Using new and innovative data sources for transport statistics and indicators

Instructions: Click on the link to access each author's presentation.

Organiser: Nikolaos Roubanis

Chair: Christophe Demunter

Participants:

Evangelia Ford-Alexandraki: Traffic and Mobility

Miriam Blumers: Early estimates of maritime traffic using innovative data sources

Frank Halmans: Multimodal container transport in the Netherlands

Michal Bis: Experience of Statistics Poland in the production of experimental statistics on road and maritime transport using innovative data sources

IAOS-ISI 2024
MEXICO CONFERENCE

Traffic and Mobility Indicators from Innovative Data

- Eurostat -

INEGI
International Statistical Institute
IAOS IMPROVING OFFICIAL STATISTICS

Traffic and Mobility lab project

Aim: Develop experimental transport indicators on traffic and mobility using innovative data & establish new ways of processing and sharing innovative data to produce statistics

- Joint project with Eurostat's Unit A5*
  Project started in 2022, expected end in 2024

- 2022: landscaping study to identified promising new data sources for meaningful transport indicators

- 2023: select 3 use cases for transport indicators,
  develop agreements with relevant partners to get access to the data and develop methodology for producing indicators

- 2024: pilot indicator methodology

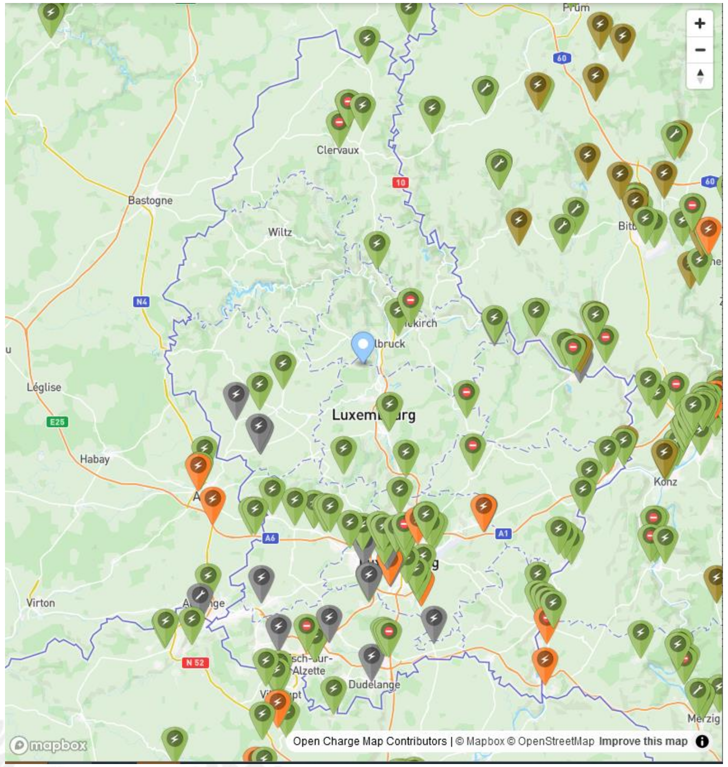
* supported by a contract with PwC

Traffic and Mobility – Selected Use Cases

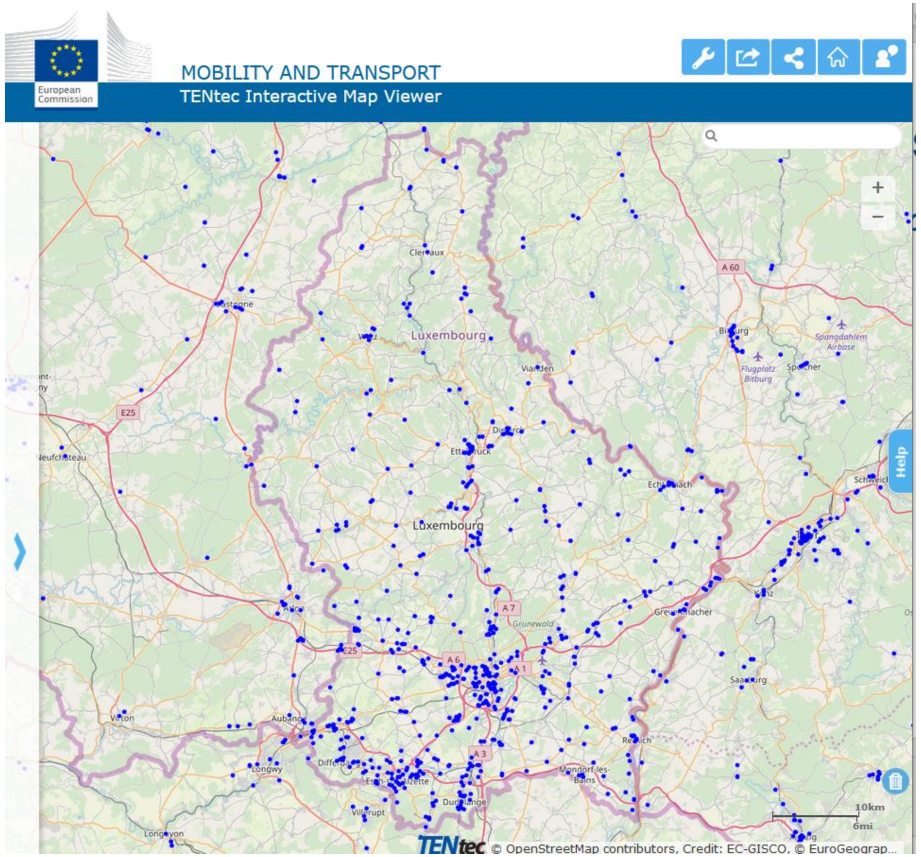
<table>
<tr><th>Use Case 1:
Adoption of Alternative Fuels</th><th>Use Case 2:
Availability of Public Transport</th><th>Use Case 3:
Air Quality Traffic Pollutants Levels</th></tr>
<tr><td>Measuring distribution and capacity of publicly available alternative fuels infrastructure (recharging stations) based on crowdsourced data in NUTS 2/3 regions.</td><td>Measuring availability of public transport using GTFS and crowdsourced data in NUTS 2/3 regions.</td><td>Measuring average concentration of selected air pollutants at peak traffic times and their variation based on the European Environment Agency air quality database and TomTom traffic data.</td></tr>
<tr><td colspan="3">Indicators:</td></tr>
<tr><td><u>Indicator 1.1</u>
➢ Charging infrastructure density
<u>Indicator 1.2</u>
➢ Charging infrastructure network capacity
<u>Indicator 1.3</u>
➢ Charging infrastructure distribution</td><td><u>Indicator 2.1</u>
➢ # of stops / (population and/or area km^2)
➢ Average # of lines serving public transport stops per NUTS 2/3 region
➢ Times of day when public transport is available
<u>Indicator 2.2</u>
➢ Travelable distance via public transport in a given time frame (in terms of % of region area and/or % of population reached)</td><td><u>Indicator 3.1</u>
➢ Average level of air pollutant at peak traffic times (over day/week/working days/ month per City or NUTS 2/3 region)
➢ Average difference of level of air pollutants between peak traffic times and baseline
<u>Indicator 3.2</u>
➢ # of Km with both high traffic and high air pollutant concentration</td></tr>
</table>

1. Charging infrastructure for alternative fuels

Crowd-sourced or commercial data?



[Source: Open Charge Map]



[Source: EAFO/ Eco-Movement data visualized in TENtec Interactive Map Viewer]

IAOS-ISI 2024
MEXICO CONFERENCE

eurostat

1. Charging infrastructure for alternative fuels

Data sources

Open Charge Map *(open-source)*

European Alternative Fuels Observatory *(proprietary)*

Open Street Map

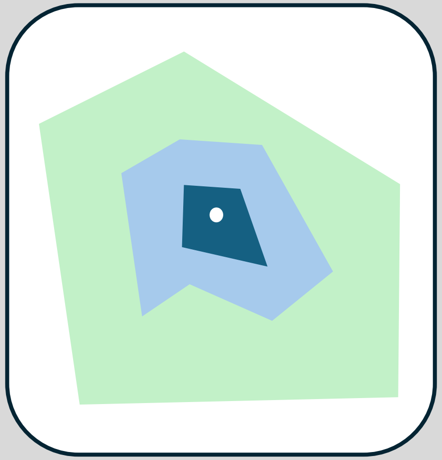
NUTS raw data

Calculation

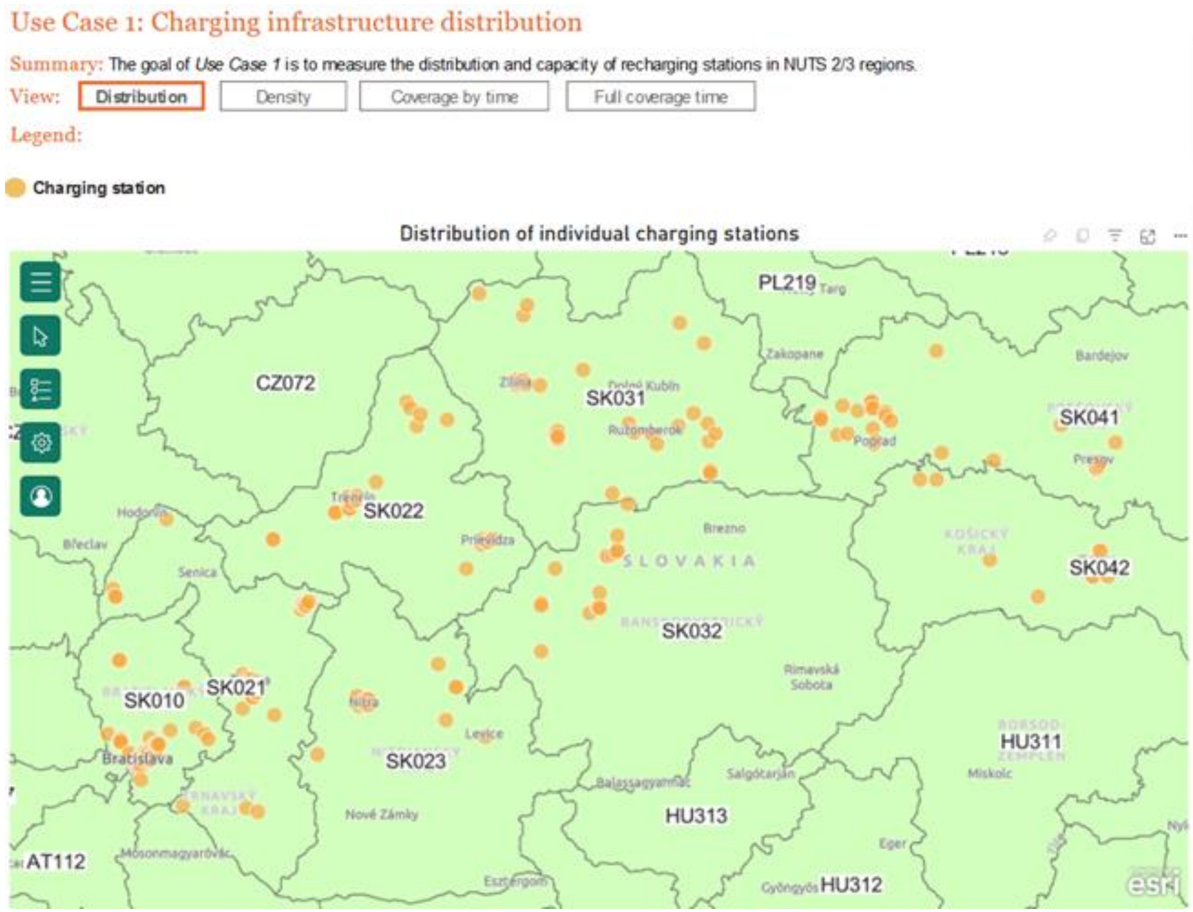
Attribution of recharging station to NUTS region

Origin-destination pairs
-> distance between recharging stations

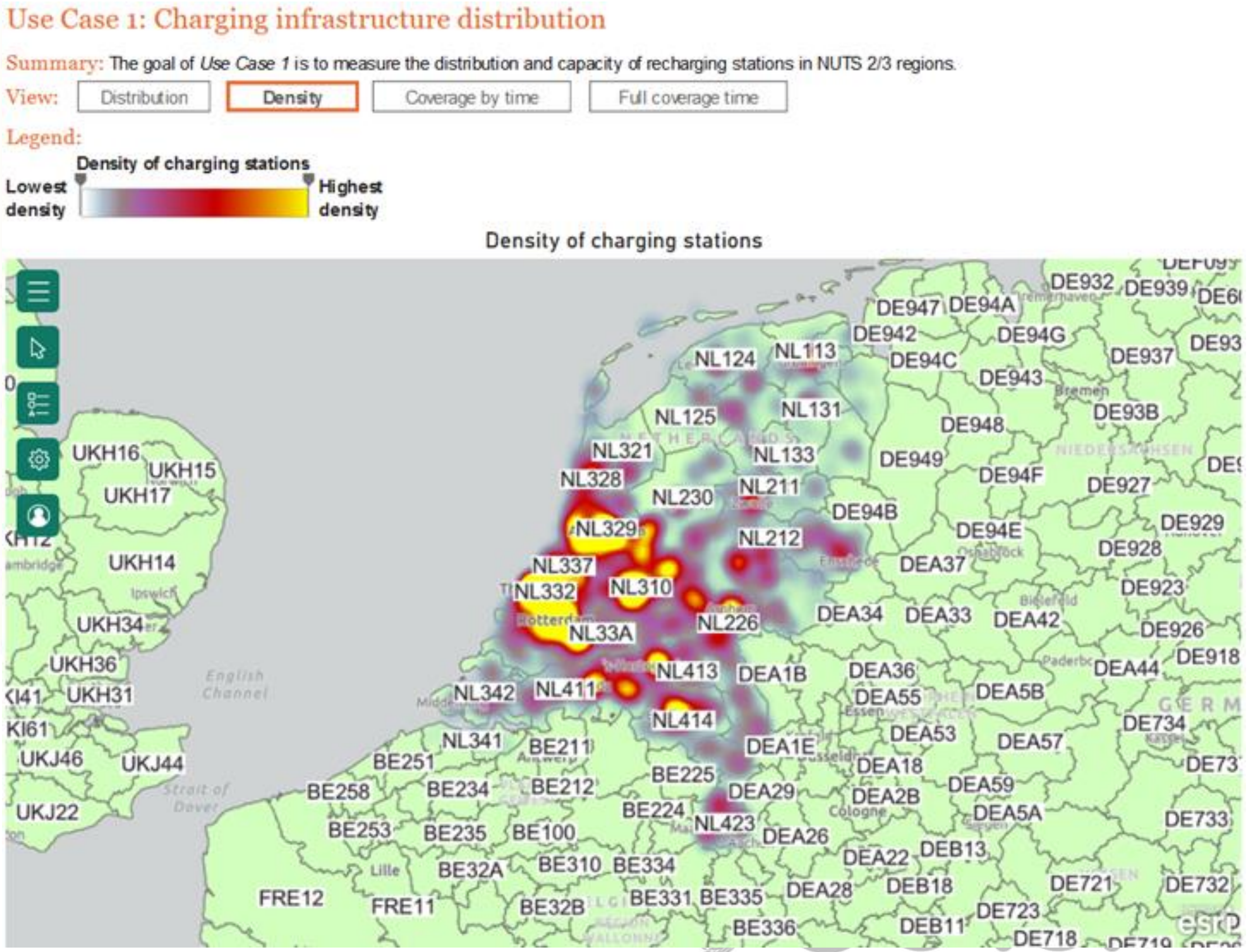
Isochrone polygons
-> Percentage of a NUTS region covered within a set travel time



Output



Source: Preliminary project results implemented in Power BI using ArcGIS and based on Open Charge Map data



Source: Preliminary project results implemented in Power BI using ArcGIS and based on Open Charge Map data

Use Case 1: Charging infrastructure distribution

Home

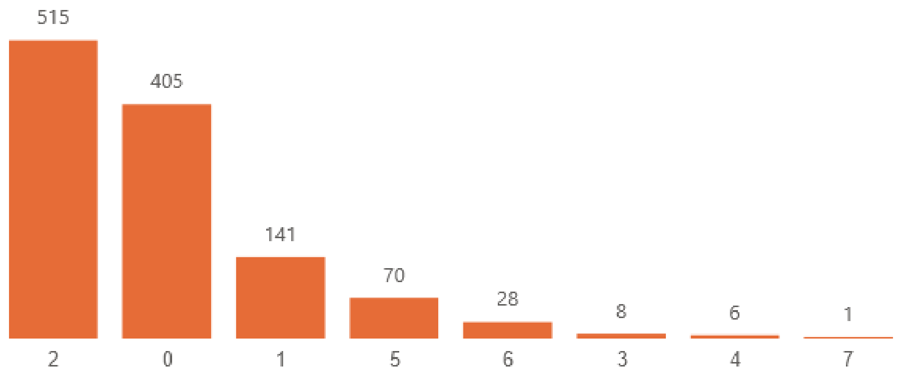
Summary: The goal of *Use Case 1* is to measure the distribution and capacity of recharging stations in NUTS 2/3 regions.

