Document title	Draft HELCOM guidance regarding best practices in data handling and definitions of certain AIS data products
Code	4-1-Rev.1
Category	DEC
Agenda Item	4 – Access to and use of HELCOM AIS information
Submission date	27.05.2016
Submitted by	Secretariat
Reference	

This is a revised version of document 4-1:

Annex 2 has been added and a general proof-reading has been made.

Background

The HELCOM Secretariat has until now produced traffic density maps and descriptive statistics from AIS data based on short period and small geographical areas. This has been due mainly to lack of human and material resources.

In order to server better the HELCOM community , the HELCOM Secretariat would like to generate and deliver high resolution data products on ship traffic and its environmental pressures/risks based on AIS.

The Secretariat participates in a Maritime Spatial Planning project that has been the starting point to design a process to handle data and produce such products. This document explains the detail on how we did it. We explain the main steps from pre-processing AIS data to making shipping traffic density maps.

Action requested

The Meeting is invited to consider the draft HELCOM guidance regarding best practices in data handling and definitions of certain AIS data products, and if feasible, put it forward as the substance of a new HELCOM Recommendation.

How to handle AIS data and make shipping traffic density maps

Contents

From AIS raw data to shipping traffic density maps

HELCOM Automatic Identification System (AIS) data is currently used for several purposes in the Baltic Sea like for example improving navigation safety, analyzing emissions from ships or mapping underwater noise. The Danish Maritime Authority (DMA) has been storing since 2005 a compilation of the data of each country in the HELCOM AIS network, available to all Baltic Sea countries and other organizations.

The HELCOM Secretariat has not been able to make use of this data due to lack of resources and many projects in the HELCOM community have not taken advantage of this valuable information. Therefore, the HELCOM Secretariat started a project last year called Baltic SCOPE to provide shipping traffic density maps (a map that highlights the intensity of the traffic at a given time) to help in the elaboration of Maritime Spatial Planning.

Learning to handle AIS data has not been an easy task but the HELCOM Secretariat managed to create a process to make available processed AIS data and shipping traffic density maps in a relatively short time. This document covers all steps, from how to process raw AIS data to how to create density maps.

In the first section of the document we will explain the data handling practices used at the HELCOM Secretariat to generate a harmonized dataset of AIS data in the Baltic Sea. In the second part we will explain how to generate shipping traffic density maps. The document does not contain information regarding the quality of the AIS data (e.g. missing signals, number of duplicated signals, etc.).

AIS data handling practices

The aim of this section is to share the practices to pre-process AIS data to create a harmonized dataset. This dataset can be used for generating statistics, maps or other relevant products about shipping activities in the Baltic Sea. After several trials, it was decided to produce monthly files of AIS data for the years 2005 to 2014 for further analysis.

Following interviews with AIS data users (Annex I), we decided to work on a dedicated server available by remote access with the following specifications: Intel Xeon E5-2630 0 @ 2,30GHz 10 cores with 48 GB RAM. This server allows several persons working at the same time.

Description and decoding of the data

The AIS data was delivered by the Danish Maritime Authority (DMA) in two batches: one batch with decoded data in yearly files and a second one with raw data sorted daily files. The decoded and the raw data contain all the AIS messages (e.g. position reports from ships, base stations reports, etc.) received from the base stations that are part of the HELCOM AIS network.

The difference between decoded and raw data is:

- Decoded data: The decoded data in CSV files contain several parameters (columns) such as the date and time when the signal was issued, the identification of the AIS message, the identification number of the AIS transmitter, etc.
- Raw data: The raw data (.txt) contain NMEA sentences always preceded by an encapsulated tag. These tags (beginning with the characters “$PGHP”) contains the information related to the moment (date and time) when the signal was issued. The

NMEA sentences contain the rest of information: the identification of the message issued, the identification of the AIS antenna, etc.).

While the decoded data was already sent in CSV files, the raw data from 2007 and 2008 went through a process of decoding to convert them to human-readable CSV files.

The figure 1 explains this process: the daily files or raw AIS data were merged into monthly files (a) and then decoded into CSV files (b).

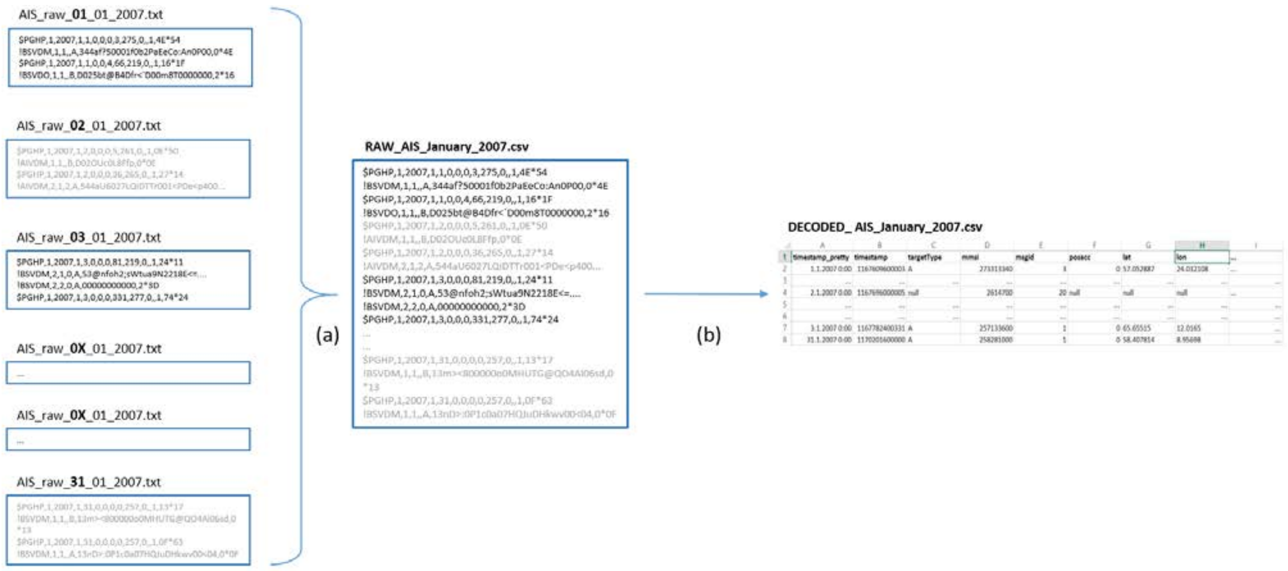


Fig. 1: Preparation of the decoded AIS data (example for January 2007)

The AIS decoder used on the step (b) is a free software available online developed in 2015 by DMA. The application is decoding each NMEA sentences with its encapsulated tag and is generating CSV files.

