United Nations | Department of Economic and Social Affairs
Statistics

UNCEBD - Task Team on AIS Data

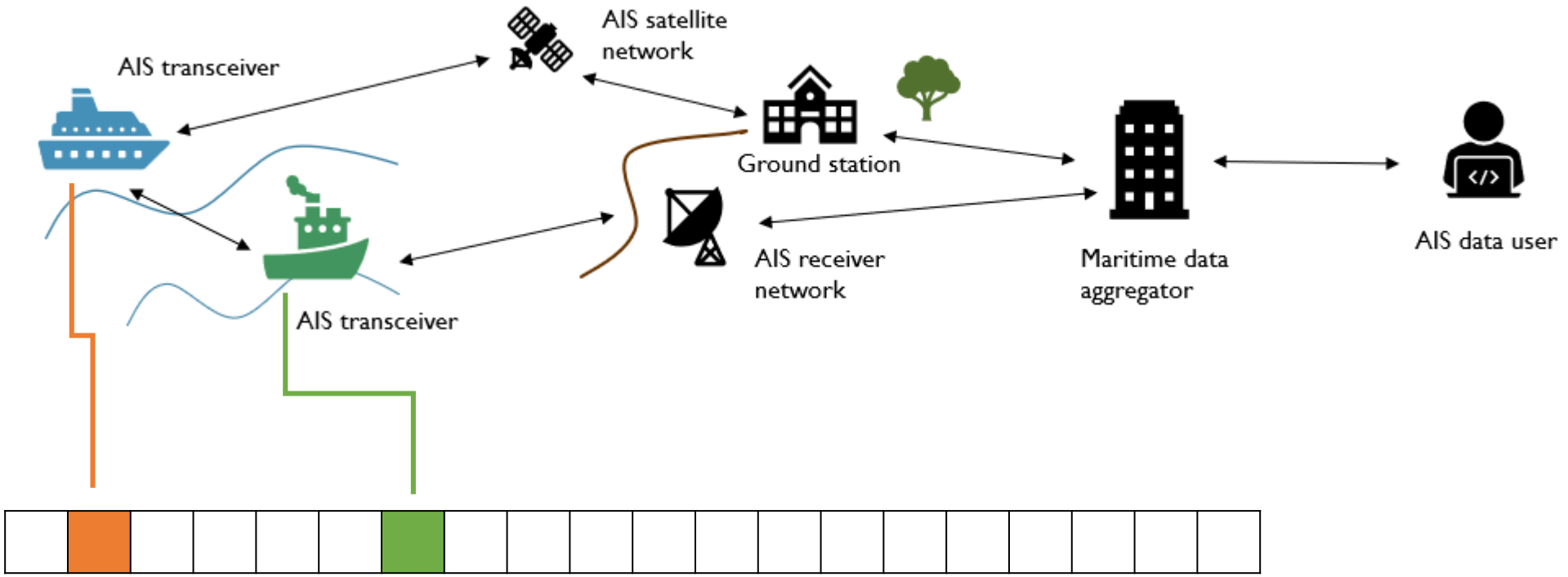
International Seminar on Big Data for official statistics in China - 9 September 2021

Photo by Markus Spiske, unsplash

Markie Muryawan
UN Statistics Division

WHAT IS AIS DATA – BASIC INFORMATION

- The automatic identification system (AIS) is a tracking system for ships, originally developed for collision avoidance



Reference: [Lloyd's List Intelligence (2017). Understanding AIS.]

Information types of AIS

27 different AIS messages types containing different types of information, identified as “static”, “dynamic” or “voyage-related” are valid for a different time periods

Static:

Dynamic:

Voyage related

Time does not exist in AIS frames. It is added by receivers

MMSI - Ship’s unique identification system

GPS data to record and track position

VHF radio signals for receiving and sending AIS information

Coverage

Which Ships

- IMO made AIS-Class A compulsory for :
  - vessels of 300 gross tonnage and upwards engaged on international voyages,
  - cargo ships of 500 gross tonnage and upwards not engaged on international voyages,
  - passenger ships (more than 12 passengers), irrespective of size for safety reasons

•AIS reporting requirements are described in Regulation 19 of Chapter V of the International Convention for the Safety of Life at Sea (SOLAS).

Time scales/Frequency of Reporting

- Few seconds to few minutes depending on the navigational status
- Global datasets : 310 billion AIS messages per year

Coverage

- Spatial : Global
- Time : as of 2005

AIS Source - Data Providers

Source

- AIS messages can be seen as another type of sensor data, regularly broadcasted by AIS responder.
- AIS messages can be received by surrounding ships, Terrestrial and Satellite AIS receiver stations.
- Terrestrial Receivers: signals ~40 sea miles

Data provider(s)

- The messages are collected and aggregated by commercial, community or port authority/coastguard as “data aggregator”
- Decoding and cleaning of AIS messages are undertaken by data aggregator
- Data aggregator may complete the coverage by exchanging data among themselves

However, some limitations

- AIS message is radio signal, it can be lost
- Message is encoded, it can be corrupted
- Transmitter has specific timeslot, in busy area not all ships can have different timeslots
- AIS transponder can be turned off
- The most important is that AIS is intended to avoid collision
- Land-based transceiver has limited coverage – needs to use satellite receiver in open sea
- Technical errors in AIS dynamic messages (due faulty equipment)
- Delay in updating AIS static messages (or no update)
- Human error when updating AIS static message