View: 1.1 - Charging stations by Region | **1.2 - Charging stations by Category** | 1.3 - Distribution of charging stations

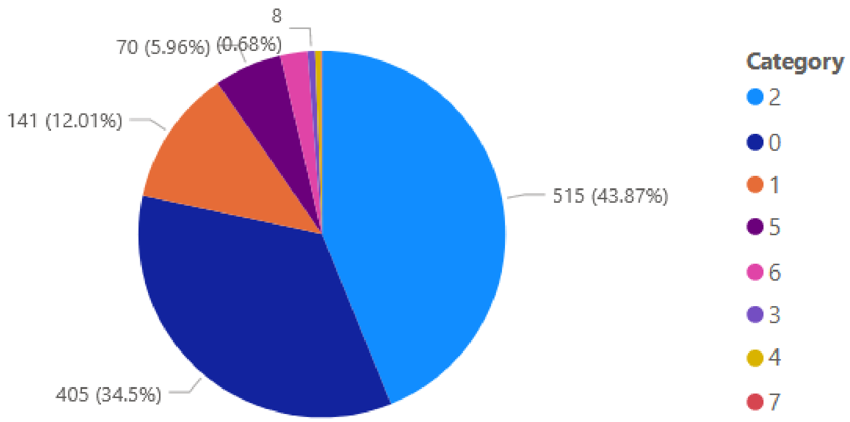
Indicator 1.2 overview

NUTS ID	Category	# Charging stations
BE234	2	97
BE211	2	79
BE100	2	73
BE234	0	63
BE211	0	44
BE241	0	39
BE100	0	37
BE251	2	31
BE242	2	27
BE211	1	22
BE241	2	22
BE242	0	21
BE224	2	19
BE328	0	19
BE212	2	17
BE224	0	17
BE212	0	16
BE251	0	15
BE310	2	15
BE332	0	14
BE213	0	13
BE213	2	13
BE254	0	12
BE223	2	11
BE211	5	10
BE213	1	10
Total		**1,174**

of Recharging stations by Category



of Recharging stations by Category



KPIs

Total charging stations	Total number of NUTS
1,174	39

Filters

Source
- EAFO
- OCM (selected)

Country
- BE (selected)
- NL
- SK

NUTS Level
- NUTS 2
- NUTS 3 (selected)

NUTS ID
- BE100
- BE211
- BE212
- BE213
- BE223

Category of Charger
- 0. Unknown
- 1. Slow AC recharging point, single-phase
- 2. Medium-speed AC recharging point, triple-phase
- 3. Fast AC recharging point, triple-phase
- 4. Slow DC recharging point
- 5. Fast DC recharging point
- 6. Ultra-fast DC recharging point (Level 1)
- 7. Ultra-fast DC recharging point (Level 2)

Source: Preliminary project results implemented in Power BI and based on Open Charge Map data

Use Case 1: Charging infrastructure distribution

Home

Summary: The goal of *Use Case 1* is to measure the distribution and capacity of recharging stations in NUTS 2/3 regions.

View: 1.1 - Charging stations by Region | 1.2 - Charging stations by Category | **1.3 - Distribution of charging stations**

Indicator 1.3 - %Coverage at current time radius

NUTS ID	Time radius (minutes)	NUTS covered%
BE100	10	100.00
BE211	10	83.92
BE212	10	92.28
BE213	10	65.43
BE223	10	56.12
BE224	10	75.10

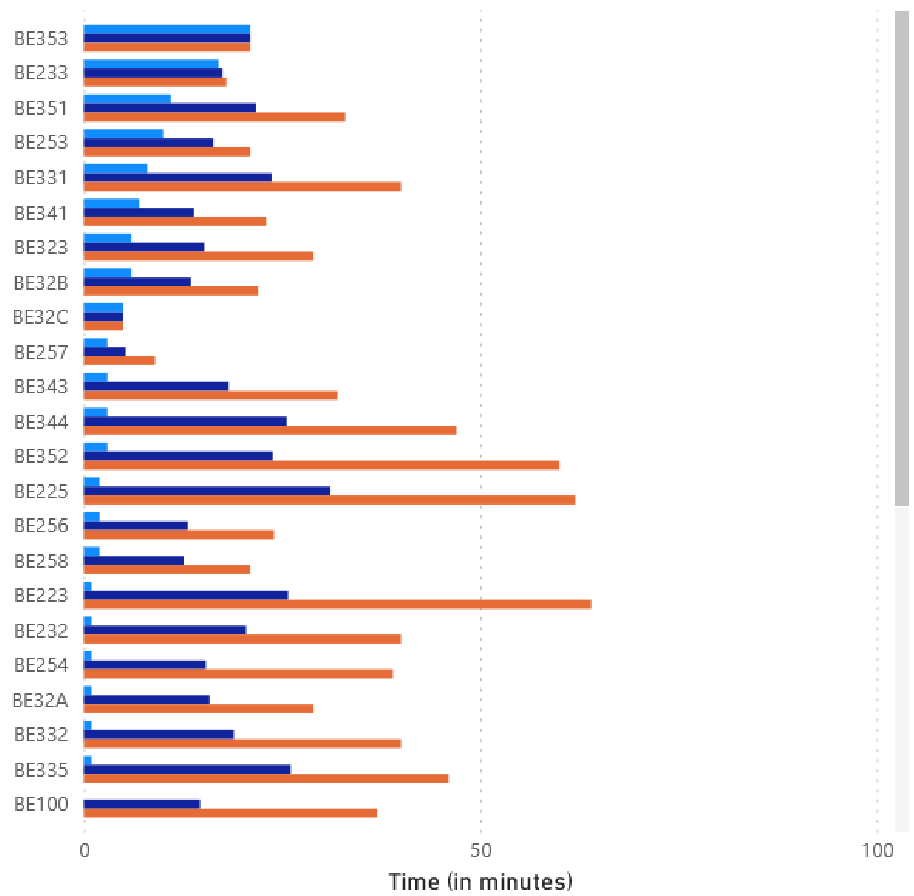
Indicator 1.3 - Time needed to cover all points

NUTS ID	Time needed (minutes)
BE100	10
BE211	30
BE212	20
BE213	30
BE223	60
BE224	20
BE225	50

Indicator 1.3 - Travel time between stations in minutes

NUTS ID	Minimum	Average	Maximum
BE310	0	25.45	81
BE213	0	26.85	72
BE223	1	25.77	64
BE241	0	22.21	63
BE225	2	31.13	62
BE212	0	22.26	60
BE352	3	23.84	60
BE224	0	19.77	51
BE234	0	15.18	49

Min, Average, and Max travel time between stations in a given NUTS



Filters

Source
- EAFO
- OCM (selected)

Country
- BE

NUTS Level
- NUTS 2
- NUTS 3 (selected)

NUTS ID
- BE100
- BE211
- BE212
- BE213
- BE223

Time radius (minutes)
- 10 (selected)
- 20
- 30
- 40
- 50
- 60

Source: Preliminary project results implemented in Power BI using Open Street Map and based on Open Charge Map data

Considerations

- NUTS delimitation:
  - precision requires scale -> scale affects calculation time
  - OSM data is structured in polygons that need to be wholly included
    -> roads segments might slightly cross NUTS borders
- Computation time for infrastructure distribution is high (OCM: 10h, EAFO: 3 days)

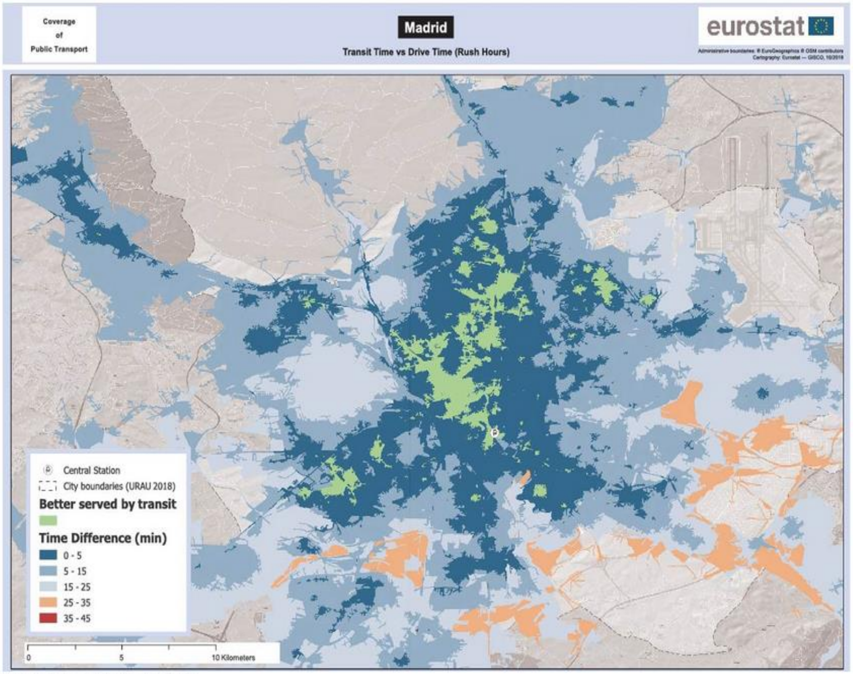
Areas for future development

- Indicators currently fully piloted for one MS (BE, partially for NL and SK)
- Data orchestration (data ingestion, transformation and indicator calculation) could also be used for hydrogen infrastructure

2. Public transport

	Indicators
2.1	**Availability of public transport stops per NUTS2 and NUTS3 region** i.e. the average number of stops and lines serving these stops, the frequency, and the percentage of time during the day that public transport is available
2.2	**Efficiency of public transport** i.e. percentage of region area and the percentage of population that can be reached in certain amount of time ***& Comparison with other mode (car)***

Comparative accessibility of Madrid Atocha train station by car and by public transport (transit)



Source: Eurostat analysis on behalf of ECA.

2. Public transport

Data sources

General Transit Feed Specifications data *(open source)*

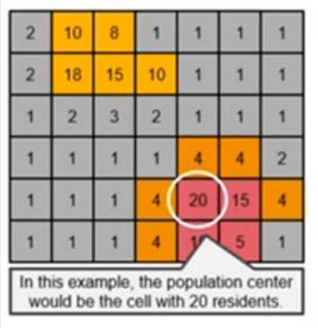
Open Street Map *(open source)*

EU population data

NUTS raw data

Calculation

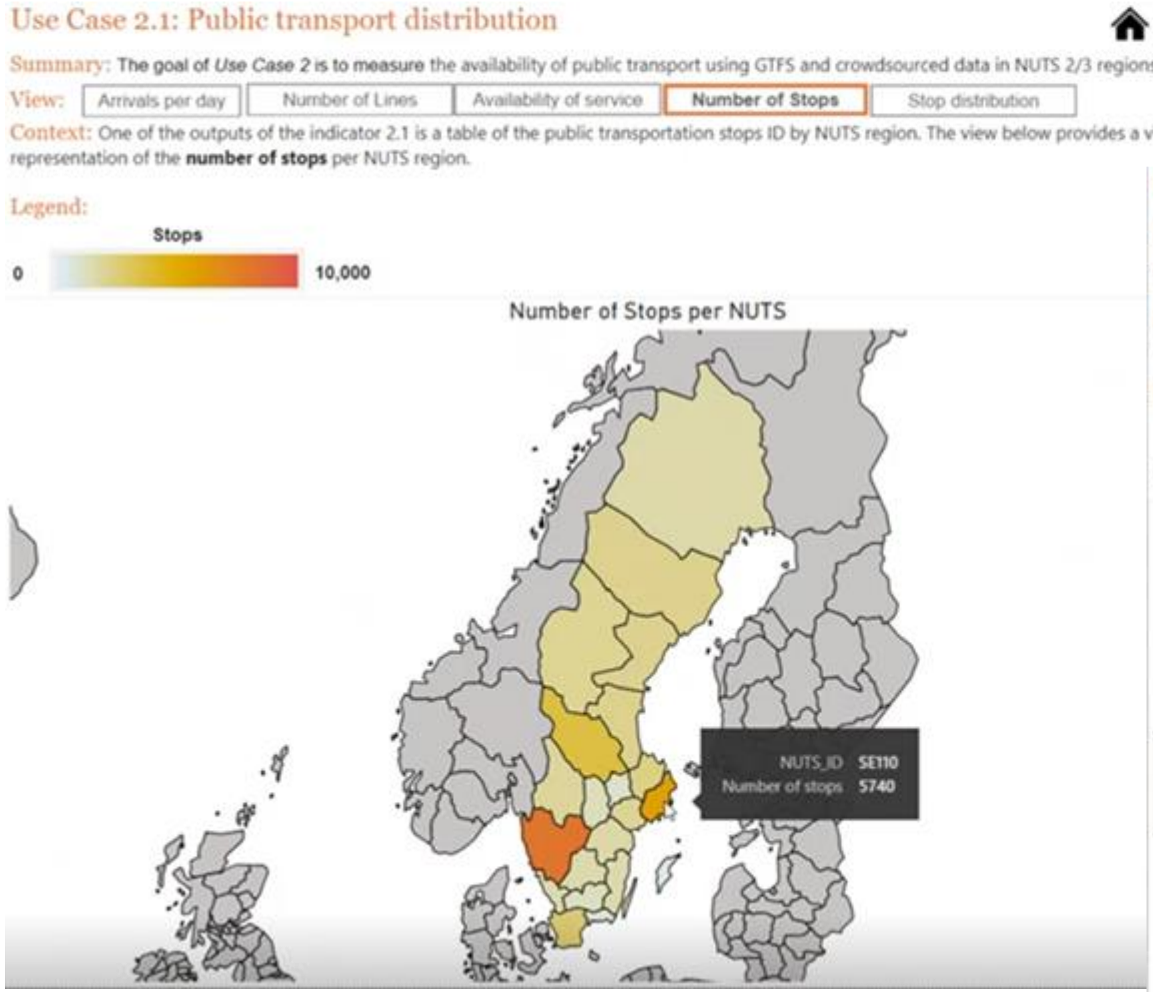
Attribution of stops to NUTS region and aggregation of lines, stops and departures