For both 2007 and 2008, the 12 decoded monthly files were finally merged into one final file: a yearly AIS decoded file that have the same format as the other years. For the years 2005 to 2014, the yearly files are between 45,3 GB (year 2005) and 202,9 GB (year 2013). These 10 yearly files of AIS records can now be pre-processed.

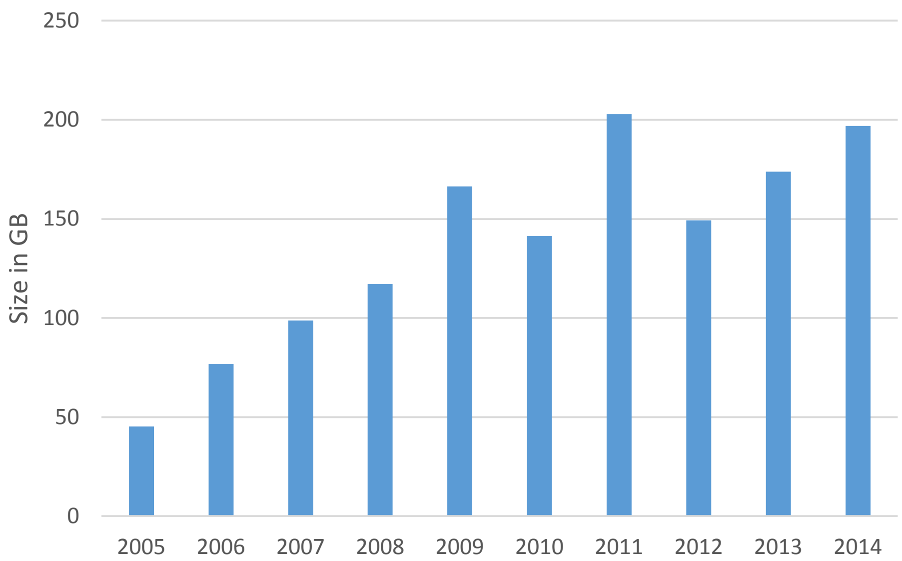


Fig. 2: Size of the AIS yearly files

The next table gives the detailed information for each years. Notice that the HELCOM AIS dataset begins in April 2005, when the network was AIS network was put in place.

Year	Format of AIS data received from DMA
2005 2006	Yearly file (.CSV), decoded
2007 2008	Daily AIS raw strings (.txt)
2009 2010 2011 2012 2013 2014	Yearly file (.CSV), decoded

Pre-processing of decoded data

The main goal of pre-processing the AIS data is to make available the relevant information to produce maps or generate statistics on shipping activities in the Baltic Sea. The inputs of the pre-processing steps are yearly files (CSV) of AIS data containing all messages. The outputs are monthly files (CSV) that contain the positions of the ships in the Baltic Sea (position reports). The pre-process was done using R language with the RStudio interface (0.99.893), the same script was applied to the yearly files one by one. Two main steps can be identified:

- **Step 1**: cleaning and selection of relevant AIS data:
  The yearly file is divided in smaller files of 1 000 000 rows to avoid running out of memory. For each divisions, a process is going through each AIS signals to select the relevant data and to remove erroneous signals.
- **Step 2**: Sorting by month the selected AIS data:
  After the step 1, each of the divisions are sorted by months and merged into monthly files.

Step 1: Cleaning and selection of relevant AIS data

This step aims to remove erroneous signals and to select the relevant data linked with shipping activities. The following actions are applied to the divisions:

- Removal of the signals that are not from the year wanted
- Removal of the duplicated signals
- Selection of AIS messages relevant for assessing shipping activities (1, 2, 3, 18 and 19).
- Removal of wrong MMSI signals. A list that can be updated (i.e. less or more than 9 digits or equal to 000000000,111111111,222222222,333333333,444444444,555555555, 666666666, 777777777, 888888888, 999999999,123456789,0,12345, 1193046).
- Correction of wrong IMO numbers: for each signals with an erroneous IMO number (not seven digits), it is replaced by “NA”.

- o Add the Maritime Identification Digits (MID) and the flag of she ships for each signals. The MID is the three initial digits of the MMSI. This action is also removing MMSI numbers that do not have a MID (erroneous MMSI).
- o Removal of special characters in all the division.
- o Addition of two columns: one for the week number and one for the month.
- o Selection of the signals within the Baltic Sea and internal waters (following the definition of the Baltic Sea in the Article 1 of the Helsinki Convention. A polygon was drawn manually around the Baltic and only the signals within this polygon were kept (Figure 2).

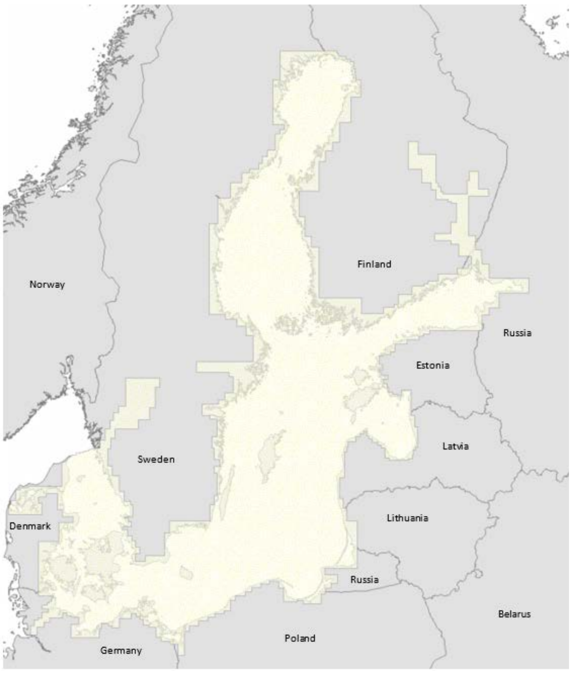


Fig. 3: In yellow the polygon covering the Baltic Sea area to discard the wrong AIS signals that are on land

- o Removal of the signals with erroneous SOG (Speed Over Ground): negative values or more than 80 knots.
- o Removal of the signals with erroneous COG (Course Over Ground): negative values or more than 360°.
- o Selection of relevant parameters to generate relevant data products.

All parameters in table 1 are kept for all of the signals. There is a lot of redundancy but, because of the processing time to create the final files, it was decided to avoid deleting information.

