliverable 4.7, but it has not any consequences for the other deliverables in this WP. Not getting AIS data from EMSA means that we cannot get AIS data on the European level for free.

Participants: Netherlands , Denmark, Greece, Poland

Task 5 – New statistical output

Aim of this task is to explore possibilities of new statistical products (for example interport statistics) by using AIS data. This task also includes analysing and elaborating scenarios for production of European and national statistics based on one single European data source of AIS data.

Participants: Norway, Netherlands, Denmark, Greece, Poland

Task 6 – Future perspectives

Aim of this task is to produce a consolidated report summarising the contents and the outcomes of WP 4. This report also includes a cost-benefit analysis of using AIS-data for official statistics. The report should also include aspects like sustainability of the data source, possibilities of improving international comparability, possibilities of data sharing (at micro- or aggregated level), possibilities of meeting the needs of both European and national statistics by one European AIS database, quality improvement of current statistics and a sketch of a possible statistical process and needed infrastructure**,** including technical skills required to generate statistical outputs from the source data.

Participants: Norway, Netherlands, Denmark, Greece, Poland

Deliverables (SGA-2 only):

4.6	Report on estimating emissions. This report will describe the investigated methodology for calculating emissions and the reason why we choose for a certain methodology. The report also describes the created model itself (inclusive other needed data sources) and the results of testing the model. Finally this report also describes the impact of carrying out these calculations at the European level on the quality of emissions calculations.	month 13
4.7	Report about the results of comparing the quality and coverage of the European AIS data from Dirkzwager and EMSA (by applying the quality framework). Also the strengths and weaknesses of both sources will be compared with existing maritime data.	Month 15
4.8	Report about possible new statistical output based on European AIS data. The report also describes analysed and elaborated scenarios for production of European and national statistics based on one single European data source of AIS data.	month 16
4.9	Consolidated report on project results including a cost-benefit analysis of using AIS-data for official statistics.	month 17

Milestones (SGA-2 only):

4.10	Progress and technical report of first internal WP-meeting	Month 9
4.11	Progress and technical report of the second internal WP-meeting	Month 14

Annex 3: description of variables in location- and identify messages

Variables in the location messages:

- MMSI
- longitude
- latitude
- accuracy
- speed
- course
- rotation
- status
- timestamp

Variables in the identify messages:

- MMSI
- IMO
- name
- callsign
- destination
- draught
- dim_a
- dim_b
- dim_c
- dim_d
- fix_type
- type and cargo
- timestamp

Annex 4: description of systematic errors in AIS data (Fluit, 2015)

Parameter	Known errors	Description
IMO	IMO missing IMO = 0 (default value)	IMO number must have a length of 7 digits. IMO = 0 (default) means, that MMSI number wasn't entered on the transponder. Main reason of IMO error is wrongly inserted on the transponder.
MMSI	MMSI = 0 (default value) MMSI = 1193046 (default/resetting value for AIS device) MMSI = 123456789 MMSI = 000000000 MMSI = 111111111 MMSI = 222222222 MMSI = 333333333 MMSI = 444444444 MMSI = 555555555 MMSI = 666666666 MMSI = 777777777 MMSI = 888888888 MMSI = 999999999	MMSI number must have a length of 9 digits. MMSI = 0 (default) means, that MMSI number wasn't entered on the transponder. The same MMSI number 1193046 is transmitted by more than one different vessel. It is possible that this number was adapted to a specific model of AIS device. Main reason of MMSI error is wrongly inserted on the transponder.
Ship name	Ship name is empty or filled with blank characters: @) "@@@@@@" = not available = default	All static parameters are manually entered in the AIS device so therefore are a common cause of errors.
Call sign	Call sign is missing @@@@@@@ = not available = default	Call signal of a vessel. Call sign of a vessel is a worldwide unique identifier for ships and boats assigned by their national licensing authorities Each call sign begins with the Call Sign alphanumeric prefix that indicates nationality. The prefix is usually followed by 2 or 3 alphanumeric characters.
Type of ship	0 = not available or no ship = default	
dim_a and dim_b dimensions to bow and stern	when dim_a = 0 and dim_b=0 Ship dimensions = 0 if not available. The length of ship is a negative value. The length of ship is above 460m	Dimensions to bow and stern. Dim_a and dim_b represent the location of the GPS antenna on board the ship. The length of the ship is sum of dim_a + dim_b and should correspond with value in the Lloyds register. Situation when one of the parameters is 0 may be correct. In AIS data, there were ships which had length above 460m which is more than the longest ship ever built.
dim_c and dim_d	when dim_c = 0 and dim_d=0 Ship dimensions = 0 if not available. The width is a negative value.	Dimensions to port and starboard The width of the ship is sum of dim_c + dim_d and should correspond with value in the Lloyds register. Situation when one of the parameters is 0 may be correct.

Destination	@@@@@@@@ = not available	Voyage data must be updated after each port visit. Crew often does not change value of destination.
Draught	0 = not available = default;	
Latitude and longitude coordinates	Value out of range	In AIS data, there is a high number of ships where value is out of range. Range of value for latitude and longitude coordinates: • **Latitude**: max/min +90 to -90 • **Longitude**: max/min +180 to -180