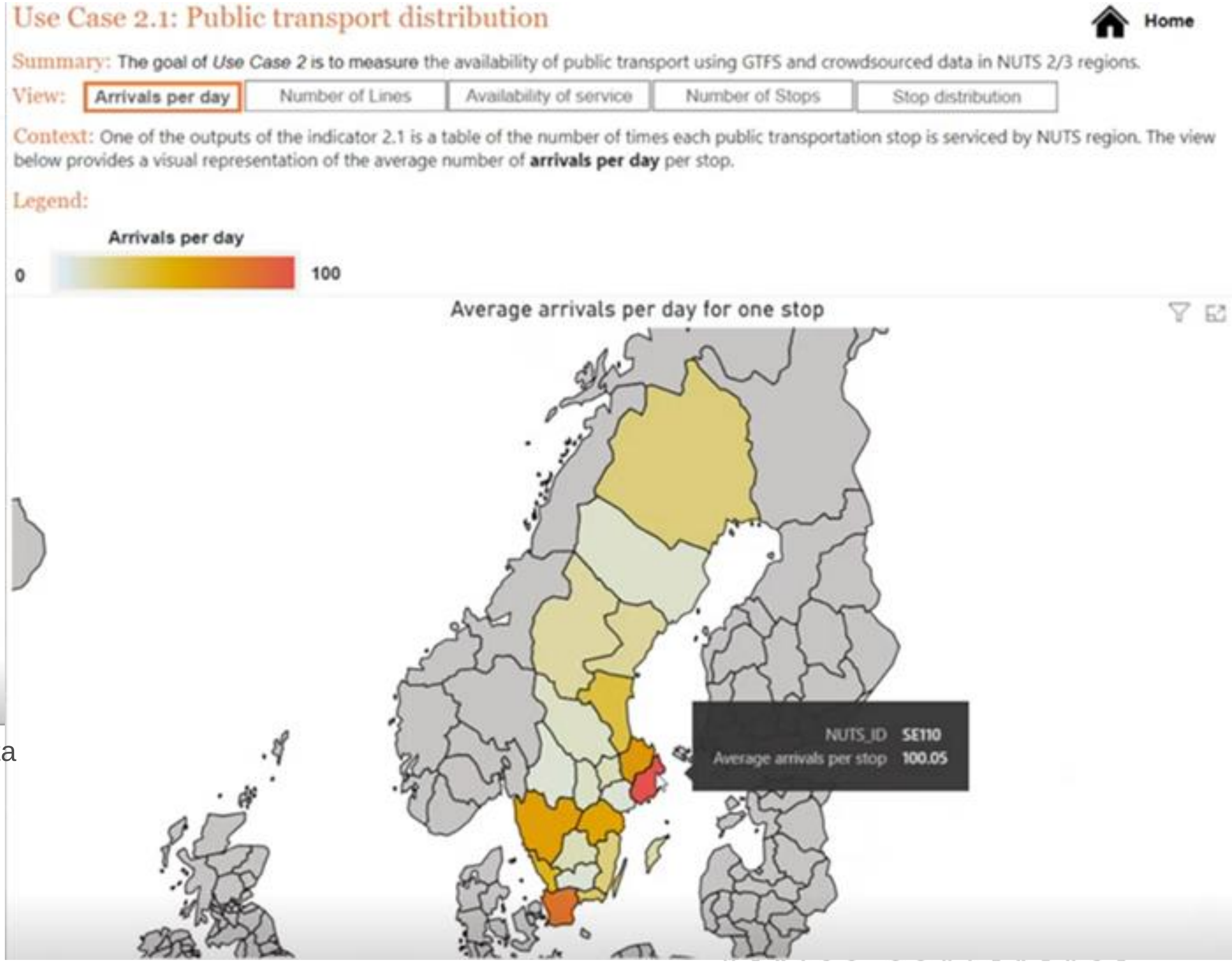
Clustering algorithm to determine population center = origin point for analysis of reach

Isochrone polygons
-> Percentage of a NUTS region and population reached within a set travel time

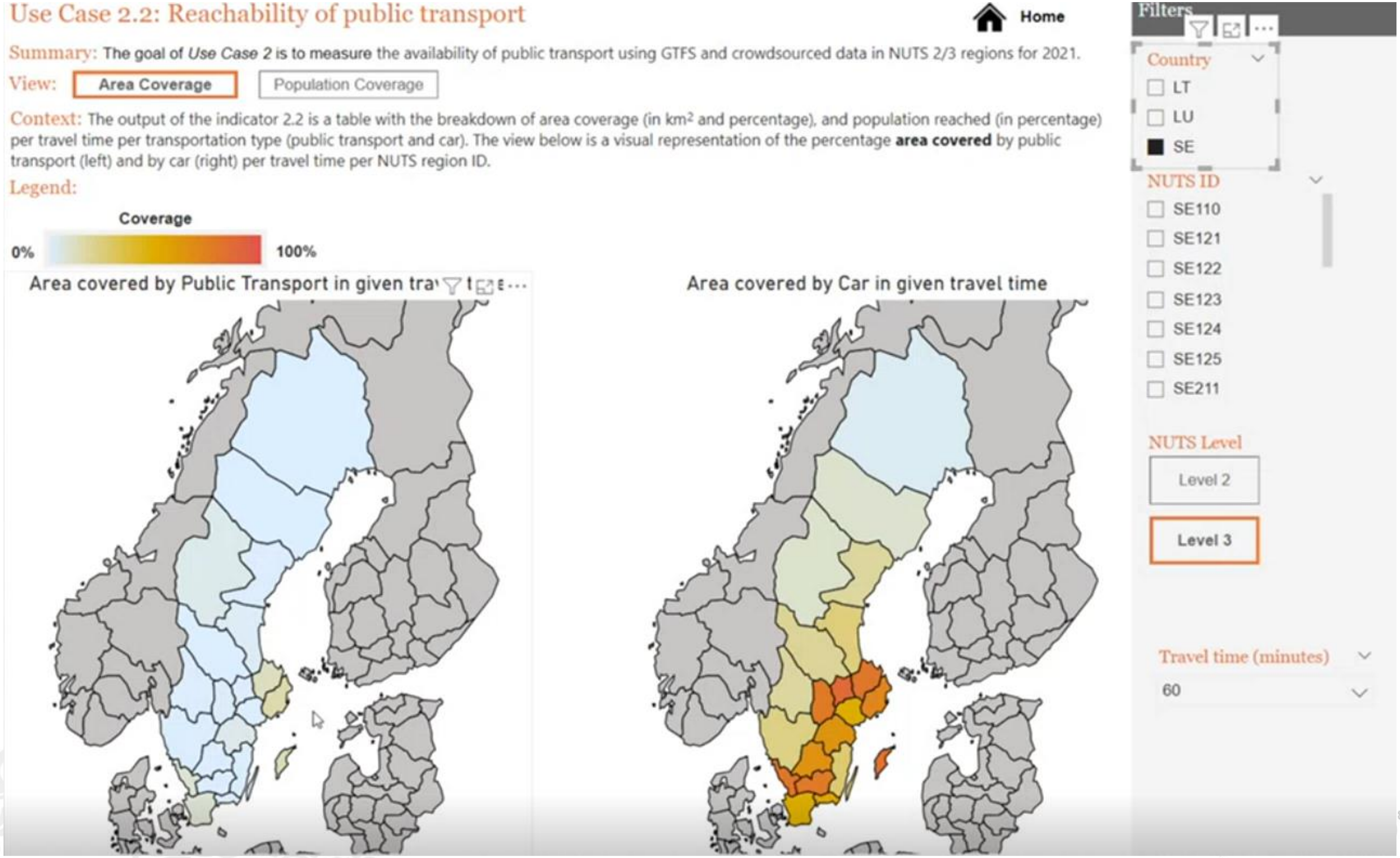
Output



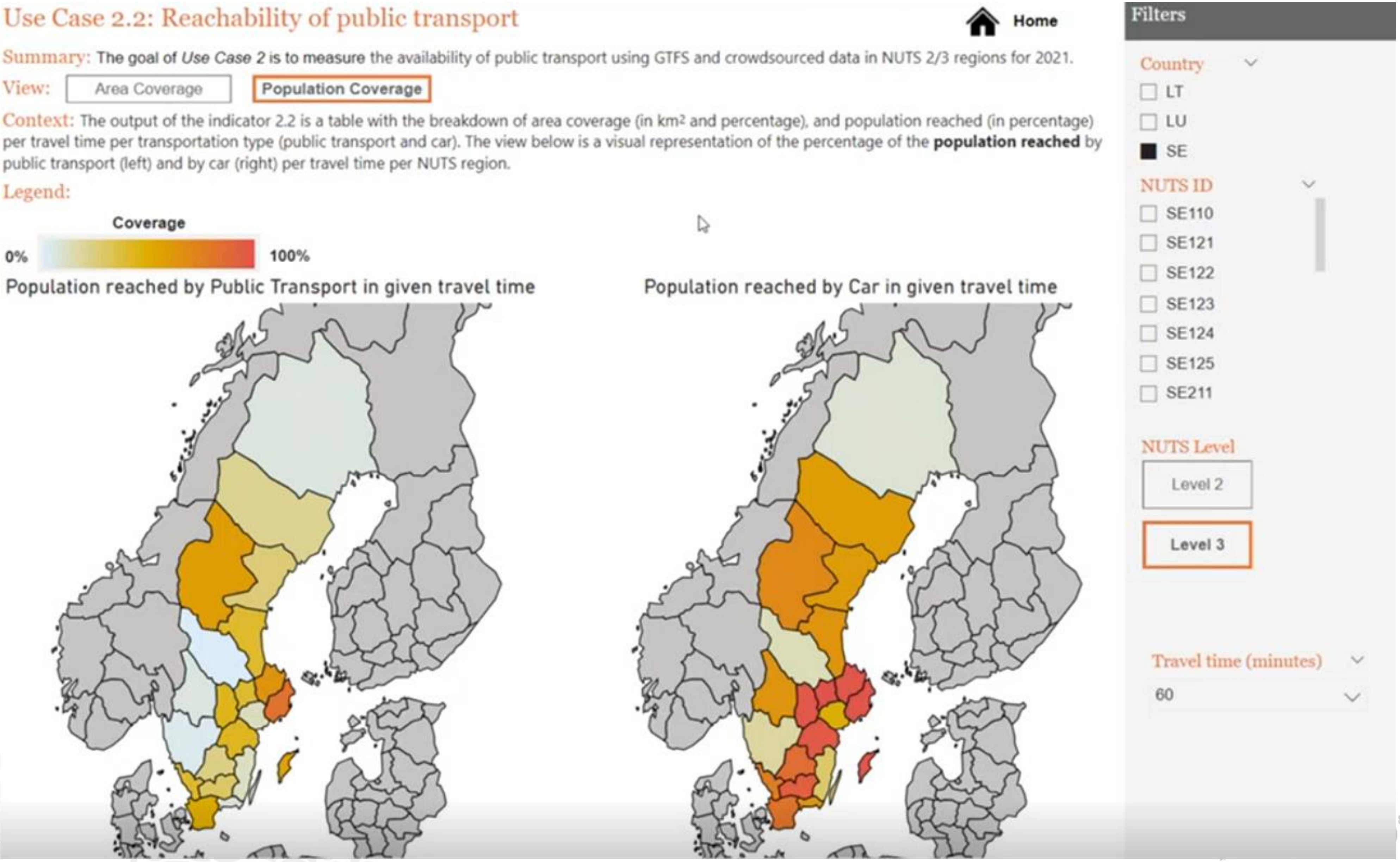
Source: Preliminary project results in Power BI using ArcGIS based on GTFS data



Source: Preliminary project results in Power BI using ArcGIS based on GTFS data



Source: Preliminary project results in Power BI using ArcGIS and Open Street Map based on GTFS data



Source: Preliminary project results in Power BI using ArcGIS and Open Street Map based on GTFS data

Considerations

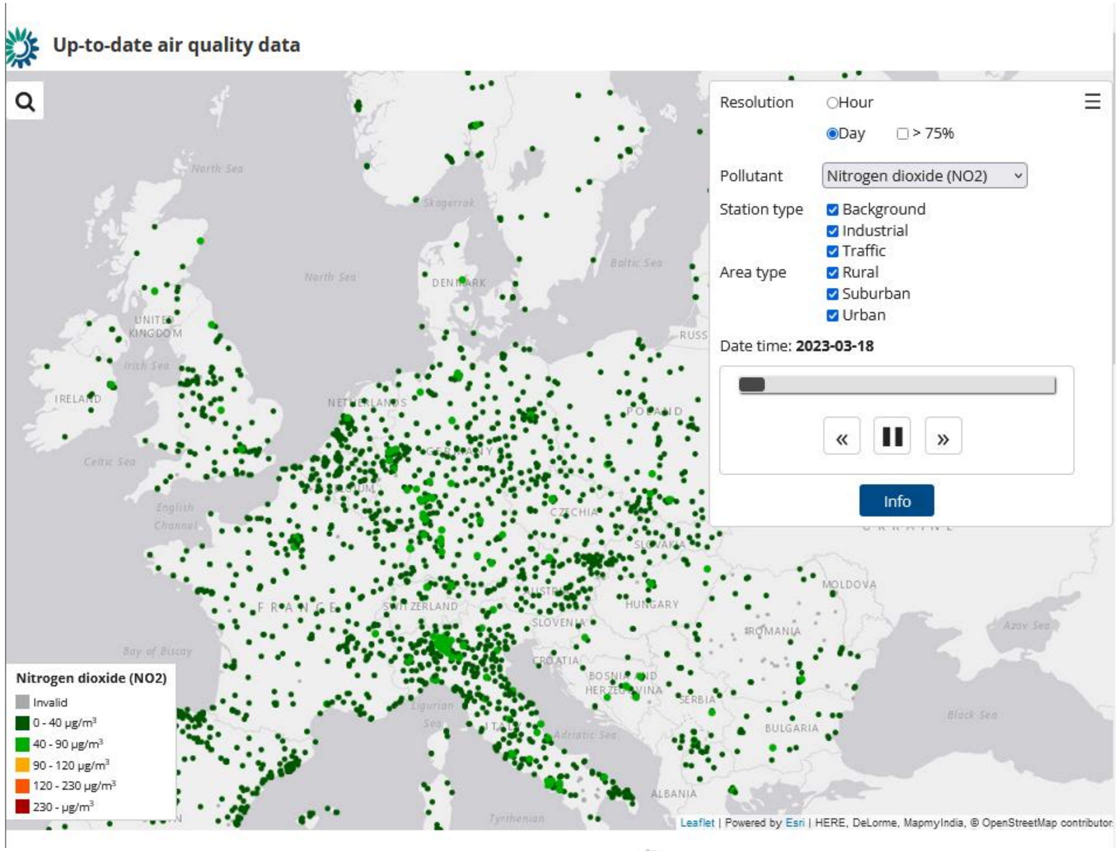
- GTFS data available for most, but not all MS
- Available GTFS data displays differing use of formats -> extra data treatment necessary
- Limitation: The determined origin point might require a “walk” to the nearest public transportation stop and a certain waiting time there for the public transport.
  This might result in XX minutes “spend” on a potentially inconsequential distance.