Parameters	Description
timestamp_pretty	time in format dd/mm/yyyy hh:mm:ss
timestamp	Unix time stamp (seconds since 01/01/1970 00:00:00)
msgid	The AIS message the signal was issued
targetType	AIS type A or B
mmsi	MMSI number of the ship
lat	Latitude in decimal format
	Longitude in decimal format
posacc	Position accuracy
SOG	Speed Over Ground in 0.1 knot
COG	Course Over Ground in 0.1°
shipType	Ship type of the vessel
dimBow	The dimension between the AIS transmitter and the bow of the ship in meters
draught	Draught of the ship in 0.1 meter
dimPort	The dimension between the AIS transmitter and the port side (left) of the boat in meters
dimStarboard	The dimension between the AIS transmitter and the starboard side (right) of the boat in meters
dimStern	The dimension between the AIS transmitter and the stern of the ship in meters
month	Month the signal was issued (between 1 and 12)
week	Week number the signal was issued
imo	IMO number of the ship
country	Flag of the ship

Table 1: Parameters in the pre-processed AIS data

Each divisions are saved as a CSV file. For each divisions, a file is also reporting the amount of signals kept after removing the erroneous signals (duplicated signals, wrong MMSI, etc.), these files are called “reports”.

Step 2: Sorting by month the selected AIS data.

This final step will go through each divisions and will write in a new CSV file the signals for a same month. This process will begin with January, then February, March, etc. The column “month” is used to sort the data into the final files. It is possible to sort into weekly files but because the processing time was too long, this solution was settled aside. The reports are also merged into monthly files per each year. The size in GB for each year ranges from about 15 GB to almost 80 GB (fig- 3)

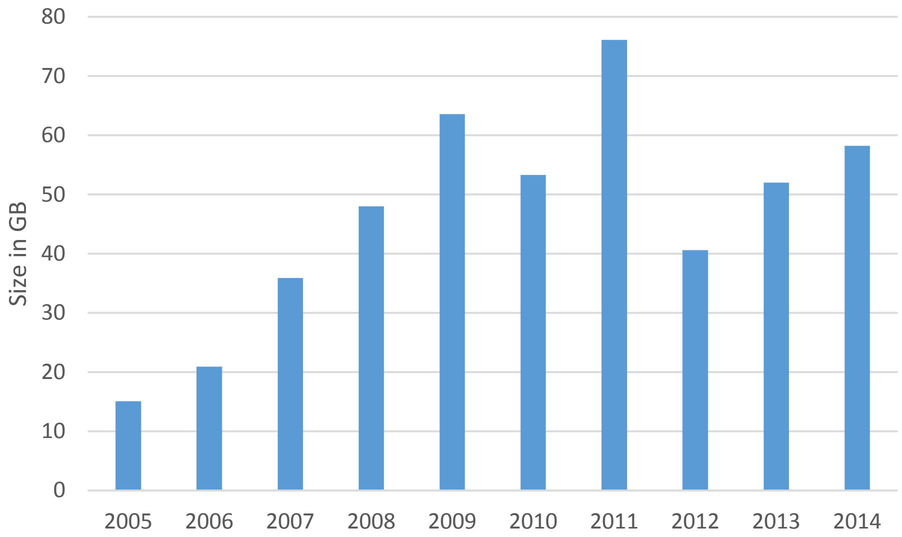


Fig. 4: Size in GB of pre-processed AIS data

The whole process is described in the figure in the next page.

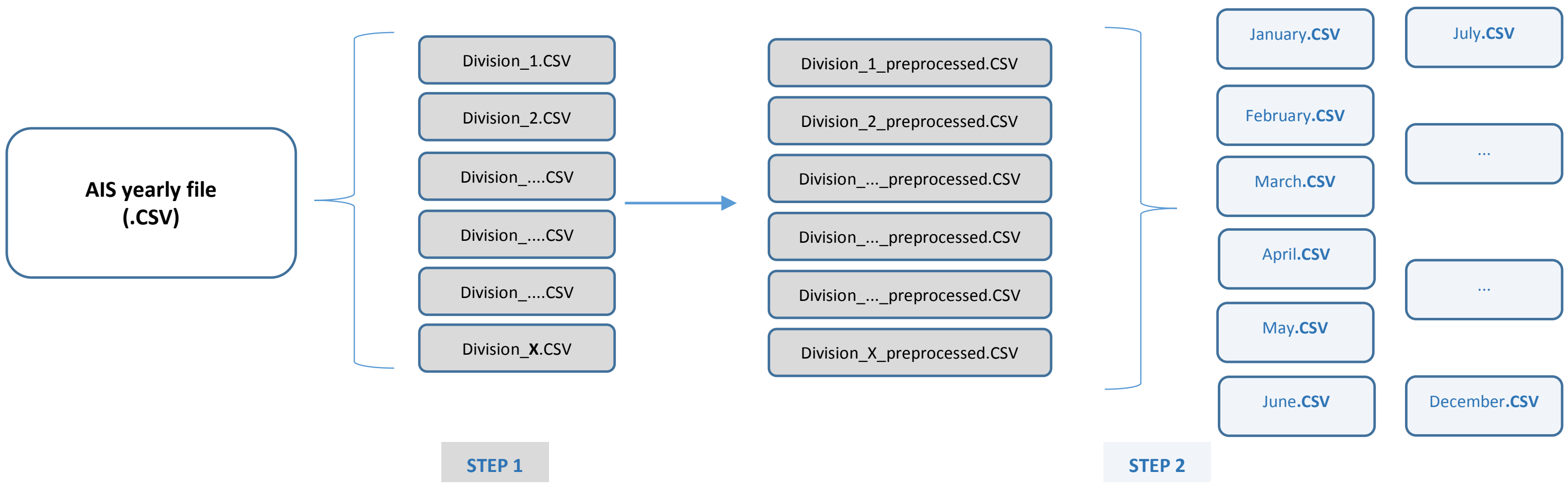


Fig.5: Steps to pre-process yearly files of AIS data

Upgrading the ship information

The ship-related information (i.e. dimensions, ship type) is not obligatory in the AIS data and therefore it is incomplete not reliable for generating relevant AIS products related to IMO-registered ships such as density maps per ship type or by ship dimensions. An upgrade of the ship information was accomplished following four steps:

Step 1: For each year, a ship list was generated. The lists include all ships (IMO and non-IMO registered ships) and contain the following parameters:

- MMSI
- IMO number
- Name of the ship
- Callsign
- Country
- Target type
- Ship type
- DimBow
- Draught
- DimPort
- DimStarboard
- DimStern

Step 2: The list of each years (except 2007 and 2008) were merged to have a unique list of ships that operated in the Baltic Sea during this period. The ship-related information was bought from a provider for 18,358 IMO registered ships. The IMO numbers were used to identify the ships and the following parameters were provided:

- IMO number
- Name
- Ship type
- Gross Tonnage
- Net Tonnage
- Length
- Width
- Draught

Step 3: Each yearly ship lists (from step 1) were edited using the new information from step 2. When the information was available, the ship information from the AIS data (i.e. ship type) was replaced by the new information from the provider (only for IMO registered ships). When the information was not available from the provider, the original data (from AIS) was kept. At the end of this step, a total of 120 ship types were available in the ship list.