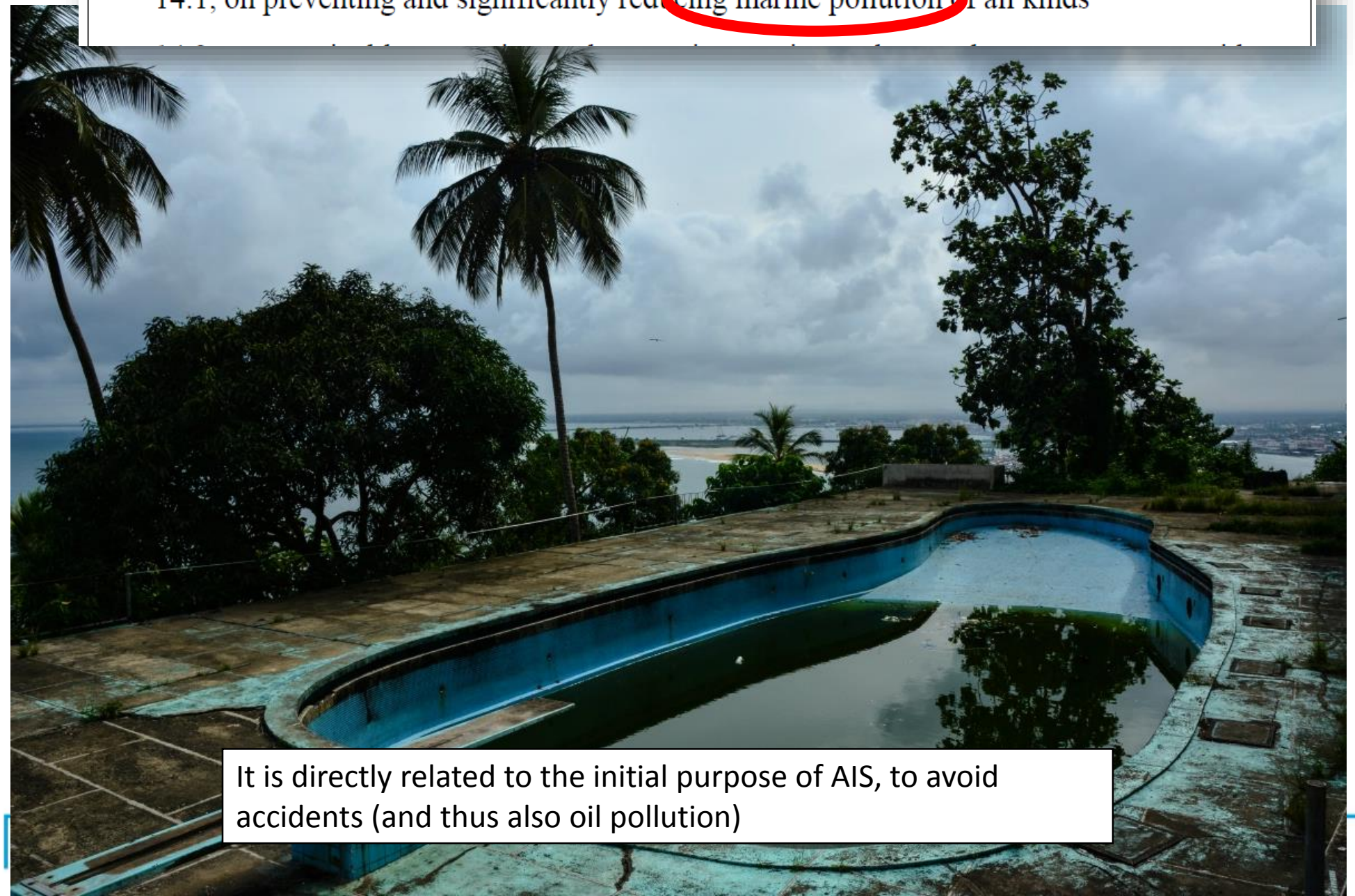
Application in real life – Use Cases

- SDGs
  - There is no specific SDGs on Maritime Transport, but Maritime Transport (and thus AIS) is important for many SDGs and targets.

- Economics, fisheries, the environment, green technology, and complement maritime/inland port authority data

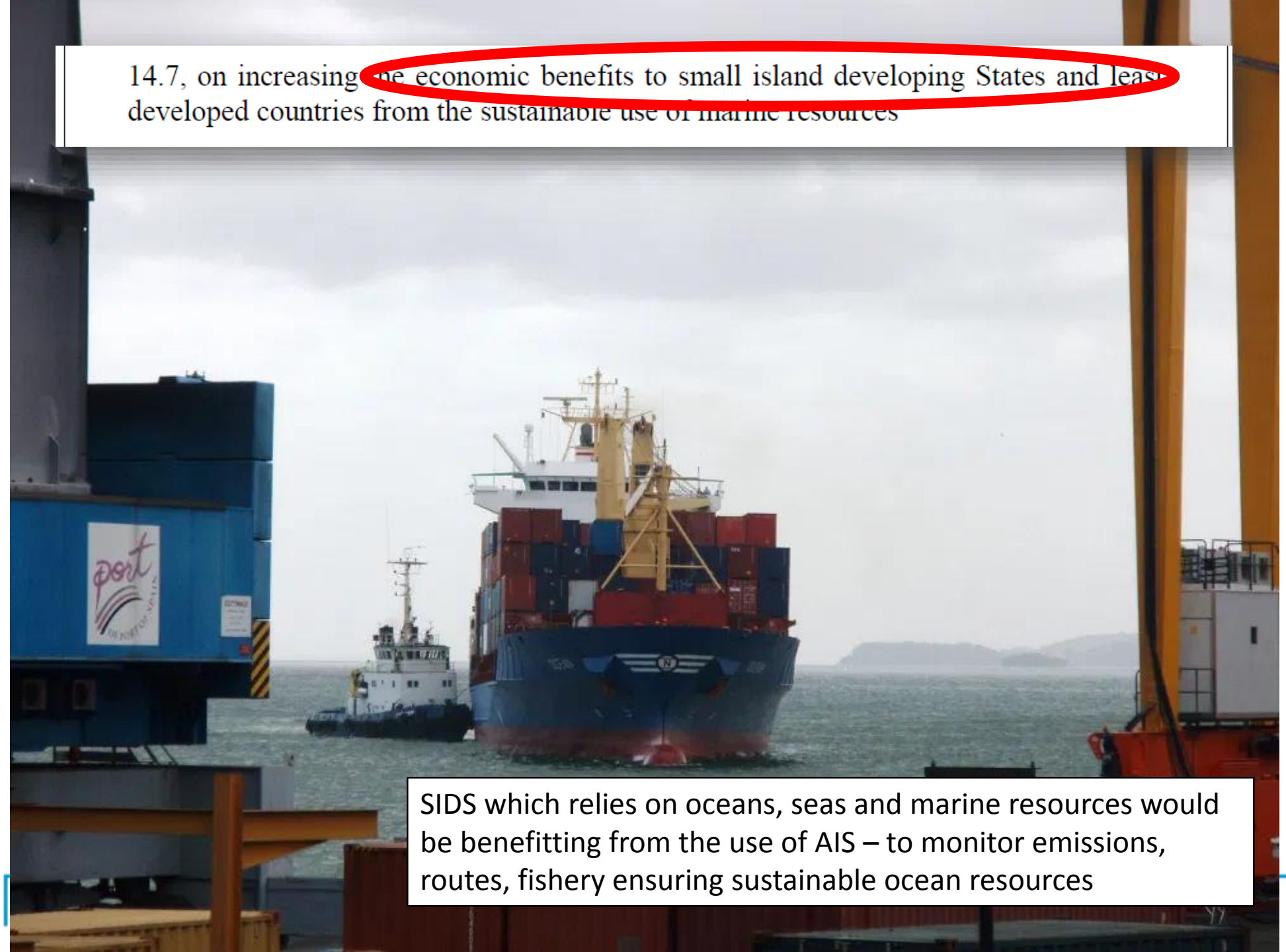
- Other emerging use cases

14.1, on preventing and significantly reducing marine pollution of all kinds



It is directly related to the initial purpose of AIS, to avoid accidents (and thus also oil pollution)