Areas for future development

- Indicators currently piloted for three MS (LT, LU, SE)
- Results are very sensitive to
  - the origin point chosen (geometric center of the population center grid cell)
  - the origin time chosen (10am weekday)

3. Traffic and Air Quality

	Indicators
3.1	Average concentration of air pollutants (e.g. PM, NO_2) at rush hours and the difference from a baseline value
3.2	Number of kilometres with traffic and high air pollutant concentrations



Source: EEA's Up-to-date air quality data

3. Traffic and Air Quality

Data sources

EEA's hourly Air Quality data

e.g. NO2

(open source)

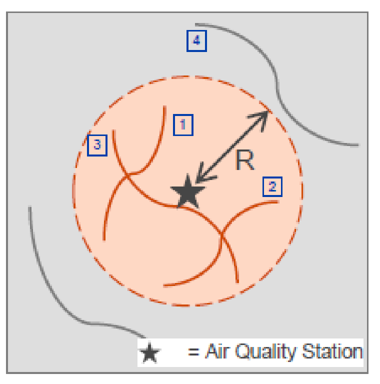
TomTom data

(proprietary)

NUTS raw data

Calculation

Calculate monthly baseline for average air pollutant concentrations per station



Identify roads around air quality station (R=100m)

Identify traffic on those roads (deviation from free flow <= 70%)

Calculate average air pollutant concentration during traffic and compute the difference to baseline

Output

Use Case 3.1: Average air quality during traffic

Home

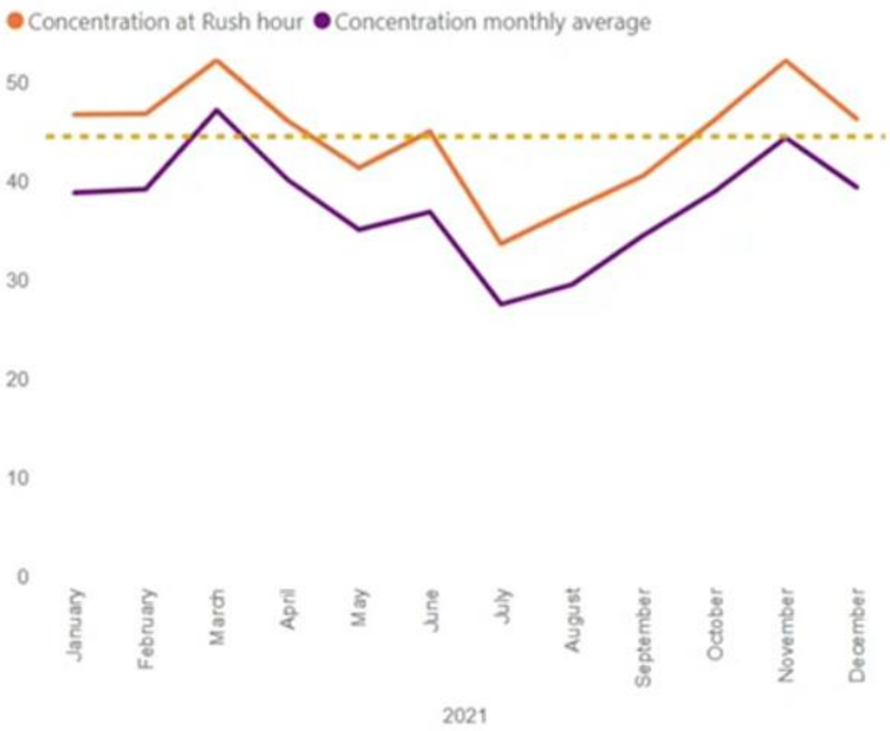
Summary: The goal of *Use Case 3* is to measure the average concentration of selected air pollutants at peak traffic times and their variation based on the European Environment Agency air quality database and TomTom traffic data.

Context: The output of the indicator 3.1 is a table with the concentration at rush hour, the monthly concentration baseline, and the difference between these two figures per air pollutant, per day, and per air quality station ID for a given year and country. The view below provides an additional visual representation of the evolution by day of the difference between the concentration at rush hour and the monthly **average concentration per air pollutant** per air quality station ID.

Date

01/01/2021 31/12/2021

Concentration by Day



Monthly comparison of Air Pollutant Concentration

Month	Average Concentration at Rush hour	Concentration monthly average
January	46.74	38.82
February	46.79	39.17
March	52.21	47.17
April	46.08	40.14
May	41.33	35.09
June	45.02	36.89
July	33.66	27.55
August	37.17	29.57
September	40.55	34.52
October	46.16	38.93
November	52.17	44.38
December	46.29	39.37
Total	**44.48**	**37.61**

Filters

Air Pollutant

- C6H6
- NO2
- SO2

Sampling Station ID

SPO-BETB001_00008_100

Location of Sampling Station

2024 MEXICO CONFERENCE

Source: Preliminary project results in Power BI using ArcGIS based on EEA data and TomTom Speedprofiles

eurostat

Use Case 3.1: Average air quality during traffic

Home

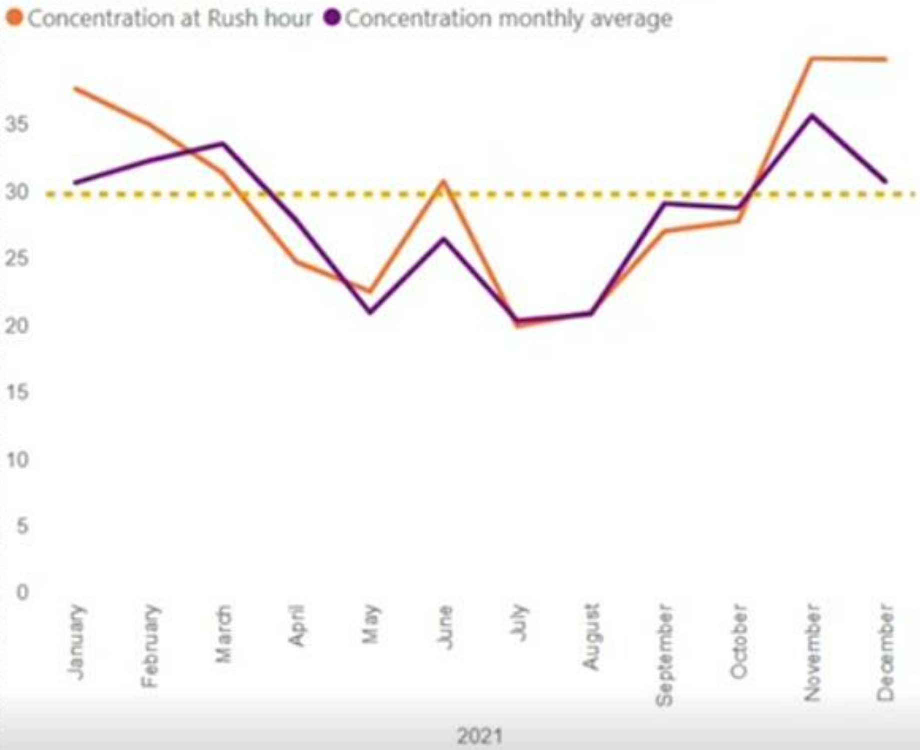
Summary: The goal of *Use Case 3* is to measure the average concentration of selected air pollutants at peak traffic times and their variation based on the European Environment Agency air quality database and TomTom traffic data.

Context: The output of the indicator 3.1 is a table with the concentration at rush hour, the monthly concentration baseline, and the difference between these two figures per air pollutant, per day, and per air quality station ID for a given year and country. The view below provides an additional visual representation of the evolution by day of the difference between the concentration at rush hour and the monthly **average concentration per air pollutant** per air quality station ID.

Date

01/01/2021 31/12/2021

Concentration by Day



Monthly comparison of Air Pollutant Concentration

Month	Average Concentration at Rush hour	Concentration monthly average
January	37.63	30.61
February	34.97	32.25
March	31.32	33.52
April	24.69	27.76
May	22.48	20.91
June	30.72	26.41
July	19.91	20.31
August	20.93	20.79
September	26.99	29.06
October	27.73	28.72
November	39.92	35.63
December	39.83	30.70
Total	**29.71**	**28.05**

Filters

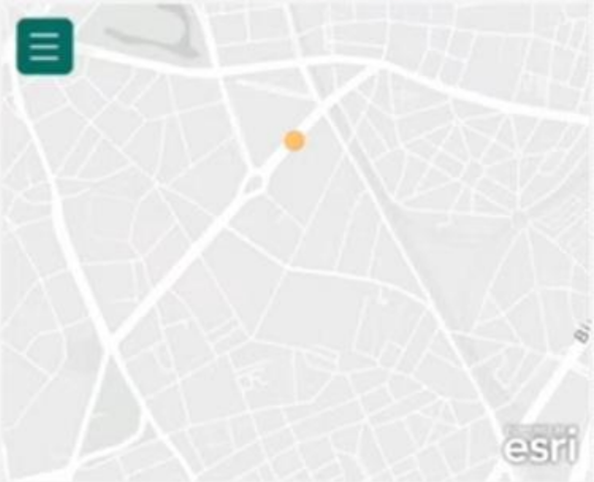
Air Pollutant

- C6H6
- NO2 (selected)
- SO2

Sampling Station ID

SPO-BETR805_00008_100

Location of Sampling Station



Source: Preliminary project results in Power BI using ArcGIS based on EEA data and TomTom Speedprofiles

Considerations

- Considerable missing data in EEA data set
  -> completeness threshold: 80% for computation of monthly average
- Number of traffic air stations rather limited (11 for BE)
  -> aggregation to e.g. city level currently not expedient

Areas for future development

- Indicators currently piloted for one MS (BE)
- Indicators piloted for year 2021

–> interpretation complicated given the unusual mobility patterns during the pandemic

Challenges & lessons learnt

- Commercial data set have better and richer content than free data sets, but they generate financial costs
- Public data is also not always easy to get, administrative agreements are time consuming
- Results are depended on harmonized input data
- Benchmarking of results and transparency of methods is key

- **Next steps:** –> scale-up
  –> present selection of indicators to METAC
  –> publish experimental statistics

Thank you

Further information:

Evangelia FORD-ALEXANDRAKI Evangelia.Ford-Alexandraki@ec.europa.eu
Matyas MESZAROS Matyas.Meszaros@ec.europa.eu
Miriam BLUMERS Miriam.Blumers@ec.europa.eu
Nikolaos ROUBANIS Nikolaos.Roubanis@ec.europa.eu

Early estimates of maritime traffic using innovative data sources

- Eurostat -

Early estimates of maritime traffic

- Objective: Use Eurostat quarterly statistics and EMSA data for improving timeliness of maritime statistics – publish port calls few weeks after a reference quarter instead of a year later

- Eurostat-EMSA cooperation started in February 2023, with a formal agreement (MoU) to provide AIS and other administrative data to Eurostat for 'Early estimates of maritime traffic'

- National experiences (IRL, DK, GR, NL) were the starting point for developing a way to assess comparability of the two data sources and estimate port calls at EU level

Project results: EMSA data

EMSA provided anonymised microdata from 3 sources:

- SafeSeaNet port calls – vessel traffic monitoring system linking maritime authorities across Europe

arrival_year	arrival_month	portofcall					
2018	12	GRPIR	12	SSN	A36A2PR	A36	GR
2018	12	GRPIR	12	SSN	A36A2PR	A36	GR
2018	12	GRPIR	12	SSN	A36A2PR	A36	GR
2018	12	GRPIR	12	SSN	A36A2PR	A36	GR
2018	12	GRPIR	12	SSN	A36A2PR	A36	GR
2018	12	GRPIR	12	SSN	A36A2PR	A36	GR
2018	12	GRPIR	12	SSN	A36A2PR	A36	GR

- MARINFO vessel information data –commercial AIS data

arrival_year	arrival_month	portofcall	ship_			
2017	5	GRPIR	1871	MARINFO	A36A2PR	A36
2017	5	GRPIR	1871	MARINFO	A36A2PR	A36
2017	5	GRPIR	1871	MARINFO	A36A2PR	A36
2017	5	GRPIR	1871	MARINFO	A36A2PR	A36
2017	5	GRPIR	1871	MARINFO	A36A2PR	A36
2017	5	GRPIR	1871	MARINFO	A36A2PR	A36
2017	5	GRPIR	1871	MARINFO	A36A2PR	A36

- EMSA detected port call data – AIS signals

arrival_year	arrival_month	portofcall					
2022	7	NOLAN	4	DPC	A36A2PR	A36	NO
2022	7	NOLAN	4	DPC	A36A2PR	A36	NO
2022	7	NOLAN	4	DPC	A36A2PR	A36	NO
2022	7	NOLAN	4	DPC	A36A2PR	A36	NO
2022	7	NOLAN	4	DPC	A36A2PR	A36	NO
2022	7	NOLAN	4	DPC	A36A2PR	A36	NO
2022	7	NOLAN	4	DPC	A36A2PR	A36	NO

Project results: Eurostat data

Directive 2009/42/EC of the European Parliament and of the Council of 6 May 2009 on statistical returns in respect of carriage of goods and passengers by sea (recast):

- **Quarterly European port vessel traffic** in main European ports, by port, type and size of vessels loading or unloading cargo, embarking or disembarking passengers (including cruise passengers on cruise passenger excursion)

- Comparison for the years 2015-2019

Data set F2: European port vessel traffic in the main European ports, by port, type and size of vessels loading or unloading cargo, embarking or disembarking passengers (including cruise passengers on cruise passenger excursion)

Periodicity: quarterly

	Variables	Coding detail	Nomenclature
Dimensions	Data set	Two-character alphanumeric	F2
	Reference year	Four-character alphanumeric	(e.g. 1997)
	Reference quarter	One-character alphanumeric	(1, 2, 3, 4)
	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, Annex VI
	Size of vessel GT	Two-character alphanumeric	Gross tonnage size classes, Annex VII

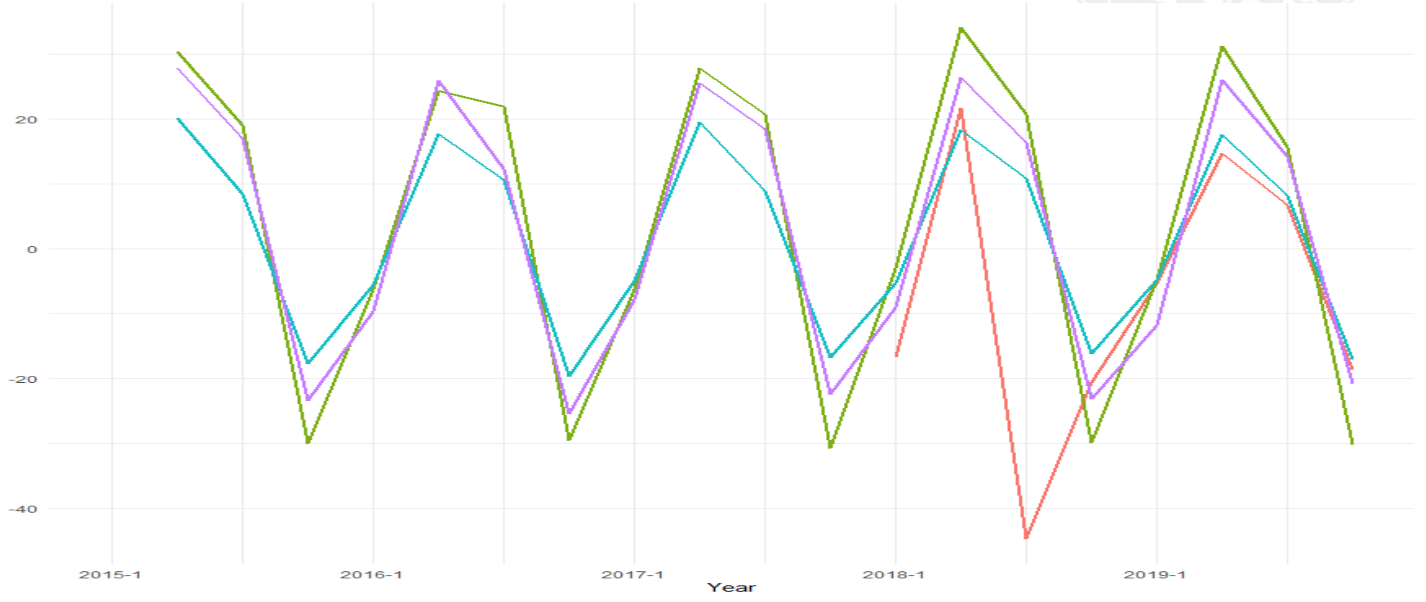
Data: Number of vessels;
Gross tonnage of vessels.