Step 4: Finally, two ship type categories were created to use with full potential the 120 available ship types. The gross ship type gives broad information about the ship, the detail ship type gives more precise information for example about the type of cargo or tanker. The table 2 defines the gross and detail categorizations:

Gross ship type categorisation	**Detail ship type categorisation**
Cargo	Bulk cargo General cargo Other cargo
Tanker	Chemical tanker Chemical/Oil tanker Gas tanker Oil tanker Other tanker
Container	Container
Passenger	Cruise Ferry ROPAX
Other	Dredger Other Tug Yatch
Fishing	Fishing
Service	Service
Vehicle carrier	Vehicle carrier

Table 2: Ship type categorisation used to update ship information

These ship types were chosen following the current work about emissions from shipping in the Baltic Sea done by the Finnish Meteorological Institute (cf. Information document 4-4 Emissions from Baltic Sea shipping in 2014 submitted by Finland for the HELCOM MARITIME 15-2015 meeting). For each 120 ship types from the previous step (step 3), a new ship type from the table 2 was assigned. The detail is available in the Annex 2.

It is now possible to generate relevant AIS products (i.e. shipping traffic density maps) per ship type or the ship dimensions for IMO registered ships. Following their IMO numbers, it is possible to filter the AIS monthly file by one or several parameters to generate relevant products.

From data to maps

A density map shows the intensity of the shipping traffic at a given time based on AIS data. We at the HELCOM Secretariat have sporadically produced such maps but the result has always been sub-optimal. The problem is that we have only used data for a very short period (maximum one week) and the method to create them was not good enough for our current purposes.

In this section, we will show the detailed process from text files in CSV format to a map. We also include what kind of software and hardware we used and discuss the problem with the traditional methods.

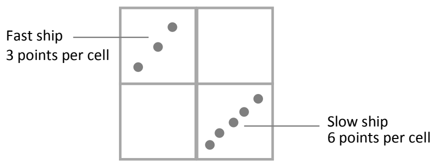
The best way we found to make these maps was to overlay a very detailed grid upon the lines that are the result of converting AIS points to lines. After overlapping lines and the grid, we could count the number lines crossing each cell and produce a reliable density map.

The problem with the point and line density

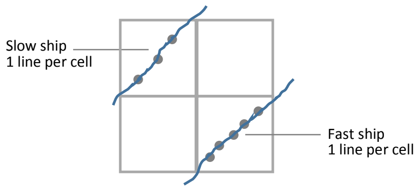
The traditional way of producing density maps has some problems. Usually, these maps are produced very fats with point or line density tools found in GIS software. The result shows density based on some calculations and algorithms but it doesn't really answer the most relevant question:

"How many ships are crossing each cell in a given grid?"

Example of a map the Secretariat has been producing in the past. It is a raster map that shows the intensity of shipping traffic during only one week in 2008.



In the traditional method we counted number of points per cell. In this example we can see the problem with this approach. If the ship goes slowly the density of points is much more than when the ship goes fast.



The optimal method is to count lines, not points, crossing each cell.

As we explain later, the only way we found was to count how many lines were crossing each cells with Spatial Joins and them summing the results with the Dissolve tool. This method is slow and consumes a lot of resources but it works.

Software used

We used ArcGIS 10.2 for Desktop advanced license with Spatial Analyst for creating raster layers and Background Geoprocessing 64-bit for running the scripts.

All layers were stored in File Geodatabases, a native ArcGIS format for storing and managing data in, according to ESRI, a fast and reliable way.

All scripts were written in ArcPython, a Python language package for geoprocessing with ArcGIS. We used PyScripter (an open source Integrated Development Environment (IDE)) for writing the scripts and IDLE (the integrated development environment that comes with

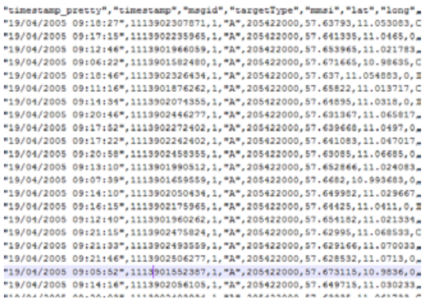
Python) for running them. Our experience running scripts directly from PyScripter was not good because they often crashed and were slower than running them from IDLE.

We found that writing stand-alone scripts was much more efficient that using Model Builder. The learning curve for writing Python scripts is higher but they are easier to customize and much more flexible.

How to make a shipping traffic density map

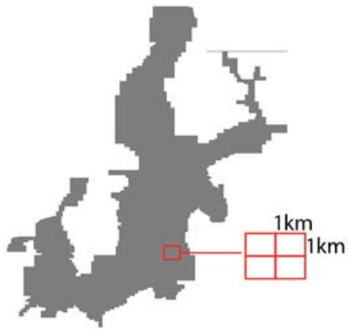
In order to make a shipping traffic density map we need:

- AIS files in CSV format that are the result of a previous cleaning process.
- A 1km x 1km grid based on INSPIRE Geographical grid systems.



```
"timestamp_pretty","timestamp","msgid","targetType","mmsi","lat","long",
"19/04/2005 09:18:27",1113902307871,1,"A",205422000,57.63793,11.053083,C
"19/04/2005 09:17:15",1113902235965,1,"A",205422000,57.641335,11.0465,0,
"19/04/2005 09:12:46",1113901966059,1,"A",205422000,57.653965,11.021783,
"19/04/2005 09:06:22",1113901582480,1,"A",205422000,57.671665,10.98635,C
"19/04/2005 09:18:46",1113902326434,1,"A",205422000,57.637,11.054883,0,
"19/04/2005 09:11:16",1113901876262,1,"A",205422000,57.65822,11.013717,C
"19/04/2005 09:14:34",1113902074355,1,"A",205422000,57.64895,11.0318,0,
"19/04/2005 09:20:46",1113902446277,1,"A",205422000,57.631367,11.065817,
"19/04/2005 09:17:52",1113902272402,1,"A",205422000,57.639668,11.0497,0,
"19/04/2005 09:17:22",1113902242402,1,"A",205422000,57.641083,11.047017,
"19/04/2005 09:20:58",1113902458355,1,"A",205422000,57.63085,11.06685,0,
"19/04/2005 09:13:10",1113901990512,1,"A",205422000,57.652866,11.024083,
"19/04/2005 09:07:39",1113901659559,1,"A",205422000,57.6682,10.993683,0,
"19/04/2005 09:14:10",1113902050434,1,"A",205422000,57.649982,11.029667,
"19/04/2005 09:16:15",1113902175965,1,"A",205422000,57.64425,11.0411,0,
"19/04/2005 09:12:40",1113901960262,1,"A",205422000,57.654182,11.021334,
"19/04/2005 09:21:15",1113902475824,1,"A",205422000,57.62995,11.068533,C
"19/04/2005 09:21:33",1113902493559,1,"A",205422000,57.629166,11.070033,
"19/04/2005 09:21:46",1113902506277,1,"A",205422000,57.628532,11.0713,0,
"19/04/2005 09:05:52",1113901552387,1,"A",205422000,57.673115,10.9836,0,
"19/04/2005 09:14:16",1113902056105,1,"A",205422000,57.649715,11.030233,
```

Exmaple of a table in CSV format with some of the fields.



The 1km grid with the whole Baltic Sea. The detail shows 1km cells

The CSV files

Each yearly file in CSV format was huge—about 50GB. Thus we needed to divide them in monthly files to avoid memory issues. Each monthly file was about 3,8GB on average which made them easier to work with.

The grid

The grid file was downloaded from the European Environment Agency (EEA) and it is based on the recommendation at the 1st European Workshop on Reference Grids in 2003 and later INSPIRE geographical grid systems (http://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2). This standard grid was recommended to facilitate the management and analyses of spatial information for a variety of applications.

EEA offers a grid in shapefile format for each country in three scales: 1km, 10km and 100km. We chose the 1km grid because we wanted as much detail as possible.

Since the grids were divided in countries we first downloaded the Baltic Sea countries and then joined all into a unique grid file.

The resulting merged grid had more than four millions cells. We needed to delete manually the cells on land to save space and make the file easier to manage. The result was a file with about 500.000 cells:

The final grid covers only the sea area.

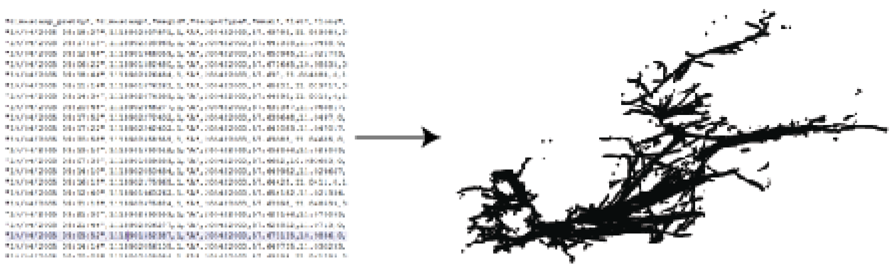
The process

Once we have the 1km grid and the CSV we can start producing the map. The process is divided in three steps

1. Make the points: We produce a point feature class for each monthly file.
2. Make lines: The points are converted to lines with a script called TrackBuilder.
3. Make maps: The lines and the grid are overlapped to count how many lines are crossing each cell.

Make points

In this step we convert the CSV files to ArcGIS point feature classes.



The table in CSV format is converted to a point feature class using a Python script

The monthly CSV files are converted to point feature classes with a script written in ArcPy. ArcGIS has its own function to convert CSV files to points (Make XY Event layer) but this script proved to be much faster.

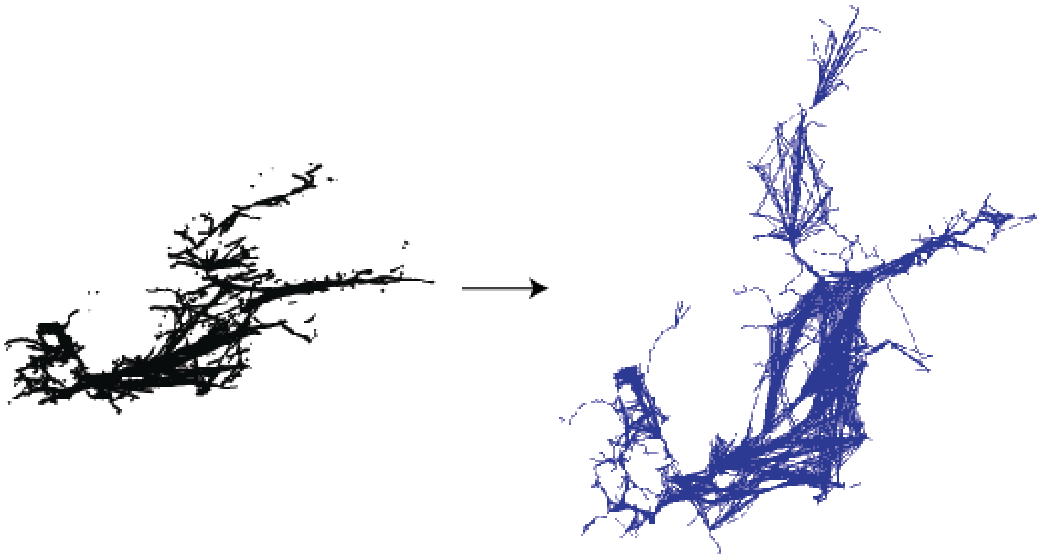
The result of this script are 12 point feature classes, one for each month.

We still need to improve the script by adding the following functionality:

- Ability to read the column headers in the csv. As it is now the headers must be hard-coded.
- Ability to find in which columns are the latitude and longitude which it is also hard-coded.

Make lines

Here we convert the point files into line feature classes.



The point feature classes are converted to lines with TrackBuilder, a Python script from NOAA. As you see in the example the projection is also changed from WGS84 to LAEA (Lambert Azimuthal Equal-Area) which is the one we use oficially in HELCOM.

The points are converted to lines with a python script called TrackBuilder made by Digital Coast, managed by NOAA's Office for Coastal Management. TrackBuilder allows to generate track lines according to date, time and identifier. It also allows users to set a distance and time filter to compensate for gaps in the data.

The advantage of using this script instead of the ArcGIS function Point to Line was that there were less errors and that each line is a trip—the script calculates the time of departure and arrival to make the line.