14.7, on increasing the economic benefits to small island developing States and least developed countries from the sustainable use of marine resources



SIDS which relies on oceans, seas and marine resources would be benefitting from the use of AIS – to monitor emissions, routes, fishery ensuring sustainable ocean resources

17.17. on encouraging and promoting effective public, public–private and civil society partnerships, building on the experience and resourcing strategies of partnerships

This is an example of UN Global Working Group AIS Task Team collaboration and availability of global AIS dataset

DESA

1. Faster economic indicators

Research question:
Faster economic indicators to act and adjust policy more quickly in response to economic changes.

So far, three economic/trade indicators are developed:

1. Time-in-port – aggregated time in seconds spent by ships in UK ports (further plans to disaggregate it by duration of stay, e.g., histogram of Time-in-port)
2. Port traffic – number of unique ships (MMSI) entering port
3. Number of visits/port calls, e.g. to capture multiple sailing, e.g. ferry route

Future research: anchorage or holding area outside port might contain information, e.g. holding pattern indicating port congestion etc.

References: https://datasciencecampus.ons.gov.uk/projects/faster-indicators-of-uk-economic-activity-improving-the-shipping-indicators/

1. Faster economic indicators

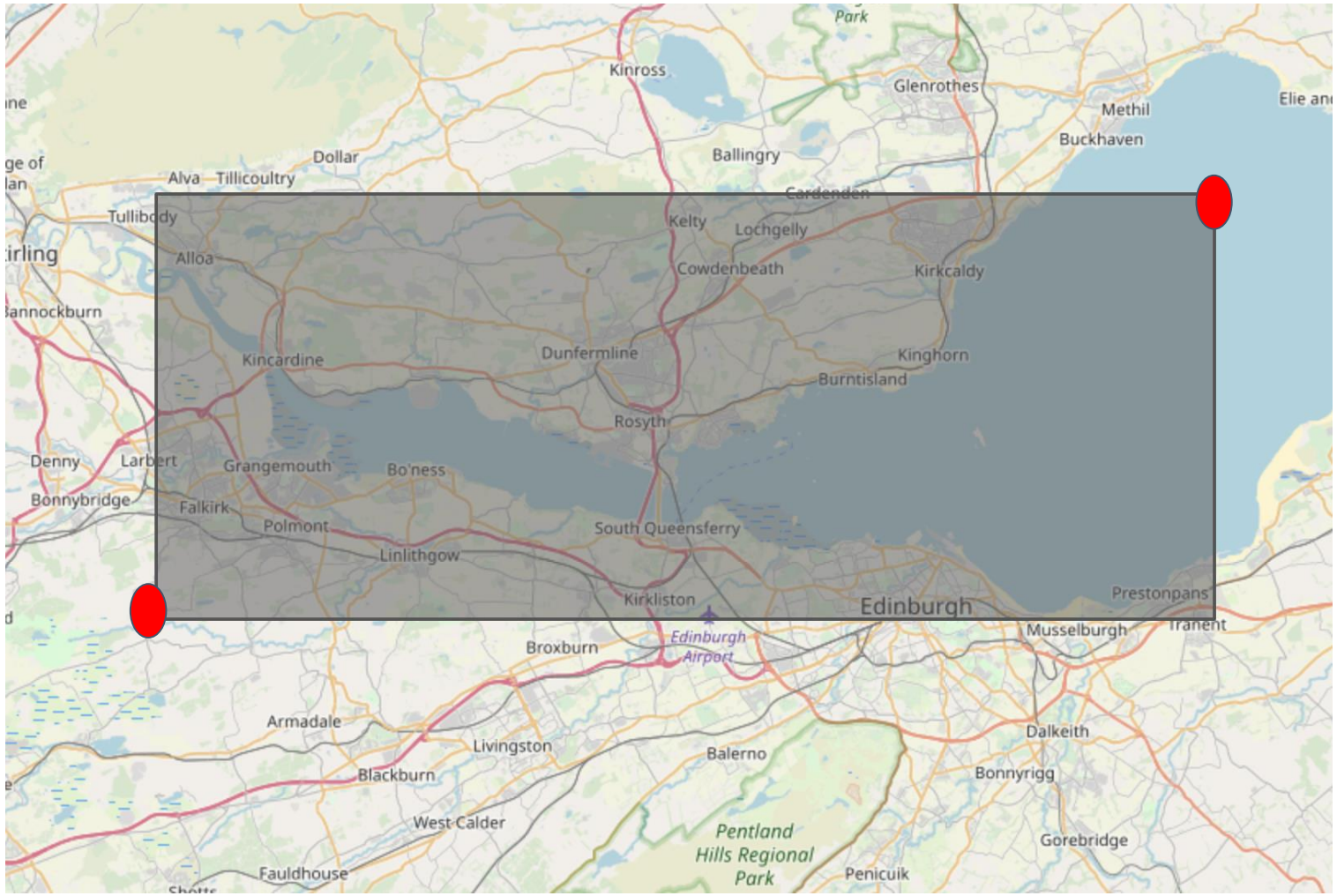
Methodology:

- Define port area using bounding box method
- Filter noises
    1. messages are not originating from ships (buoys, water level meters etc.)
    2. Ships are not considered transporting goods over certain distances

Data cleaning and preparation

- High signal/Noise ratio - in some port areas appr. 50% of messages are not originating from ships (buoys, water level meters etc.)
- “Moving ships” filter - based on condition that a ship has to travel more that a certain distance over certain period of time to be included in the computation, considerations for setting of parameters
- Ship type inference (60-80% reported) => classifier using areas of docking/maneuvers

1. Faster economic indicators



2. Trade volume estimates

Research question:
Can AIS serve as a fast and granular indicator for trade and maritime activity which could help to detect turning points on the economic cycle?

Cargo **number** indicator:

- Number of cargo ships visiting a port (filtered)
- Comparable with official maritime statistics

Cargo **load** indicator:

- Volume of cargo loaded/unloaded at a port (filtered)
- Comparable with official trade statistics

References: https://www.imf.org/en/Publications/WP/Issues/2019/12/13/Big-Data-on-Vessel-Traffic-Nowcasting-Trade-Flows-in-Real-Time-48837

2. Trade volume estimates

Methodology:

Cargo ships are identified by a filter and static and voyage-related information for the identified ships is aggregated.