Project results: Comparison

Comparison at EU level:

Strong similarity of trends between the EMSA-SSN and EMSA-MARINFO and Eurostat dataset - allows proceeding with their use for estimations of Eurostat F2 data at EU level.

% Variation of port calls over time



Data source:

- Eurostat statistics
- EMSA AIS detected port calls
- EMSA MARINFO - AIS
- EMSA SafeSeaNet - AIS

IOS-ISI 2024 MEXICO CONFERENCE

eurostat

Project results: Comparison

Comparison of the number of vessel's calls by type of vessel at port and country level showed differences for specific years and vessel types

Potential reasons for differences:

- Classification by type of vessel
- Definition of statistical ports
- Scheduled traffic between two ports
- Activity of the vessel
- Exemption in reporting to EMSA

EMSA datasets SafeSeaNet (SSN) port calls and MARINFO datasets most useful - detected AIS port calls dataset is of limited use for the time being

Project results: Modelling

Comparison of two estimation methods in order to select the most reliable one

1) Multiple Linear Regression
2) Auto-Regressive Integrated Moving Average with Exogenous variables (ARIMAX)

First method yielded differences in annual totals ranging from-7,4% to 0,5% for EU level 2015-2019 and much larger differences at disaggregated level by type of vessel.
Further limitations include: the model assumes linearity and may not capture all interactions between predictors.

Project results: Modelling

Comparison at EU level for:

Auto-Regressive Integrated Moving Average with Exogenous variables (ARIMAX)

Uses Eurostat data for previous quarters and data from both SSN and MARINFO for the quarter to be estimated (exogenous variables)

Annual deviation per type of vessel - ARIMAX					
	2015	2016	2017	2018	2019
total	5.3%	0.5%	-1.3%	-1.8%	0.5%
10	5.8%	-3.3%	0.2%	-1.2%	-2.3%
20	0.2%	1.7%	-4.2%	-3.2%	3.0%
31	-4.2%	-1.3%	-3.6%	1.1%	-0.8%
32	-2.8%	-0.9%	-5.4%	-0.9%	0.0%
33	6.9%	1.2%	-2.2%	-3.2%	0.4%
35	1.3%	-0.2%	2.6%	3.5%	1.1%
36	-7.9%	-16.9%	14.2%	-0.8%	5.4%

Type of vessels
10 Liquid bulk
20 Dry bulk
31 Container
32 Specialised
33 General cargo, non-specialised
35 Passenger
36 Cruise Passenger

Next steps

Publication of the 2015-2019 results as experimental statistics: EU estimates by vessel type

Drafting an agreement with EMSA on the regular provision of the data needed to produce the estimates

Automation of the EMSA data aggregation process

Publication of the estimates on regular basis in Eurobase

Possible further work:

Improve the classification of vessels by vessel type to produce more reliable data at country level

Thank you

Further information:

Boryana MILUSHEVA Boryana.Milusheva@ec.europa.eu
Nikolaos ROUBANIS Nikolaos.Roubanis@ec.europa.eu

IAOS-ISI 2024
MEXICO CONFERENCE

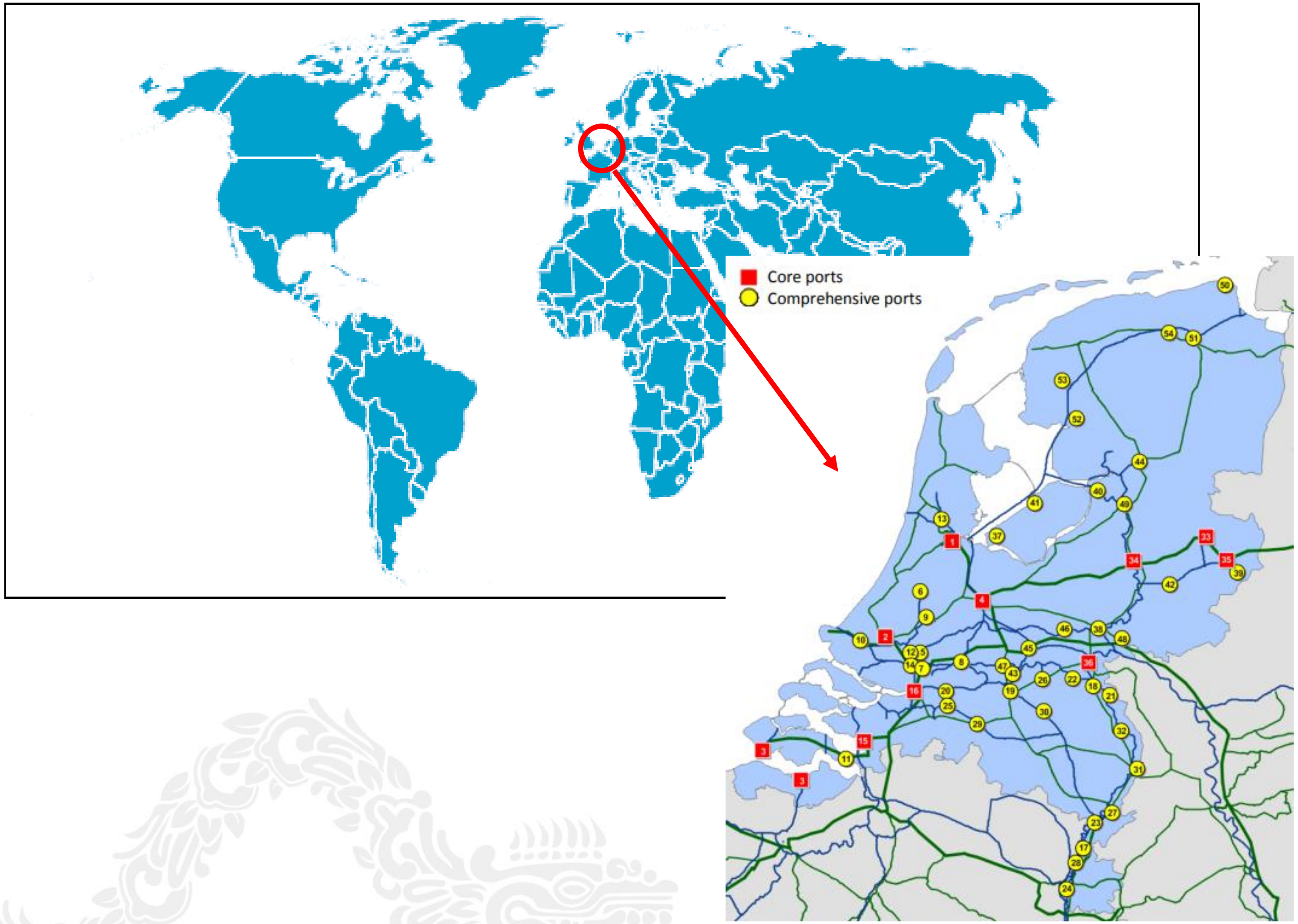
Using new and innovative data sources for transport statistics and indicators

- Statistics Netherlands -

INEGI
International Statistical Institute
IAOS IMPROVING OFFICIAL STATISTICS

Statistics Netherlands

Statistics Netherlands



Secondary data collection

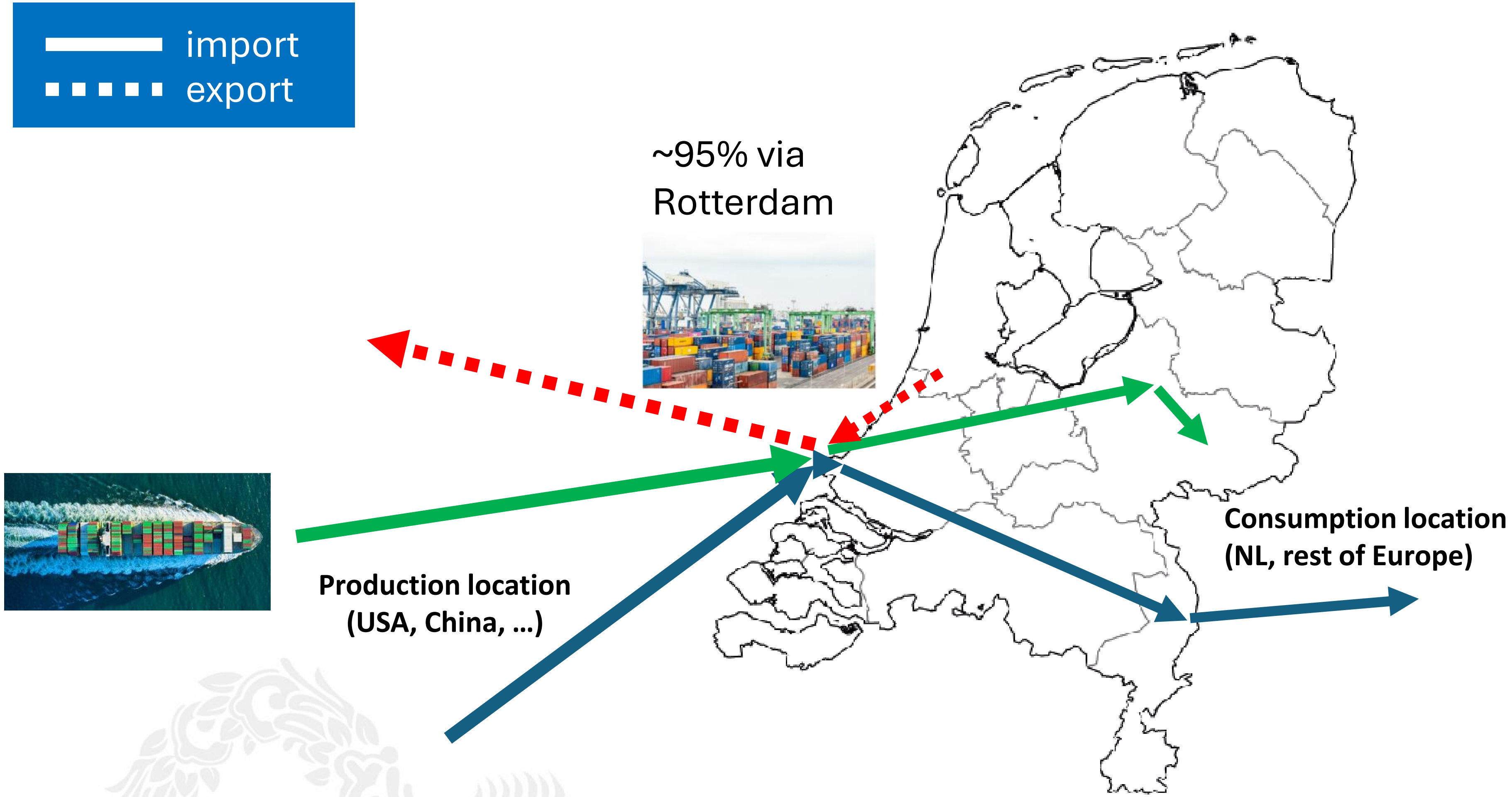
- Registers
- Backbones:
    - Population register
    - Company register
- Big data

Primary data collection

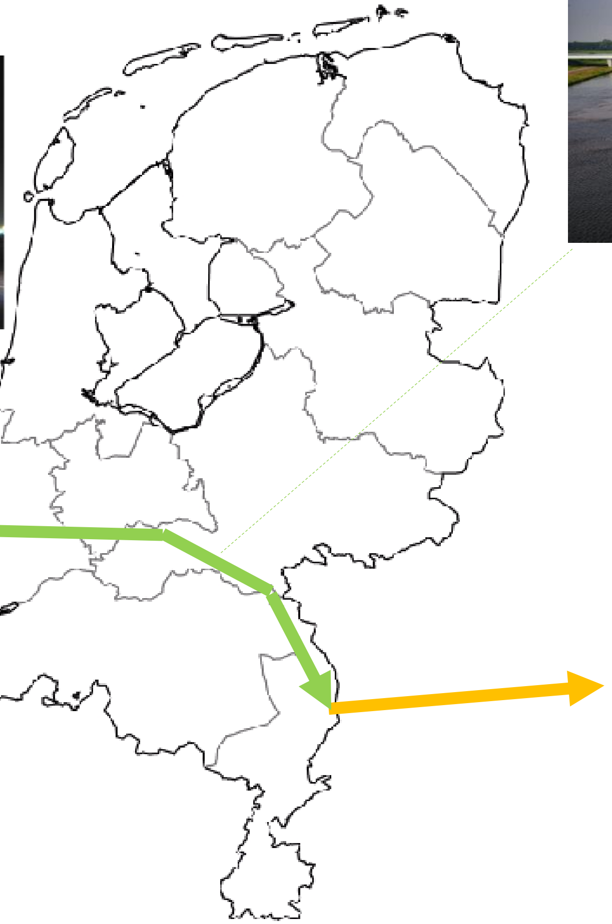
- Only when necessary

INTRODUCTION

Container transport



Multimodal container chains

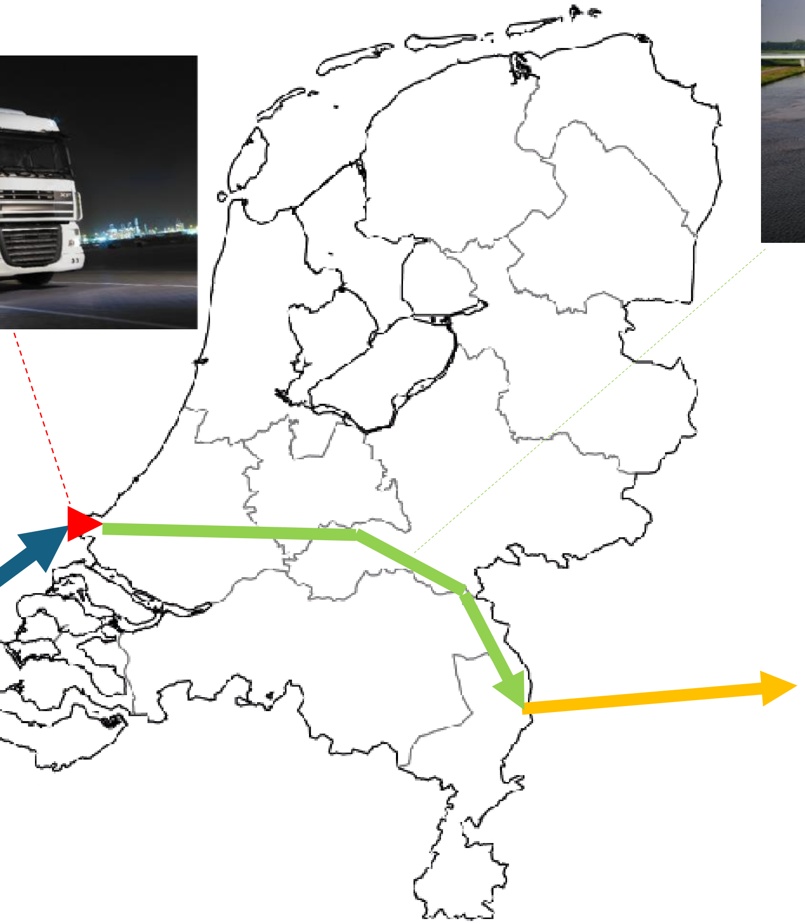


Multimodal container chains

Road transport
sample survey

IVS Next
(AIS)