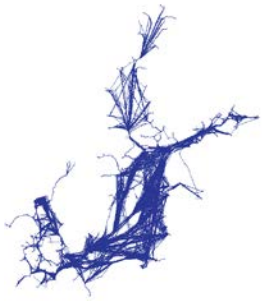
Make map

We finally make a density map overlapping the 1km grid and the lines.

To make the map we need three layers:

Grid division
Five big polygons covering the whole Baltic Sea. A Spatial Join between lines and the 1km grid is done in each polygon. All polygons are merged afterwards. The reason to do this is to avoid memory issues.

1km grid
An INSPIRE compliant grid downloaded from EEA. All land areas have been previously deleted

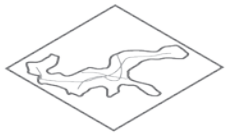


Lines
There is a line feature class for each month.

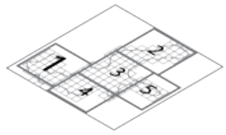
Steps for creating a map

There are six steps for creating the maps:

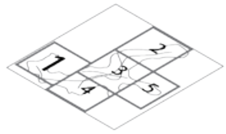
Monthly lines file: We start with a monthly line file



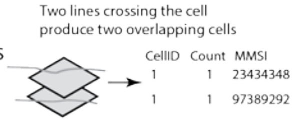
1. We **select cells** inside Grid Division: in order to avoid running out of memory we divide the Baltic Sea area in five big squares, which we call Grid Division. First, we need to select the 1km cells that are inside the first one.



2. Now we **select lines** inside that first Grid Division



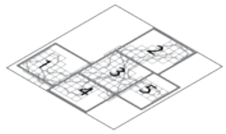
3. Once we have both 1km cells and lines selected in the first Grid Division we **overlap them** with a function called Spatial Join. The result is a 1km grid. If for example there are two lines crossing one cell then there will be two overlapping cells



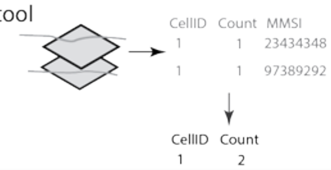
Now we **check** if the spatial join in all Grid Divisions is ready

NO YES

4. When all Spatial Joins in all Grid Divisions are ready then we can **merge** them into one



5. Now we **sum** all crossings in each cell with the Dissolve tool



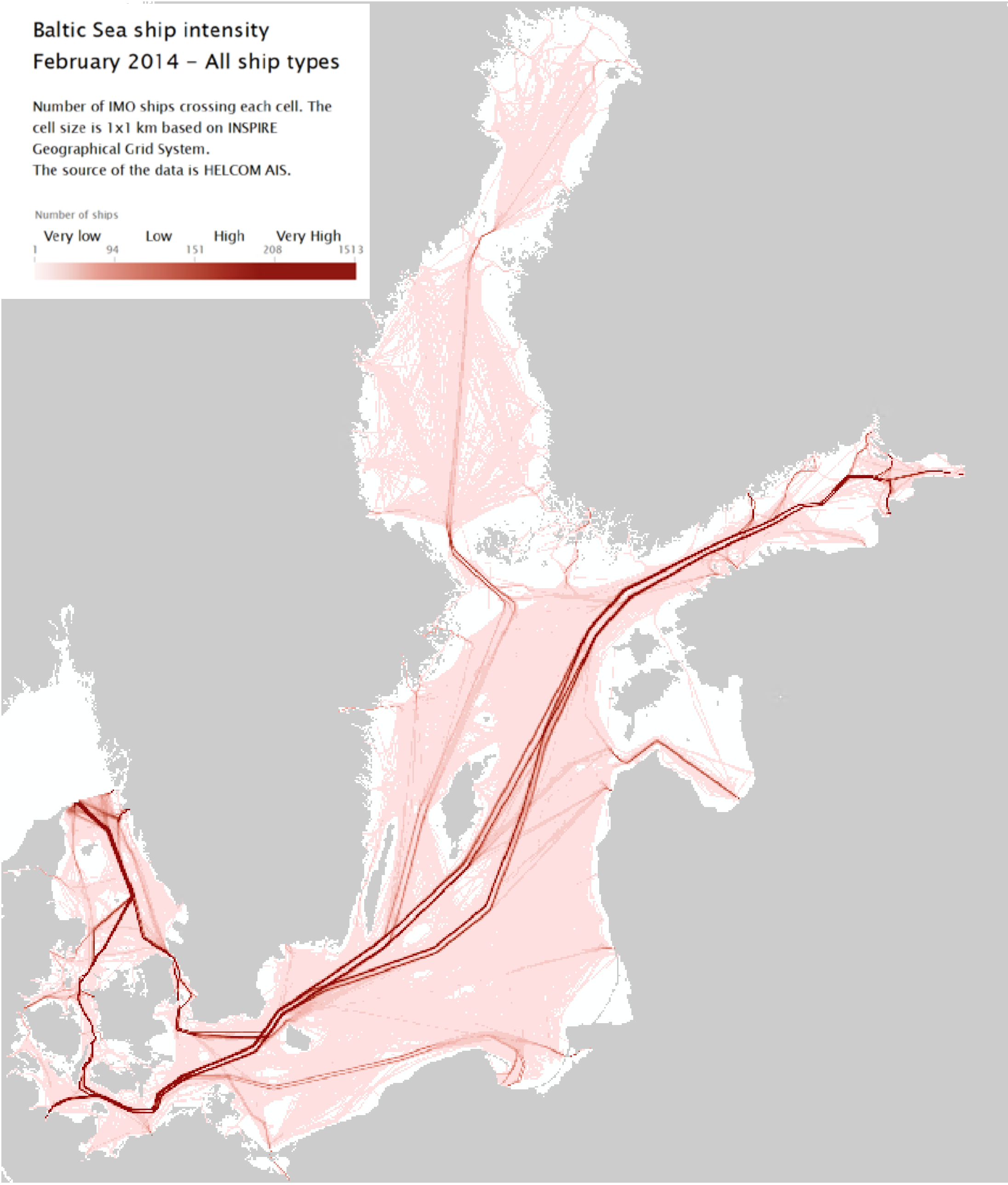
6. The result of the Dissolve tool is a huge shapefile (about 80 MB). We need to **convert it to raster** to make it lighther and more practical to work with.

Now we **check** if all monthly files are ready

NO YES

READY

When the raster is ready the symbology must be created. The final result is a map showing the shipping traffic intensity at a given time:



Annex 1: Results of AIS users interviews

This is the result of the interviews carried out by the HELCOM Secretariat at the beginning of the Baltic SCOPE project. The purpose of these interviews was to get to know how AIS data experts work with data and learn from their experiences. The focus of the interviews is on finding out what type of storage and analysis equipment they use.