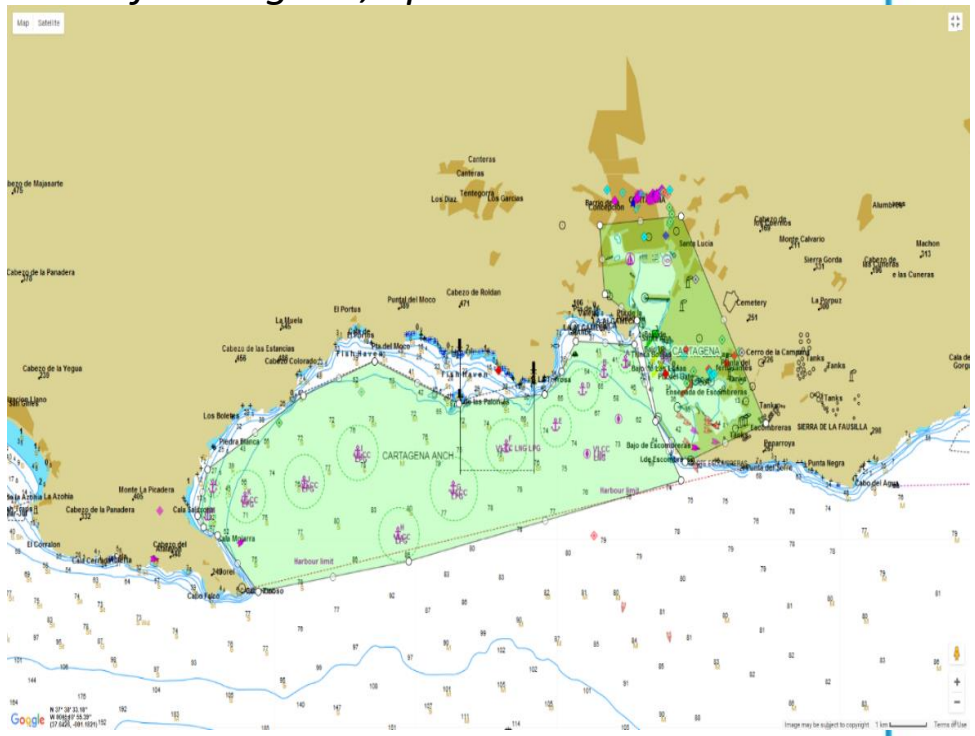
The filter identifying the cargo ships follows three rules:

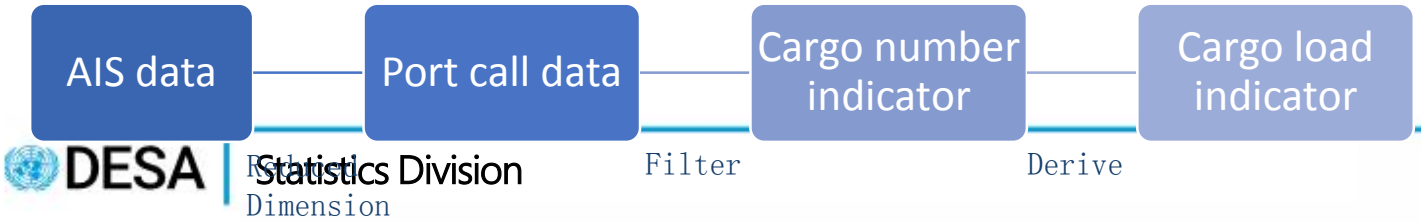
1. Bunkering tankers providing fuels to vessels located at seaports
2. Ships arriving but not departing
3. Ships that stay in the port boundaries only for a short time or for too long are omitted

High-frequency (weekly) indicators:

- Cargo number indicator that counts the number of incoming ships
- Cargo load indicator based on information on the ship's deadweight tonnage and the reported draught are calculated.