ICS
ECS
Customs



ProRail (rail infra)
Rail operator

StatLine mono-modal container transport




European mandatory statistics

Commodity flows	Transport modes	Periods	Estimated gross weight transported goods Gross weight Estimate Total	 Containerized goods	 Non-containerized goods	 Share in commodity flow Total	 Containerized goods	 Non-containerized goods
			1000 kg			%		
Imports; total	Maritime transport	2022*	203,972,649	25,046,294	178,926,354	57	7	50
	Inland shipping	2022*	51,478,933	4,377,241	47,101,691	14	1	13
	Road transport	2022*	49,878,641	1,682,680	48,195,961	14	0	14
	Rail transport	2022*	4,964,637	3,359,346	1,605,291	1	1	0
Exports; total	Maritime transport	2022*	93,850,100	25,802,976	68,047,123	33	9	24
	Inland shipping	2022*	78,873,207	5,560,280	73,312,927	28	2	26
	Road transport	2022*	36,505,920	1,799,063	34,706,857	13	1	12
	Rail transport	2022*	7,516,576	4,081,287	3,435,289	3	1	1

Source: CBS

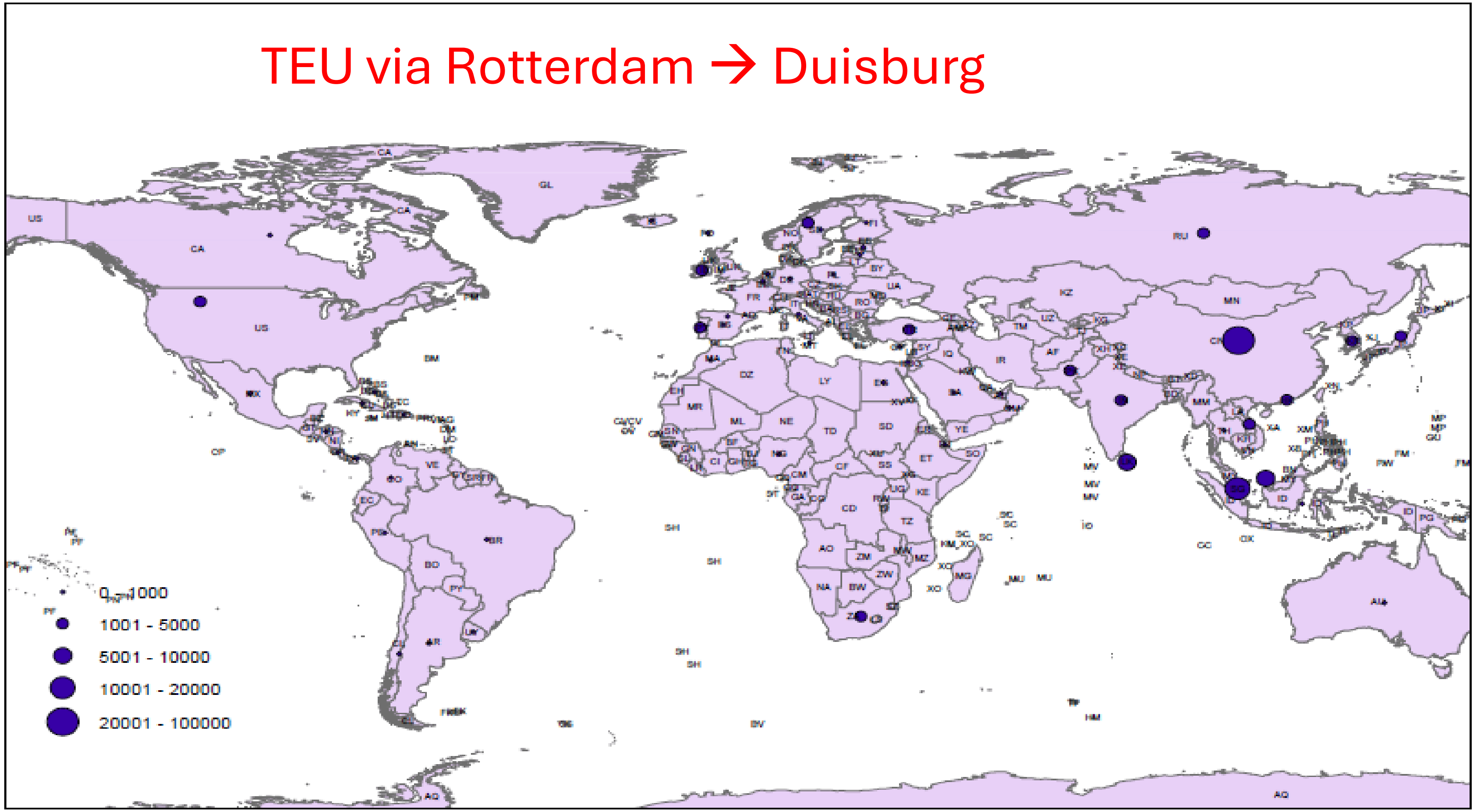
IOS-ISI 2024 MEXICO CONFERENCE

 8

Policy questions / motivation

- **Infrastructure policy substantiation** (Rijkswaterstaat - BasGoed)
- **Monitoring modal shift containers** (European Green Deal - sustainable transportation)
- **Improve transport statistics** at Statistics Netherlands:
  - Better and more detailed information about goods transported in containers;
  - Improve mono-modal statistics
  - First step towards **ultimate goal**: integrate all modalities in 1 statistic

Possible (additional) output



MEXICO CONFERENCE

History container project

2018 Eurostat grant constructing container chains:
Combine registrations and collected data sources at micro level
→ *Too many gaps*

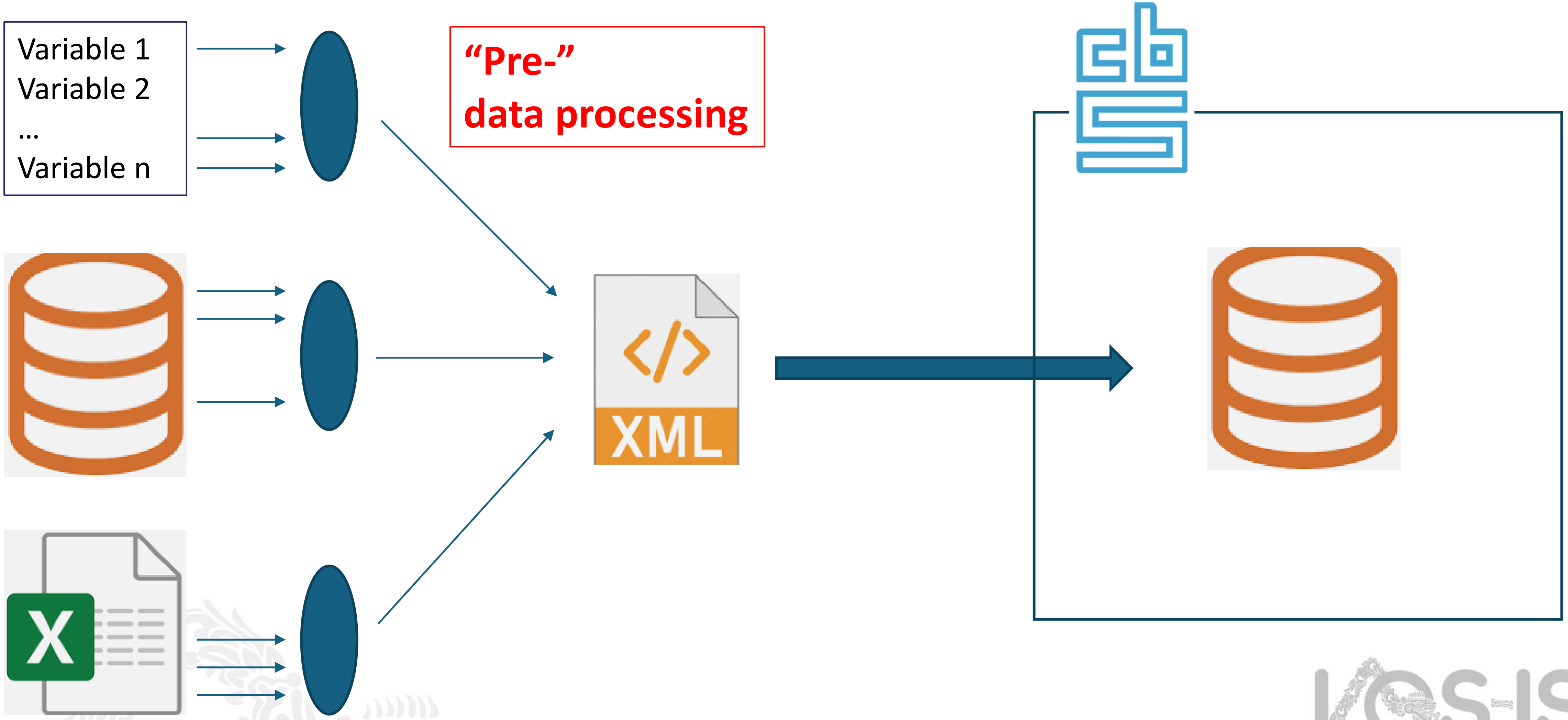
2020 Pilot adding private data:
Add 10 private data sources from different modalities in open format
→ *Data collection and preparation feasible; added value!*

2022 Scale up private data:
Integral data collection of sea, inland terminals and rail operators;
Develop sampling design for road transport (~2500 companies)
→ *A lot of convincing and patience necessary*

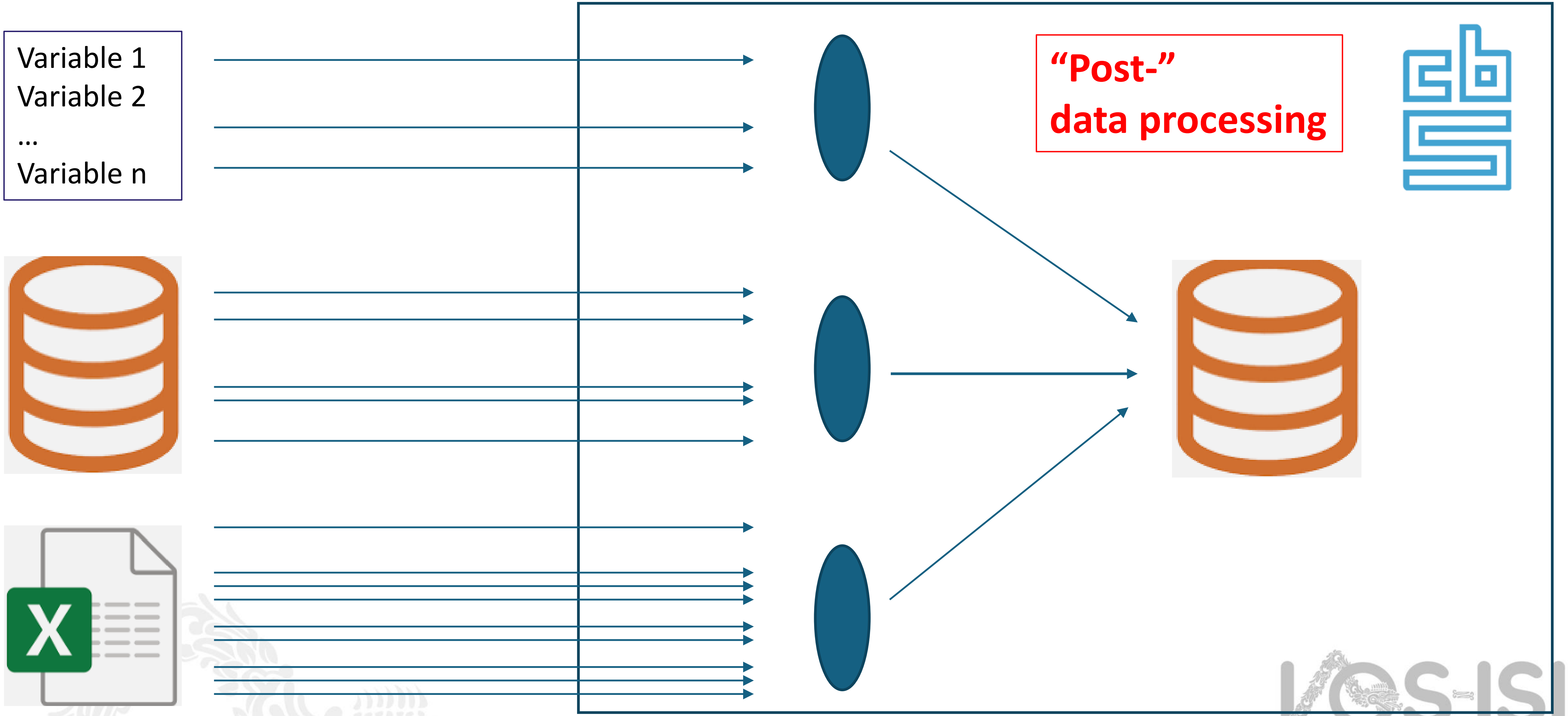
Innovative data collection

- Private data
  - voluntary
- Open data format
  - Extract from traffic management systems
- Different formats
  - JSON
  - XML
  - Excel / csv
  - API
  - PDF