Contact		General information			Storage of AIS data		Analysis of AIS data		
Country / Organization	**Name and contact**	**International Data exchange networks**	**Ideas on what HELCOM should provide**	**Additional info**	**Purpose of the storage**	**Software / Hardware**	**Types of analysis**	**Products from AIS data**	**Software / Hardware**
Denmark / Danish Maritime Authority (DMA)	*Omar Frits Eriksson, OFE@dma.dk*	HELCOM, North Sea and IALA-net. Participating in North Atlantic, American AIS (AISAS?). They provide some European member states data to EMSA.			To provide the countries a compiled version of AIS data in the Baltic Sea.		No advanced analysis.	Online map service for HELCOM Contracting Parties.	They recommend using just a powerful laptop.
EMSA	*Yann Le Moan, yann.le-moan@emsa.europa-eu*	*(This interview took place after the submission date of documents to the AIS EWG 26-2015 meeting)*	-	-	-	-	-	-	-
Finland / Finnish Meteorological Institute (FMI)	*Jukka-Pekka Jalkanan, jukka-pekka.jalkanen@fmi.fi*		To show that AIS data is relevant and the use will be increasing.	No provider of AIS data for external users. Their main focus is on writing science to produce papers.	The have all HELCOM AIS datasets from year 2005 to 2013. They run models and statistics when needed.	Basic storage in CSV divided per month on an external conventional hard disk.	STEAM – models using AIS data for emissions from ships. They use the HIS Fairplay database to have ship properties.	Analysis of emissions from ships, costs of regulations to reduce emissions, etc.	STEAM, they built their own software. No database, they use a simple workflow. Basic laptop, 16GB RAM and 64 bits.
Finland / Finnish Transport Agency	*Kaisu Heikonen, kaisu.heikonen@trafi.fi*			Bilateral agreement with Sweden, Estonia and Poland. No provider of AIS data for external users.	There is a pilot project to store national data.	Cognos for reporting and Hadoop for handling the storage.			
Lithuania / Marine Safety Agency (MSA)	*Edmundas Trusys, edmundas.trusys@msa.lt*	EMSA SSN (SafeSeaNet) and HELCOM	It would be very good to have an additional source of data than the AIS national network. Few time the connection with the AIS national network was lost and thanks to the HELCOM AIS data network, it was possible to still have data coming in.	No bilateral agreement with other countries for now, but probably in the future. Work with AIS is new in Lithuania. MSA provides AIS data or giving access to the Lithuanian SafeSeaNet (full access or with some limits, depends of the contract) for government authorities, but they must sign a contract.	Monitoring, 1 year history review in the national waters	They developed their own software. IBM blade server + virtual servers (Citrix, VMware), 20 Gb RAM + 1 To for storage. This is enough for national data.	Analysis for different purposes: ships course history, statistics, etc.	To develop graphical interface (homemade). There is already a map service for public and institutions.	Software are developed in-house. IBM Blade server, virtual servers (Citric, VMware).

Contact		General information			Storage of AIS data		Analysis of AIS data		
Country / Organization	**Name and contact**	**International Data exchange networks**	**Ideas on what HELCOM should provide**	**Additional info**	**Purpose of the storage**	**Software / Hardware**	**Types of analysis**	**Products from AIS data**	**Software / Hardware**
				Residents confirm their identity (only LT residents) to be able to monitor vessel movements with AIS (very "limitless" (public)).					
Norway / Norwegian Maritime Administration	*Harald Åseii, harald.aasheim@kystverket.no*	HELCOM and North Sea (UK, DK, etc.) international data exchange networks- Norway is running the North Atlantic network (Norway, Iceland, DK and satellites data.	These questions are more and more present in this field. It would be good for HELCOM to provide a report to answer basic questions about AIS. It would also be interesting to explain how AIS data could be merged to other data (e.g. environmental studies) to fulfill the needs of research	No bilateral agreement with other countries. AIS data provider: they provide raw data to companies (i.e. oil industry). These companies managed the data (pre-processing, database, etc.).	Storage of Norwegian AIS data network and North Atlantic. For online data viewer (2 years of data from the Norwegian network and from 2010 for the satellites).	Microsoft software – simple folders in windows explorer. Storage on conventional hard disks.		Map service available: havbase.no	
	Jon-Arve Røyset, jon.arve.royset@kystverket.no	HELCOM and North Sea (UK, DK, etc.) international data exchange networks- Norway is running the North Atlantic network (Norway, Iceland, DK) and satellites data.	HELCOM could get data flows from Havbase.	No bilateral agreement with other countries. AIS data provider through havbase. With an account, one could go deeper in filtering the ship types.	Storage of Norwegian AIS data network and North Atlantic. For online data viewer (2 years of data from the Norwegian network and from 2010 for the satellites).	PostGres (open source DB) with PostGIS extension for handling spatial data. They use a powerful laptop, but nothing special. Could not give the specifications. RAM is important for AIS data.	Pre-processing AIS data to create maps for havbase.no. There is update of the database every night to add more data. The pre-processing is automatic and new layers are generated for havbase. Use of Fairplay and ShipInfo to cross IMO number and ship caracteristics (for emissions calculation).	Everything is going on havbase. The next step is to cross havbase with weather data to be able to predict risks?	PostGres (open source DB) with PostGIS extension for handling spatial data to filter data / calculate distance, etc. Also work with google earth, ArcGIS and google API to produce graphs. A good laptop, but nothing special. Could not give the specifications. RAM is important for AIS data.
Sweden / Linköping University	*Anders Grimvall, anders.grimvall@liu.se*	They get all of the AIS data of the Baltic Sea area from the Swedish Maritime Administration.	Traffic intensity maps and shipping statistics for Helcom Assessment Areas In addition, I think that Helcom could play an important role by developing:	No bilateral agreement with other countries. No AIS data provider for external users. Data are only available for partners of the Swedish Institute for Marine Environment.	They have HELCOM AIS data for 2013 and 2014. They supply AIS data to the working group on shipping at the Swedish Institute for the Marine Environment, Uni-	SAS 9.3 (Data mining module) and QGIS Desktop 2.4. SAS because it is more documented compared to R and because they are more familiar with this software HP Elite Book	Traffic intensity maps, shipping statistics by sea area and type of ship for the entire Baltic Sea and Swedish assessment areas according to the MSFD.	Popular report (in Swedish) about the impact of shipping on the marine environment. In a relatively near future: scientific article about the use of AIS for MSP and data	SAS 9.3 (Data mining module) and QGIS Desktop 2.4. HP Elite Book 2570p, 64 bits, 16 Go RAM and GB of Hard disk. 1000 GB external hard disk.