Defining Port and Anchorage Boundaries: Port of Cartagena, Spain

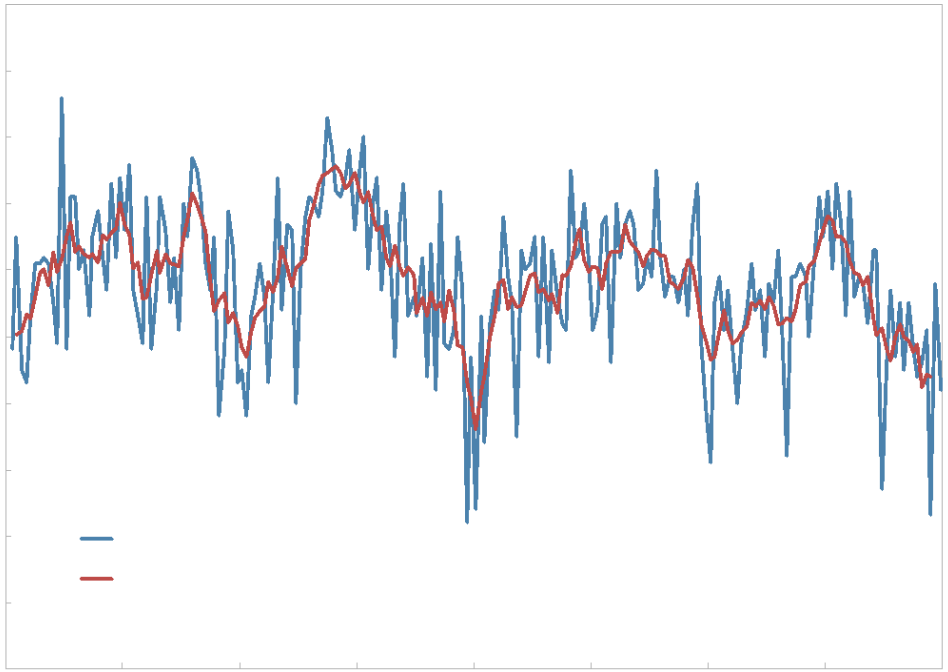




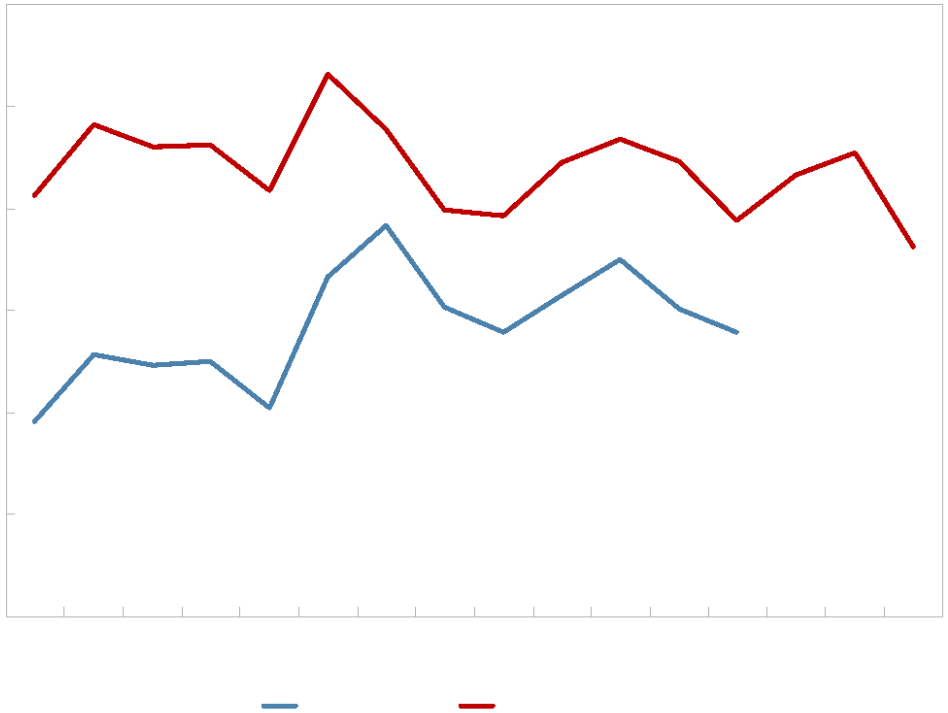
2. Trade volume estimates

Malta: Cargo Number Indicator (weekly)

Number of cargo ships visit using AIS-based port calls data, 2015-2018



Sources: MarineTraffic, staff estimates.



Sources: Eurostat, MarineTraffic, staff estimates.

3. Cruise Tourism

Research question:
Is it feasible to combine Maritime registers (Customs Declarations of Ships), Automatic Identification System (AIS-data) and web scraped data (Cruise Ship Calendar) to estimate the number of cruise ships and number of cruise passengers arriving at Bonaire?

- Focus on port of Kralendijk (Bonaire)
- Using platform STEALTH: browser-based dynamic spatial visualization tool

Motivation:

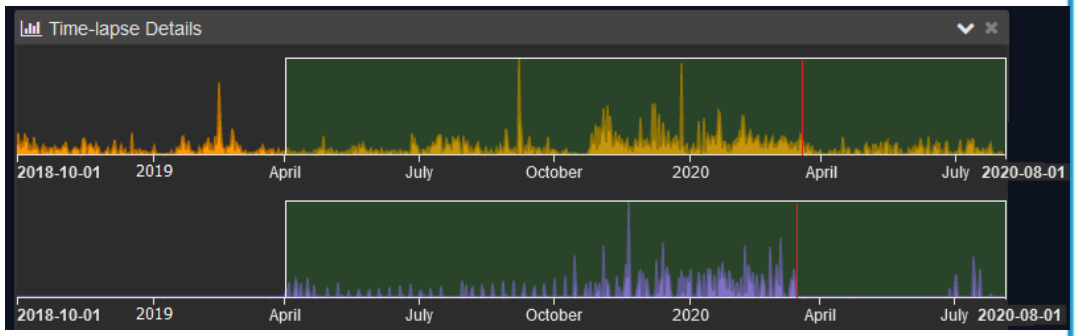
- Cruise tourism important economic activity: 458 000 passengers in 2019 vs. 158 000 passengers by air
- Reliable statistics is important for policy issues
- Validation of register of the Customs Office for Bonaire (current source of official statistics)

Information about cruises also interesting for other islands.

3. Cruise Tourism

Methodology:

- Setting a port area bounding box
- Calculate port calls for ships type cruise
- Benchmark the data with administrative records
    - Link the three data sources using either MMSI-numbers and/or IMO-numbers.
    - Implement reliable quality measures to assist an efficient output validation across these three data sources.