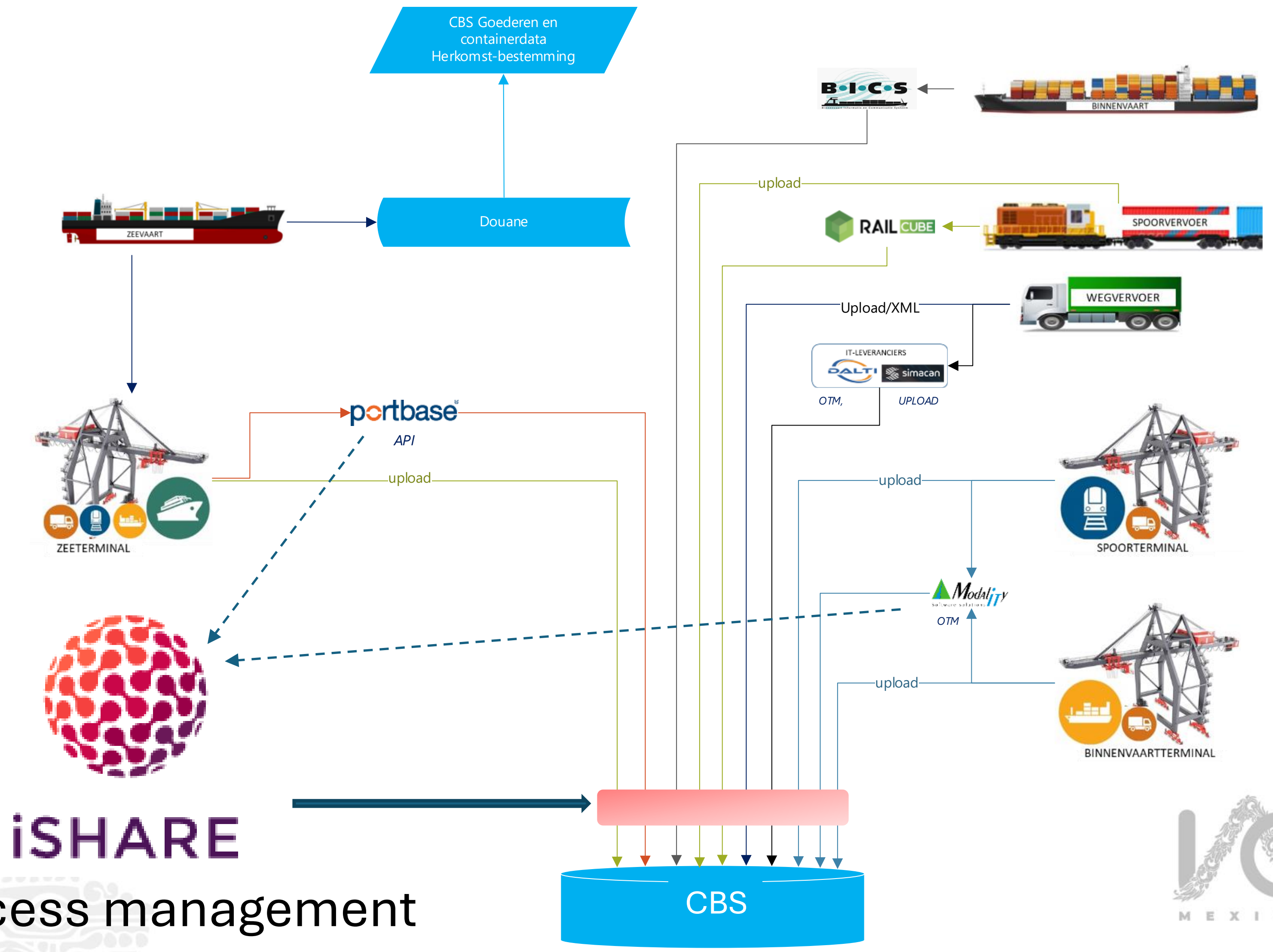
Peprocessing data to fixed format



Free format with post-data processing



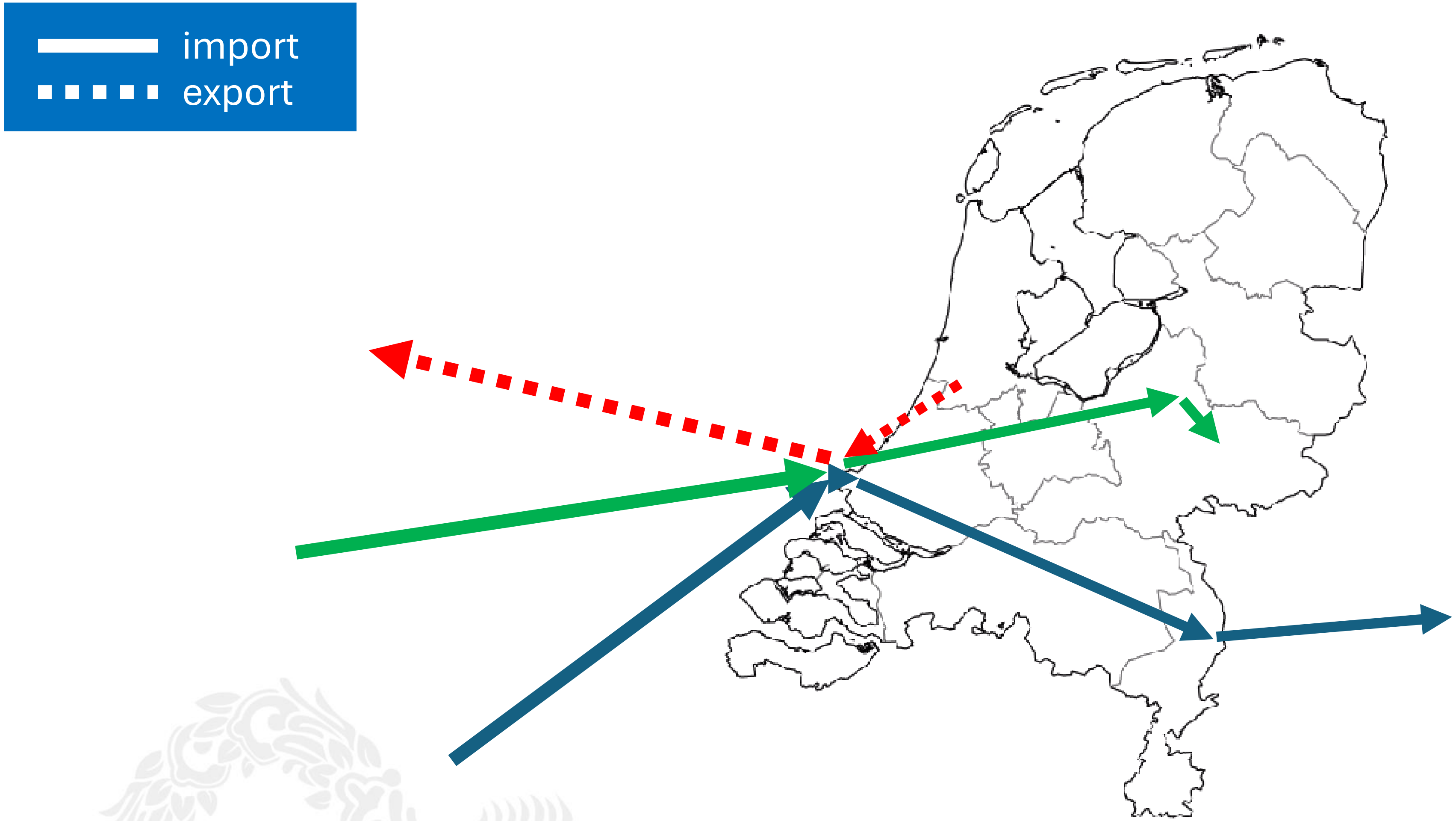
Data Collection infrastructure

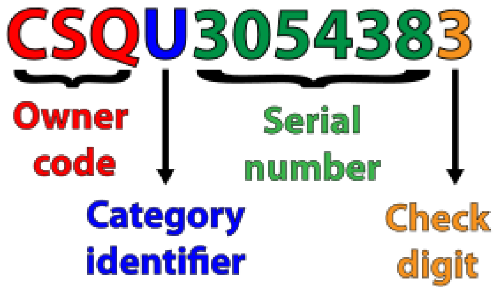


Data access management

DATA PROCESSING

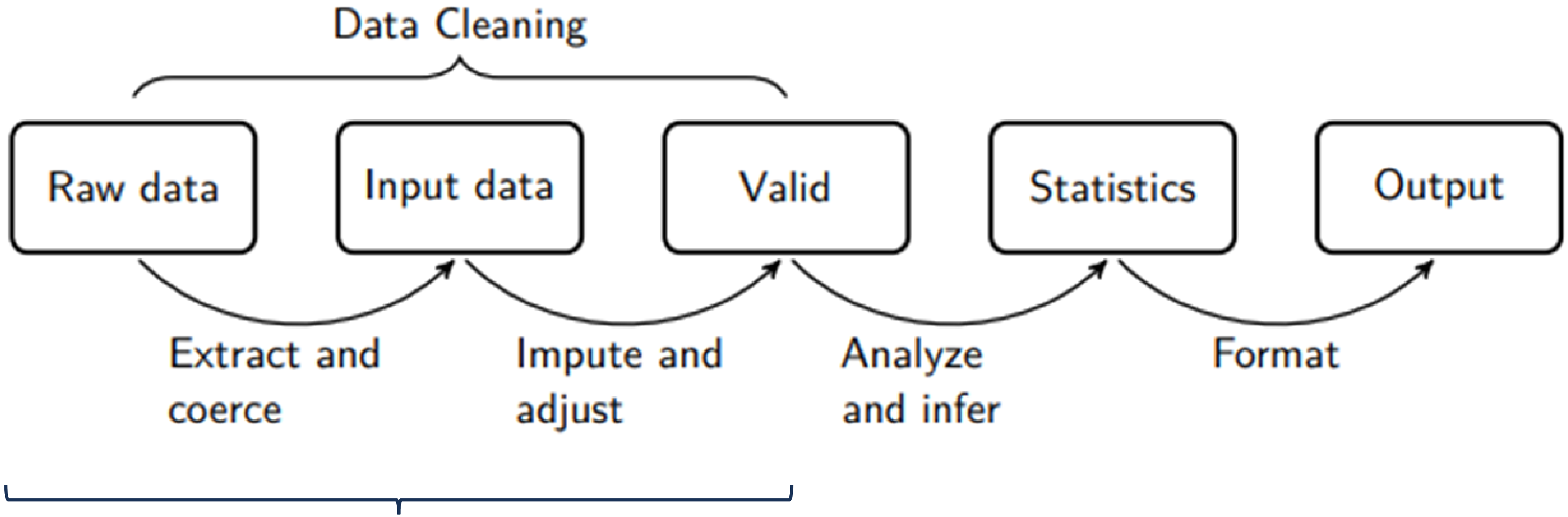
Container number essential





→ Unique 'tracking' id

Statistical value chain

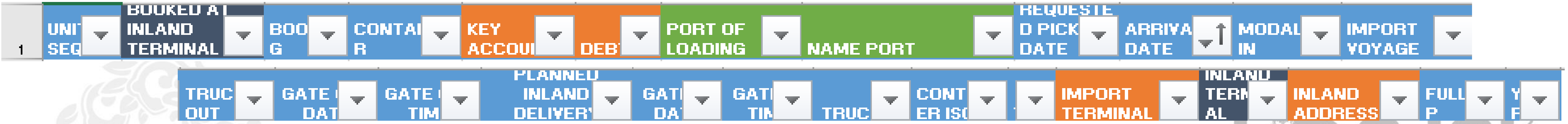


- Harmonize data formats
- Standardize variables

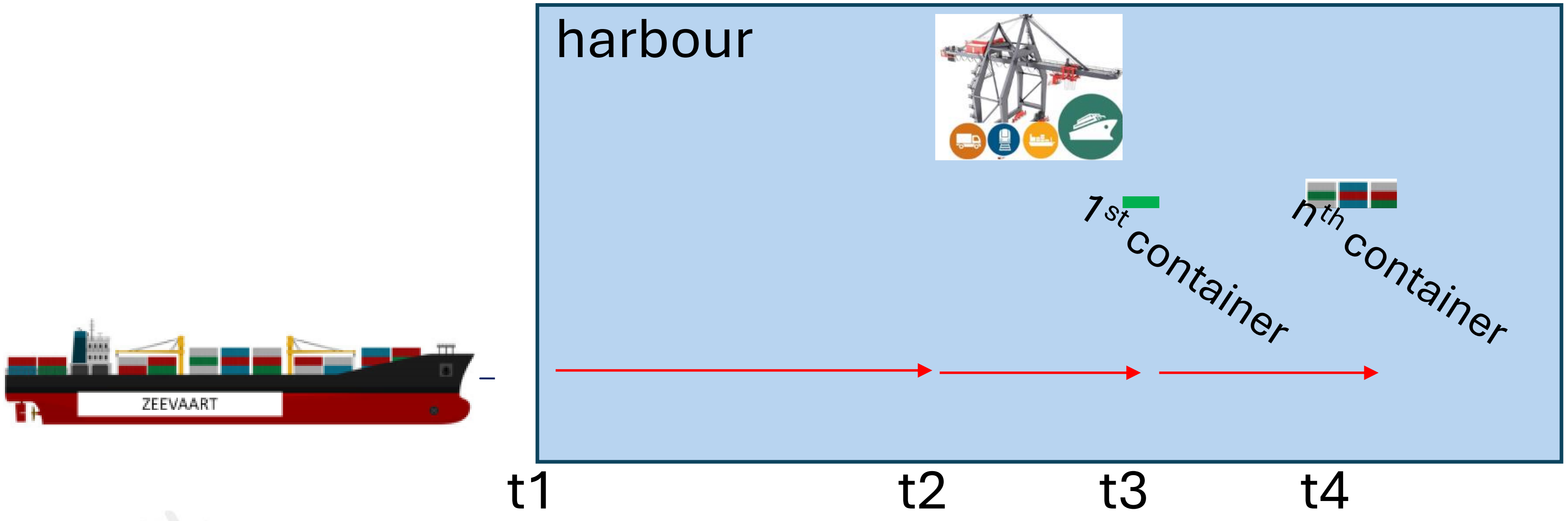
Challenges with input data

- Data from carriers and from terminals
- Different input data formats
- Different variables/columns
- Different level of detail

```
1    terminal,equipment_number,size,timestamp
```



Different definitions



actual time of arrival?

Standardizations

- Standardize all variables, column names and data types.

- Locations:
  - Use geo-coding software to get lat/lon and UN/LO codes.
  - Use routing software to get the travelled distance.

Input location	Standardized
Mexico city	coordinates = [-99.07, 19.43] UN/LO = MXMEX

Standardizations (2)

- Good descriptions:
  - Use text classification with cosine similarity to get NST2007 classification.

Raw text	Cleaned text	Classification
1764 CARTONS PALLETIZED WITH 26460 KG NET WEIGHT OF FROZEN HALF CHICKEN BREAST BONELESS SKINLESS WITHOUT INNERFILLET SALTED	FROZEN HALF CHICKEN BREAST BONELESS SKINLESS WITHOUT INNERFILLET SALTED	NST2007 = 04.1 (Meat, raw hides and skins and meat products)

How to create container chains?

- A chain starts abroad (maritime/rail) and ends with road transport to the consumption location in NL (or vice versa).

- Merge all different standardized data sources together.