Contact		General information			Storage of AIS data		Analysis of AIS data		
Country / Organization	**Name and contact**	**International Data exchange networks**	**Ideas on what HELCOM should provide**	**Additional info**	**Purpose of the storage**	**Software / Hardware**	**Types of analysis**	**Products from AIS data**	**Software / Hardware**
			- Guidelines for data quality assessment (identification and handling of missing values and obviously erroneous records) - Guidelines for producing traffic intensity maps in spatial scales relevant for different types of assessments and supplying AIS data in a form that is useable for environmental scientists who have a moderate experience of analyzing big data. If possible, it would also be nice if AIS data could be coupled to detailed ship data for further analysis (e.g. emissions, etc.).		versity of Gothenburg, Chalmers University of Technology (GOT) and Linnaeus University (Kalmar).	2570p, 64 bits, 16 Go RAM and GB of Hard disk. 1000 GB external hard disk.	Grid between 100 m x 100 m and 500 m x 500 m. There is a big pre-processing step before analysis: few % of AIS signals contain mistakes.	quality issues in using AIS data.	

Annex 2:

List of the ship type information from the step 3 and the creation of gross and detail ship categories.

Final ship types after Step 3	**HELCOM categorisation**	
	Gross ship types	**Detail ship types**
Aggregates carrier	Cargo	Bulk cargo
Anchor handling tug supply	Other	Tug
Anti pollution	Service	Service
Asphalt/bitumen tanker	Tanker	Other tanker
Barge carrier	Other	Other
Bucket ladder dredger	Other	Dredger
Bulk carrier	Cargo	Bulk cargo
Bulk/oil carrier	Cargo	General cargo
Bunkering tanker	Tanker	Oil tanker
Buoy tender	Service	Service
Buoy/lighthouse vessel	Service	Service
Cable layer	Service	Service
Cement carrier	Cargo	Other cargo
Chemical tanker	Tanker	Chemical tanker
Chemical/oil products tanker	Tanker	Chemical/Oil tanker
CO2 tanker	Tanker	Gas tanker
Container ship	Container	Container
Container Ship (Fully Cellular with Ro-Ro Facility)	Container	Container
Crane ship	Service	Service
Crew boat	Service	Service
Crude oil tanker	Tanker	Oil tanker
Deck cargo ship	Cargo	Other cargo
Diving	Other	Other
Dredger	Other	Dredger
Dredging	Other	Dredger
Drilling ship	Service	Service
Edible oil tanker	Tanker	Other tanker
Exhibition vessel	Service	Service
Fire fighting vessel	Service	Service
Fish carrier	Cargo	Other cargo
Fish factory ship	Fishing	Fishing
Fishery research vessel	Fishing	Fishing
Fishing support vessel	Fishing	Fishing
Fishing vessel	Fishing	Fishing
FPSO (floating, production, storage, offloading)	Service	Service
FSO (floating, storage, offloading)	Service	Service
General cargo ship	Cargo	General cargo
Grab dredger	Other	Dredger
Heavy load carrier	Other	Other
Hopper dredger	Other	Dredger
Hospital vessel	Service	Service
Hsc	Passenger	Ferry
Icebreaker	Other	Other
Inland waterways passenger	Passenger	Ferry
Inland waterways tanker	Tanker	Other tanker
Landing craft	Other	Other
Law_enforcement	Other	Other
Limestone carrier	Cargo	Bulk cargo
Liquefied gas	Tanker	Gas tanker
Live fish carrier	Cargo	Other cargo

Final ship types after Step 3	HELCOM categorisation	
	Gross ship types	**Detail ship types**
Livestock carrier	Cargo	Other cargo
LNG tanker	Tanker	Gas tanker
LPG tanker	Tanker	Gas tanker
Medical	Service	Service
Military	Other	Other
Minehunter	Service	Service
Minesweeper	Service	Service
Molasses tanker	Tanker	Other tanker
Motor hopper	Service	Service
Naval/naval auxiliary	Service	Service
Non propelled barge	Other	Other
Nuclear fuel carrier	Cargo	Other cargo
Offshore support vessel	Service	Service
Offshore tug/supply ship	Service	Service
Oil products tanker	Tanker	Oil tanker
Ore/oil carrier	Cargo	Other cargo
Palletised cargo ship	Cargo	General cargo
Passenger	Passenger	Undefined passenger
Passenger (cruise) ship	Passenger	Cruise
Passenger ship	Passenger	Ferry
Passenger/general cargo ship	Passenger	ROPAX
Passenger/landing craft	Passenger	Ferry
Passenger/ro-ro cargo ship	Passenger	ROPAX
Patrol vessel	Other	Other
Pilot vessel	Other	Other
Pipe burying vessel	Service	Service
Pipe layer	Service	Service
Platform	Service	Service
Pleasure	Other	Yatch
Pollution control vessel	Service	Service
Pontoon	Service	Service
Port_tender	Other	Other
Power station vessel	Service	Service
Pusher tug	Other	Tug
Refined sugar carrier	Cargo	Bulk cargo
Refrigerated cargo ship	Cargo	Other cargo
Research vessel	Service	Service
Ro-ro cargo ship	Container	Container
Sail training ship	Other	Other
Sailing vessel	Other	Other
Salvage ship	Other	Other
Sar	Other	Other
Seal catcher	Other	Other
Search & rescue vessel	Other	Other
Self discharging bulk carrier	Cargo	Bulk cargo
Ships_according_to_rr	Other	Other
Standby safety vessel	Service	Service
Stern trawler	Fishing	Fishing
Supply tender	Service	Service
Tanker	Tanker	Other Tanker
Towing	Other	Other
Towing_long_wide	Other	Other
Training ship	Service	Service

Final ship types after Step 3	HELCOM categorisation	
	Gross ship types	**Detail ship types**
Trawler	Fishing	Fishing
Tug	Other	Tug
Tug	Other	Tug
Undefined	Unknown	Unknown
Unknown	Unknown	Unknown
Utility vessel	Service	Service
Waste disposal vessel	Service	Service
Water tanker	Tanker	Other tanker
Vegetable oil tanker	Tanker	Other tanker
Vehicles carrier	Vehicle carrier	Vehicle carrier
Well stimulation vessel	Service	Service
Vessel (function unknown)	Other	Other
WIG	Other	Other
Wine tanker	Tanker	Other tanker
Wood chips carrier	Cargo	Bulk cargo
Work/repair vessel	Service	Service
Yacht	Other	Yatch