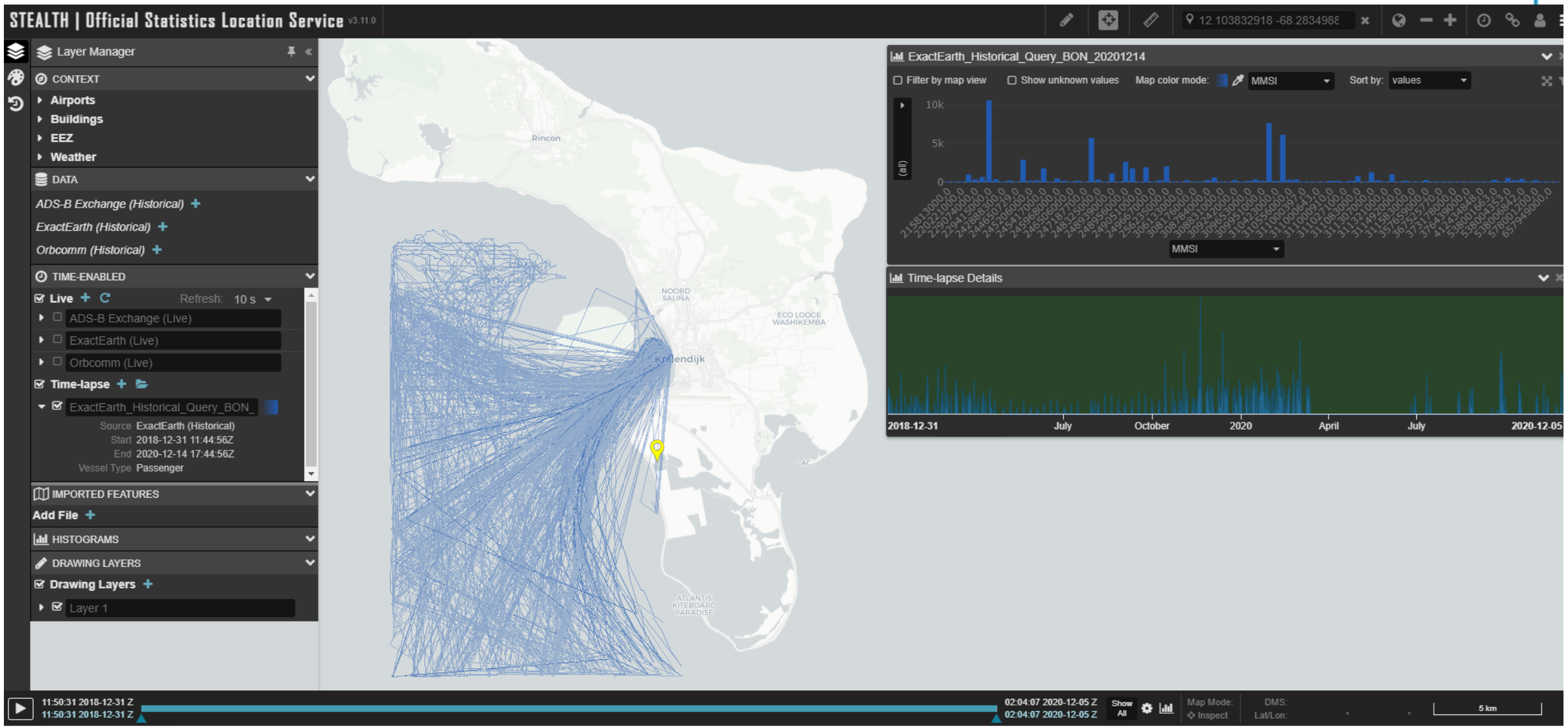


Combination of data sources is work in progress.

Limitations:

- Effect of COVID-19
- Hard to evaluate if passengers visit the island
- While AIS data does not contain actual number of passengers,
- it can be used to get insight in the route and stops of cruise ships.

3. Cruise Tourism



4. Fisheries

Research question:
Can we disseminated fisheries using a geography that covers the entire surface of the Earth.

- Official fishing locations = sensitive information
- High aggregation level
- First time at EU level

Goals:

- Conservation and protection (i.e. collisions with whales).
- Illegal, unreported and unregulated fishing (IUU) Fishing
- Transshipments

The data available are not fit to assess the local impact of fishing on the seafloor.

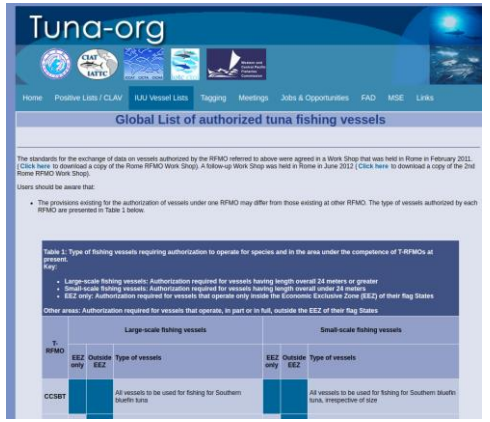
https://unstats.un.org/wiki/display/AIS/Case+studies

Powered by EMODnet

4. Fisheries

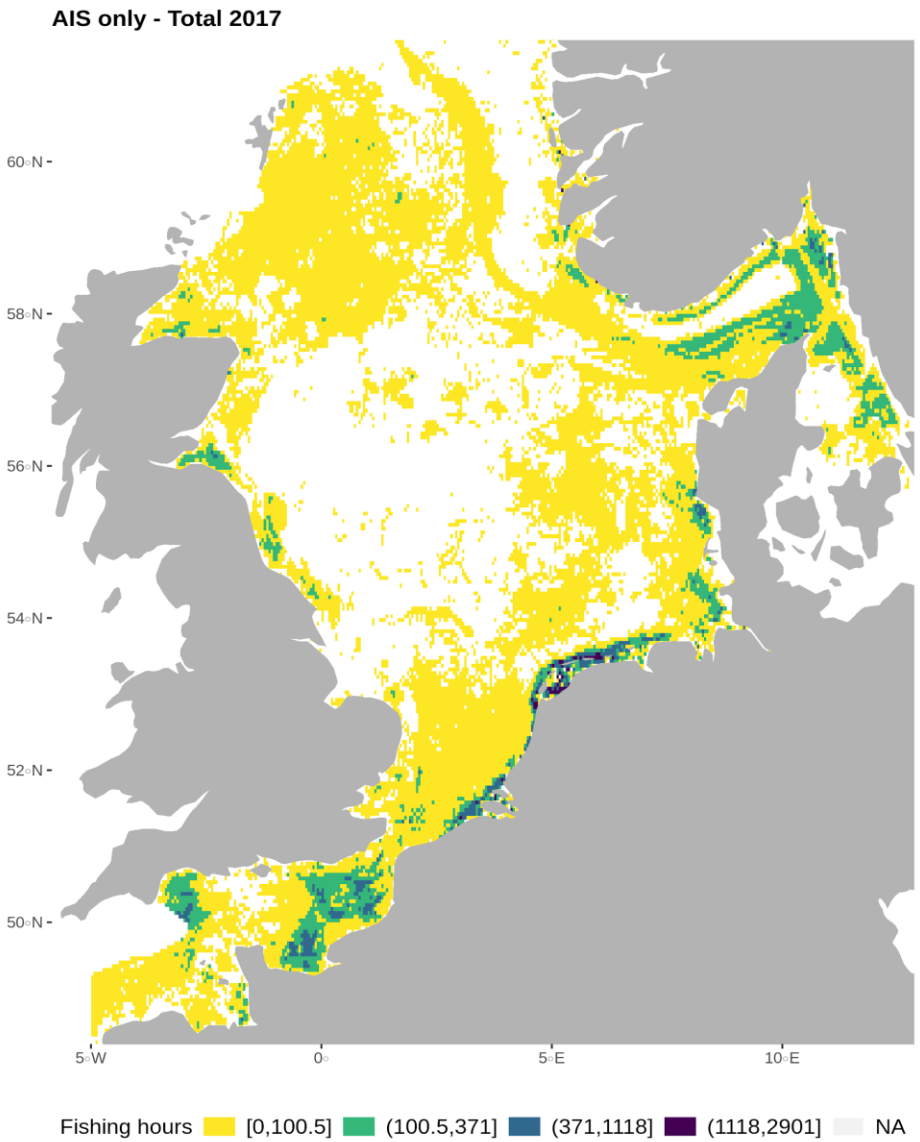
Methodology:

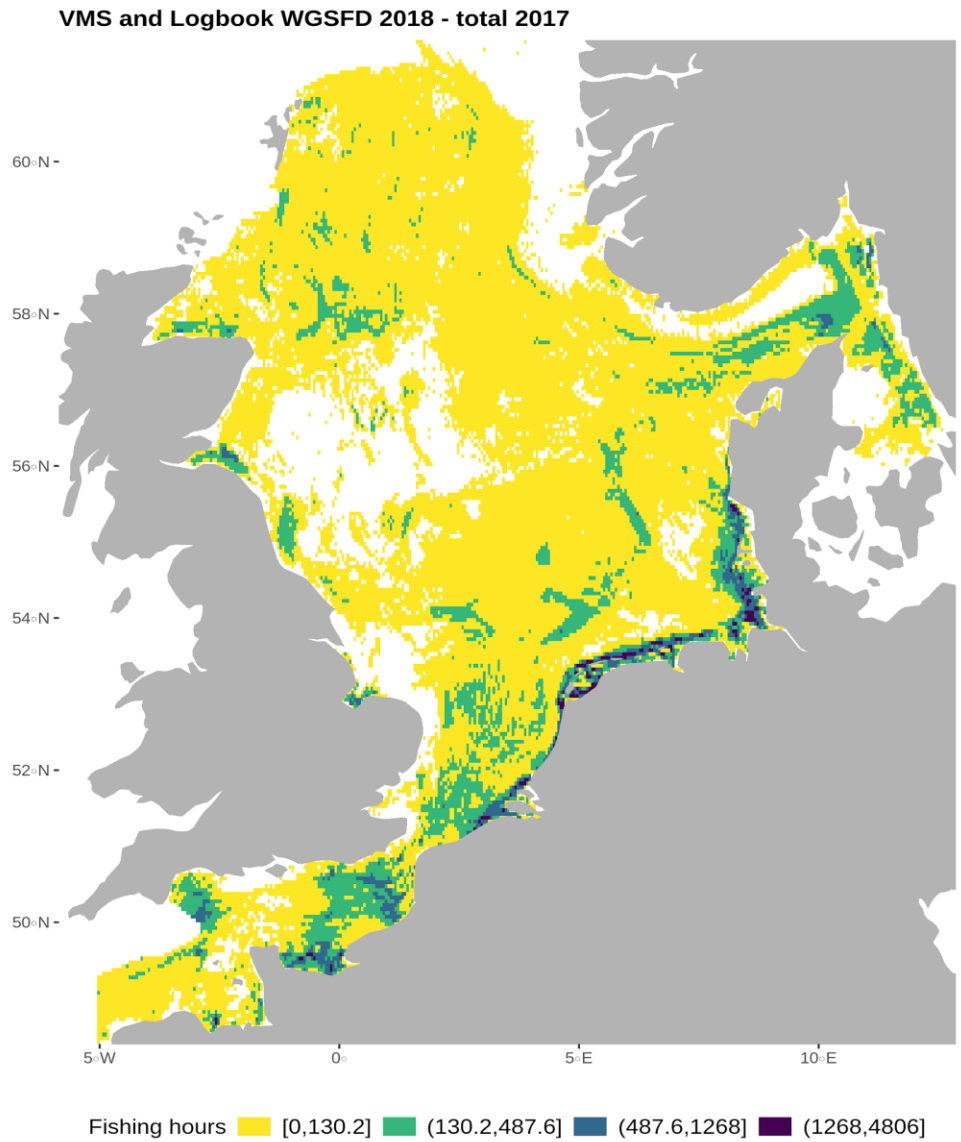
1. Extract fishing fleet by filtering shiptype = 30
2. Clean the dataset:
   - Visible errors (lat, lon, speed, course)
   - Zero speed in port: point in polygon (ports dataset)
   - Filter out zero speed
3. Obtain gear information
   - Fleet registers: EU Fleet register, Regional Fisheries Management Organizations (noisy, duplications)
   - Model: Google models gear and effort
4. Calculate fishing effort
   - Mobile bottom contacting gears
   - Speed filter in combination with direction, some models need labels: Gaussian Mixtures, HMM, Random Forests, Deep learning
   - Other gears
5. Validate and assess uncertainty
6. Aggregate and publish data



Powered by EMODnet

4. Fisheries





Powered by EMODnet

5. Maritime emissions

Research question:
Can we create a global map of estimated shipping emissions, visualized in an easy-to-use dashboard?

- Geographical distribution of emissions according to vessel locations and activity using a machine learning model
- Interactive map, various zoom levels
- Users: data scientists, general public

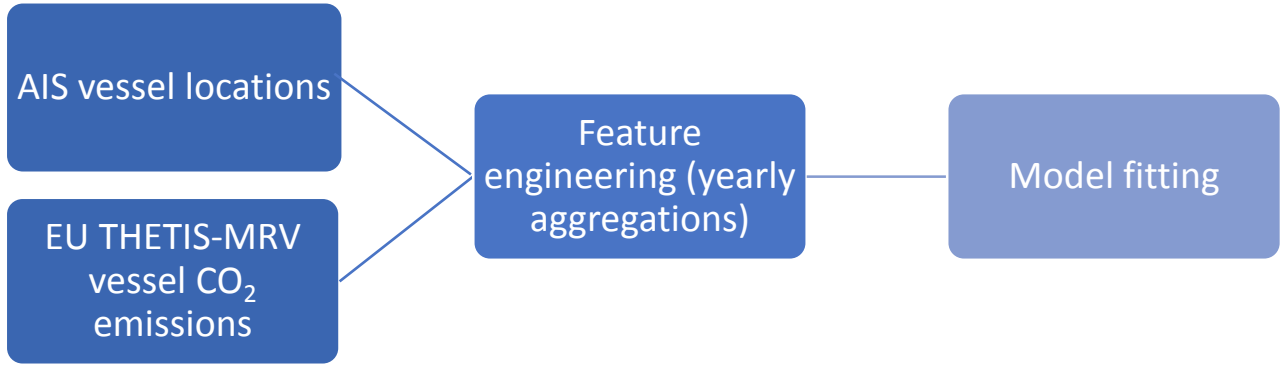
Context: Winner AIS hackathon 2020

https://www.unglobalpulse.org/2020/10/un-hosts-first-ais-big-data-hackathon/

5. Maritime emissions

Methodology:

- Machine Learning: estimate daily geographical vessel emissions
- Spatial data aggregation



IMO	Name	Ship Type	Technical efficiency		Repor... Period	Total CO_2 emissions [m tonnes]
			Ty...	(gCO_2/t·nm)		
5383304	ASTORIA	Passenger ship	EIV	169.16	2019	24512.83
5383304	ASTORIA	Passenger ship	No...		2018	20080.25
6417097	MARCO POLO	Passenger ship	EIV	68.95	2019	26799.64
6417097	MARCO POLO	Passenger ship	No...		2018	25689.03

imo	annual_co2_m_tonnes	avg_speed	avg_draught	length	width	category_name
9241061	151825,28	15,0506	10,3048	345,03	48,7	Passenger
9351490	150190,98	18,1703	6,8219	225	30,41	Passenger
9227417	120551,01	15,751	6,4158	199,9	25	Passenger
9214276	114103,05	12,946	7,2438	211,5	30,4	Passenger
9198927	113410,93	14,4973	6,6251	203,9	25	Passenger
9333175	112949	11,2708	8,4932	288,61	36,05	Passenger
9141065	108956,53	10,5485	8,2694	268,6	32,3	Passenger
9351476	108952,61	12,4126	6,8524	253,96	30,41	Passenger
9760512	107284,8	11,4852	8,6272	315,83	43	Passenger
9595321	107165,58	9,6841	8,6058	333,33	37,92	Passenger
9351488	106431,2	10,9741	6,912	224,97	30,41	Passenger

Big Data Curriculum - Core

Foundation Level Intermediate Level Expert/Advanced Level	Core* Curriculum													
Knowledge Areas >>	Coding				Coding Good Practice	Github	The UNGP	Methodology	Project Management	Big Data Leadership		Mentoring	Data Visualisation	
Courses >>	Intro to Python	Intermediate Python	Intro to PySpark	Intro to SQL	Coding Good Practices (1.4)	Intro to Github	Optimising use of the UNGP	Intro to ML	Agile	International Data Masterclass	International DS Knowledge Exchange	Virtual International DS Accelerator Mentoring	Data Visualisation for Dissemination (1.10)	

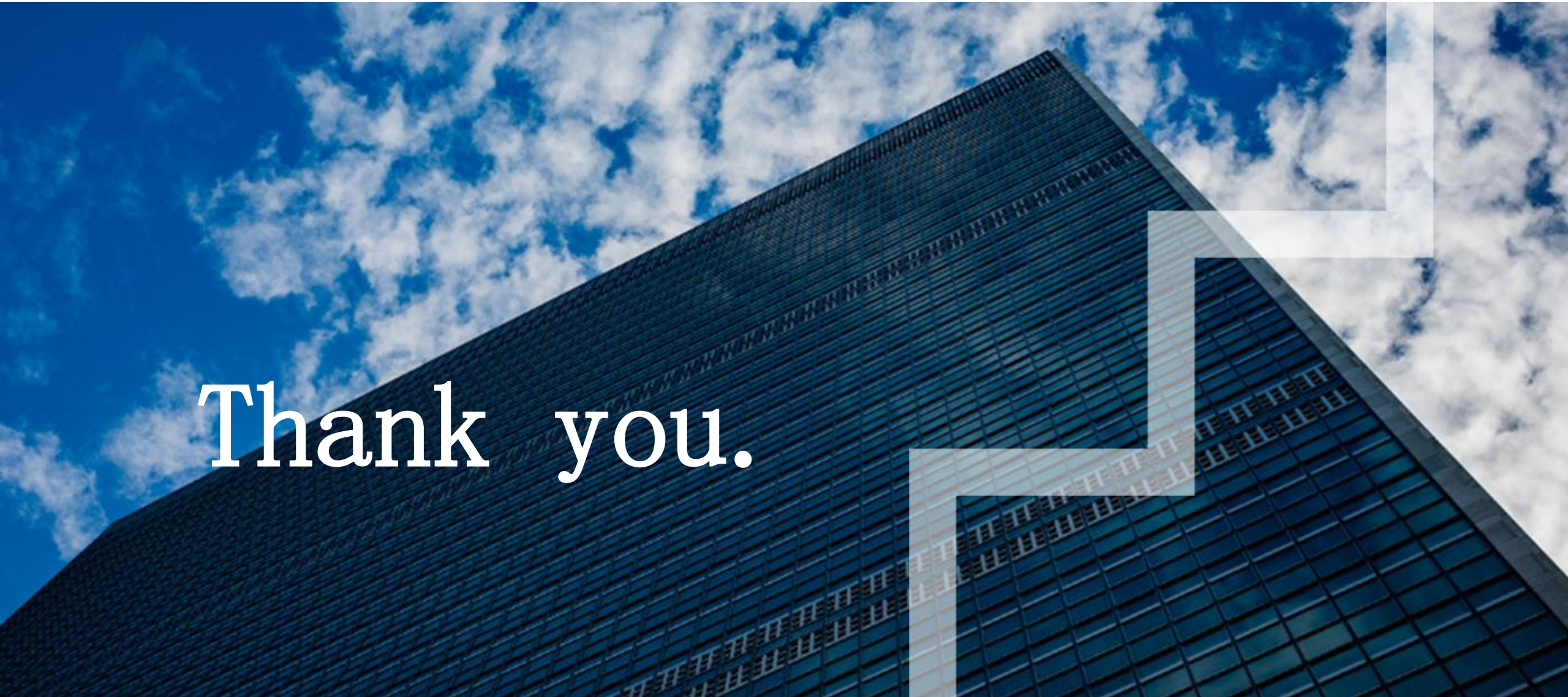
Core* Refers to learning events that have wider applicability and will be placed into a central part of t

Availability	*UK Campus*	*UK Campus*	*External*	*External*	*External*	*External*	*External*	*External*	*External*	*UK Campus & UN*	*UK Campus & UN*	*UK Campus & UN*	*External*	
Audience														
Manager/Decision Maker										(X)	(X)			
Project Manager						X	X		X			(X)		
Data Scientist	X	X	X	X	X	X	X	X	X			X	X	
Analyst/Statistician	X	X	X	X	X	X	X	X	X			X	X	
Other NSO staff	X	X							X					

Big Data Curriculum - AIS

Foundation Level Intermediate Level Expert/Advanced Level	AIS Big Data Training Curriculum							
Knowledge Areas >>	The Art of the Possible	Data Acquisition		Data Cleaning	Data Integration & Analysis			
Courses >>	What is possible with AIS data? (1.1) Priority	An Introduction to utilizing AIS data for real life use case (1.2)	Acquiring AIS data via the UNGP (1.3)	Methods for cleaning AIS data (1.5)	Linking AIS data with other data sources (1.6)	Developing a Faster Economic Indicator using AIS data (1.7)	Calculating Maritime Emissions using AIS data (1.8)	Analyzing Fishery activities using AIS data (1.9)
Core* Refers to learning events that have wider applicability and will be placed into a central part of the overall Big Data Training curriculum								
Availability	*UNGP LMS*	*In Dev*	*In Dev*	*In Dev*	*In Dev*	*In Dev*	*In Dev*	*In Dev*
Audience								
Manager/Decision Maker	X							
Project Manager	X		X					
Data Scientist	X	X	X	X	X	X	X	X
Analyst/Statistician	X	X	X	X	X	X	X	X
Other NSO staff	X							

United Nations | DESA
Statistics Division



Thank you.