- Select one container number and sort by date and time:
  - If two actions are consecutive and both in the data, then this is part of the chain.
  - Missing actions can sometimes be imputed:
    - Within X hours
    - And use same UN/LO code

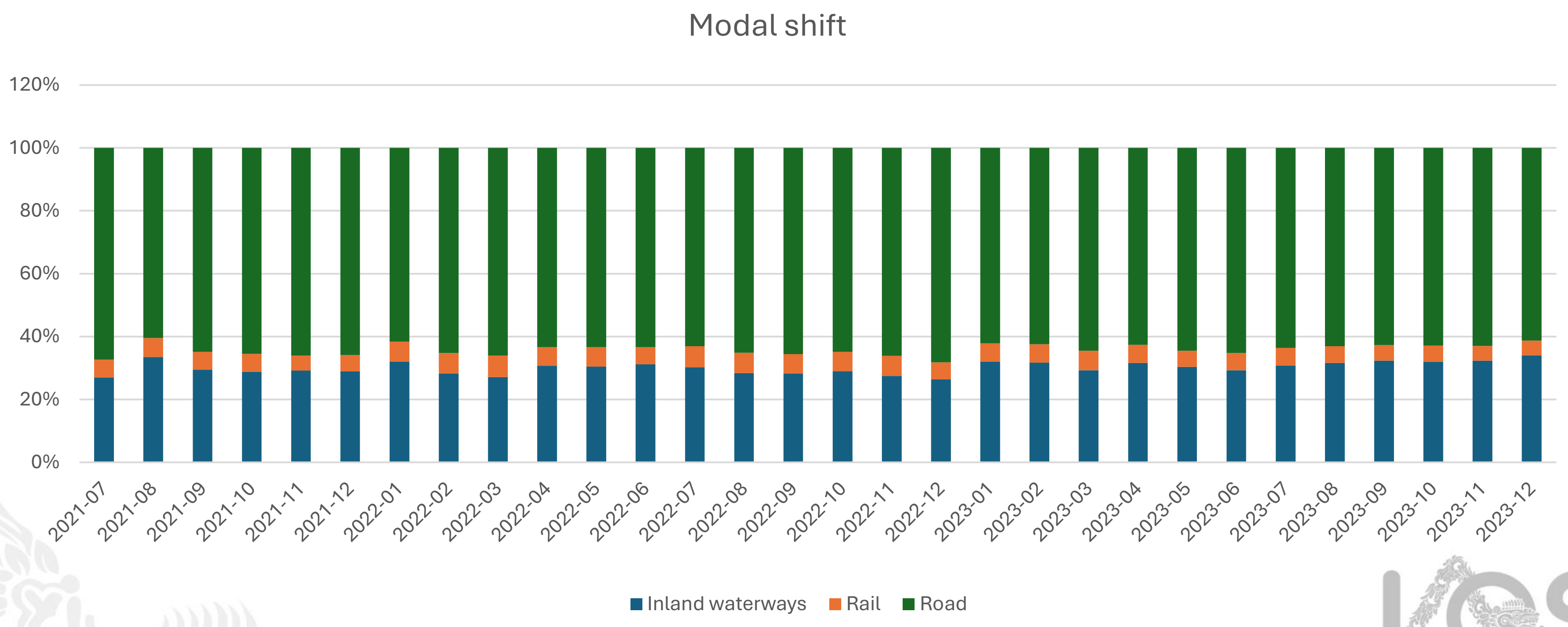
Challenges with output data

- What to do with inconsistencies between two data sources?

- How to create an estimation method?
  - Correct for missing input data
  - Impute missing parts of the chain
  - Using statistics per transportation mode as total
  - Work in progress

First experimental results

- Modal shift for containers per month from the port of Rotterdam to the inland



Contact details

- For questions contact our researchers:
    - transport@cbs.nl

- More information: See attached document

Thank you

IAOS-ISI 2024
MEXICO CONFERENCE

Experience of Statistics Poland in the production of experimental statistics on road and maritime transport using innovative data sources

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS

Experience of Statistics Poland in the production of experimental statistics on road and maritime transport using innovative data sources

Michał Bis Statistics Poland

IAOS-ISI 2024 MEXICO CONFERENCE

Agenda

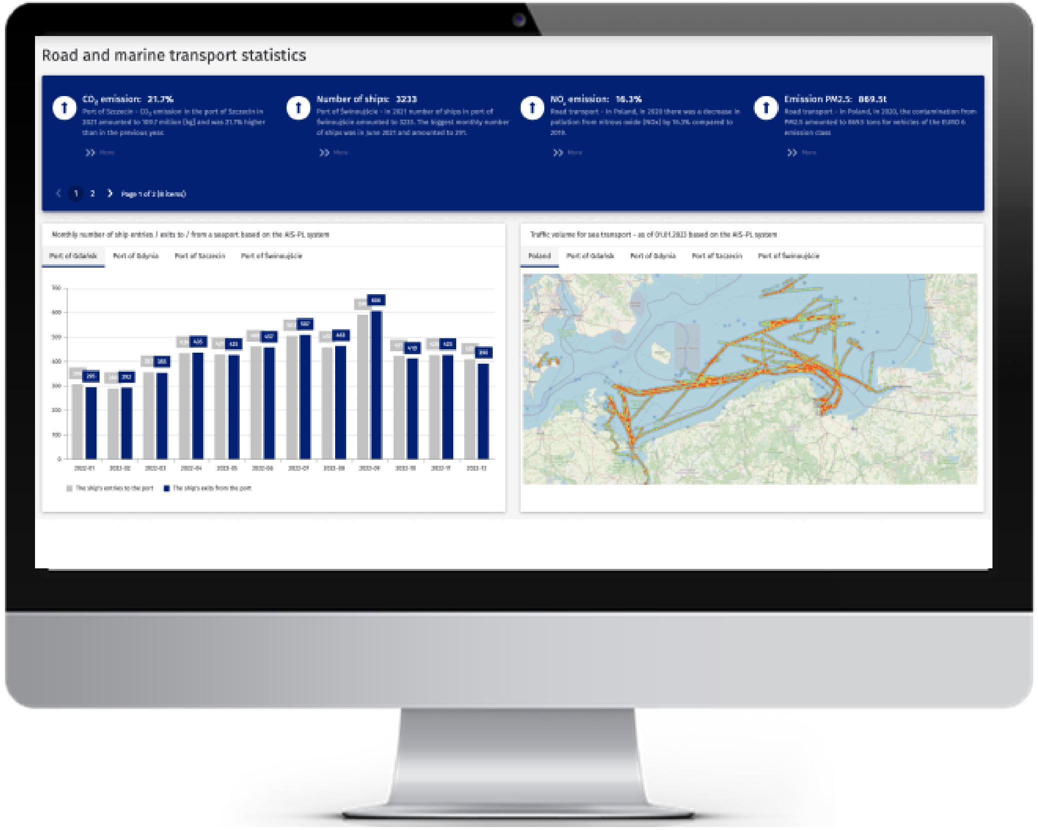
- TranStat – overview, assumptions, architecture, data flow and processing
- AIS - Automatic Identification System in a nutschell
- Statistics of traffic intensity, transportation volume in maritime transport – assumptions, results
- e-TOLL – electronic toll collection system in a nutschell
- Statistics of traffic in road transport – assumptions, results
- Conclusions

TranStat - overview

https://transtat.stat.gov.pl

The project focused on:

- obtaining access to sensory data from the Automatic Identification System (AIS) and the e-TOLL electronic toll collection system;
- adaptation of modern Big Data methods and tools;
- development a methodology for estimating traffic intensity, transportation volume and the amount of emissions;
- implementation of experimental statistics in domain of road and maritime transport;
- lower costs by using non-statistical sources;
- speeding up the publication of statistics.



Project term: 2019 - 2021

Consortium: Statistics Poland, Maritime University of Szczecin, Cracow University of Technology

TranStat - assumptions

The general requirements for modern IT systems, including:

- implementation of open standards;
- technological neutrality (vendor lock-in);
- compliance with applicable laws;
- modular construction;
- easy expansion with new system functionalities in the future;
- ensuring an appropriate level of security

and **the requirements for scalable Big Data solutions**:

- model for the 3 V's of big data: volume, velocity and variety.

TranStat - architecture

The TranStat IT system has been developed and implemented in the production environment.

The following functional subsystems have been developed as part of the system:

- **Data collection and processing subsystem responsible for the following subprocesses:**
  - decoding AIS data,
  - processing stream data from sensors,
  - integration, validation, transformation and aggregation of data.

- **Data presentation and analysis subsystem - internal**, the purpose of which is to enable data exploration and visualization as well as statistical analyzes using the RStudio and Apache Zeppelin tools.

- **Data presentation and analysis subsystem - external**, intended for an external users, operating on the basis of calculated aggregates and indicators - https://transtat.stat.gov.pl

TranStat - the process of data flow and processing

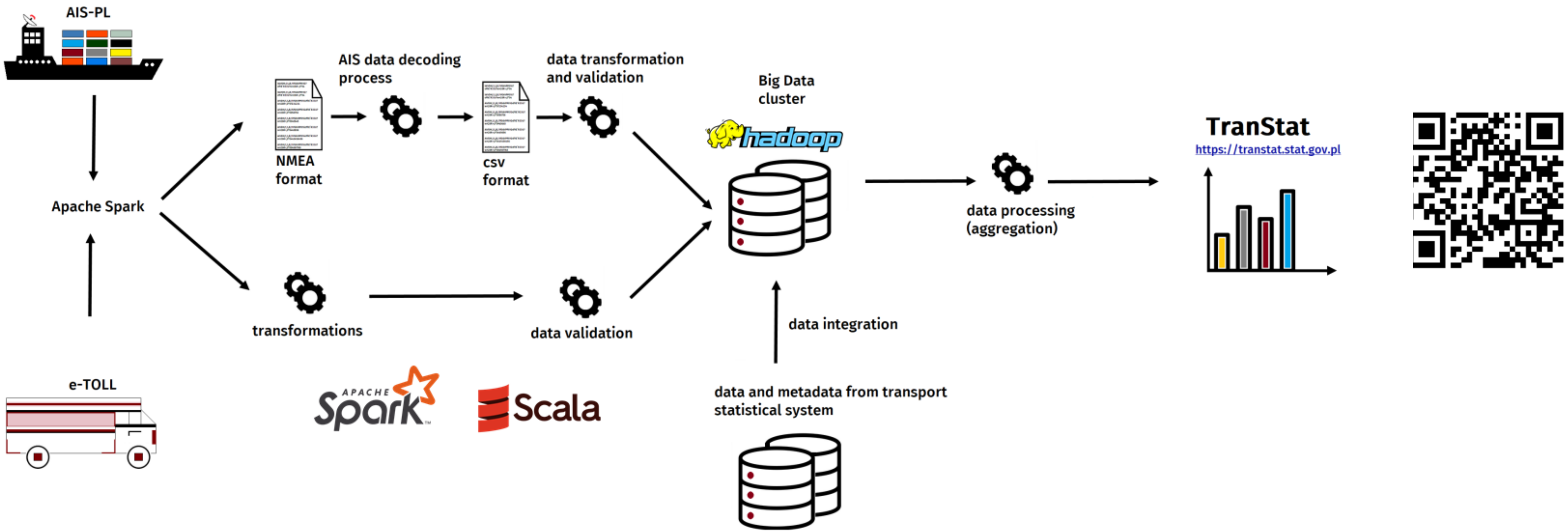


Figure 1. The proces of data flow and processing for TranStat system

Source : own study

Automatic Identification System in a nutschell

Automatic Identification System (AIS) - it is a system of automatic identification used on ships for the electronic exchange of information between nearby ships, AIS base stations and satellites. According to the requirements defined in Chapter V of the SOLAS Convention developed by the IMO, the AIS system should be installed on:

- all ships of 300 gross tonnage and more used in international shipping,
- all ships of 500 gross tonnage and more not used in international shipping,
- all passenger ships, regardless of size.

The basic application of the AIS system:

- strengthening of navigation safety (anti-collision system),
- vessel traffic management support for coastal Vessel Traffic Service (VTS).

Data source availability - legal basis

Regulation of the Minister of Maritime Economy and Inland Navigation of September 26, 2018 on the National System for Monitoring Vessel Movement and Information Transmission.

Automatic Identification System in a nutschell

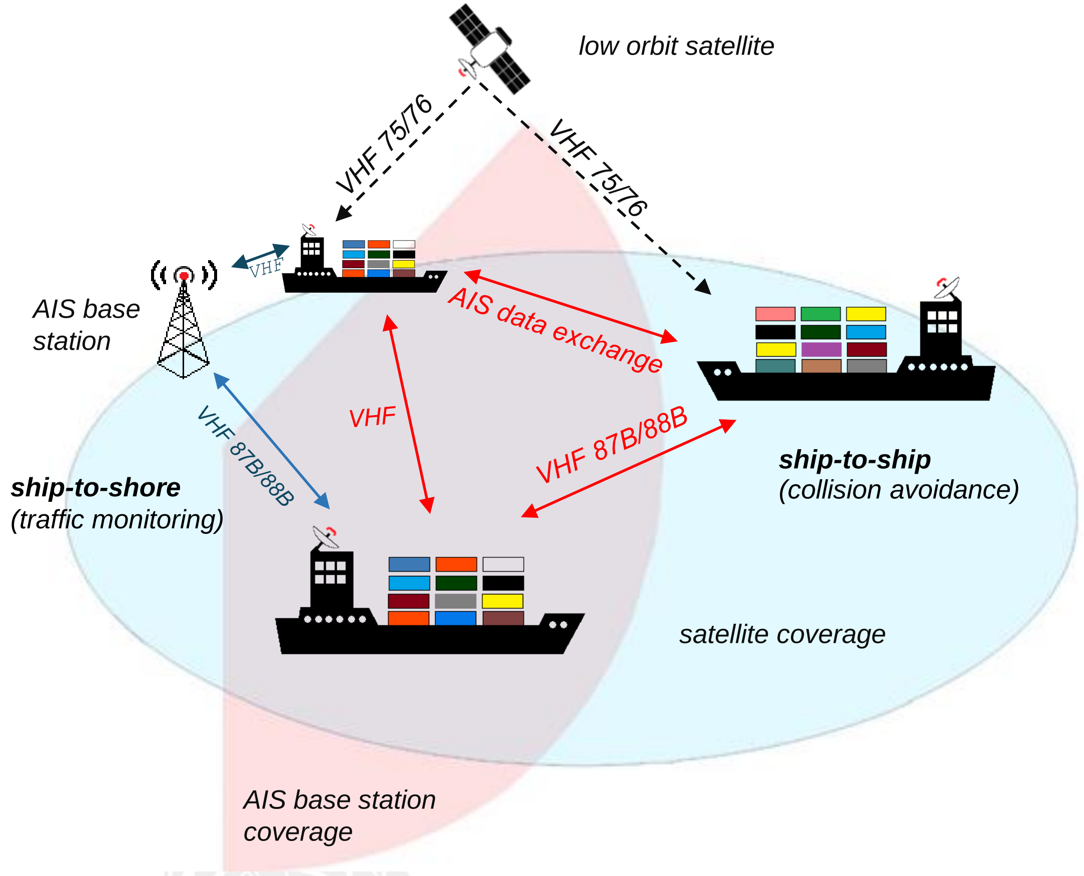


Figure 2. AIS – the principlr of work

Source : own study

Dynamic data

- Information on ship movements
- Automatically transmitted
- Every 2 to 10 seconds depends on vessel's speed
- Every 3 to 6 minutes when anchored

- Maritime Mobile Service Identity number (MMSI)
- AIS navigational status
- Rate of turn
- Speed over ground
- Position coordinates (longitude/latitude)
- Course over ground
- Heading
- Bearing at own position
- UTC second

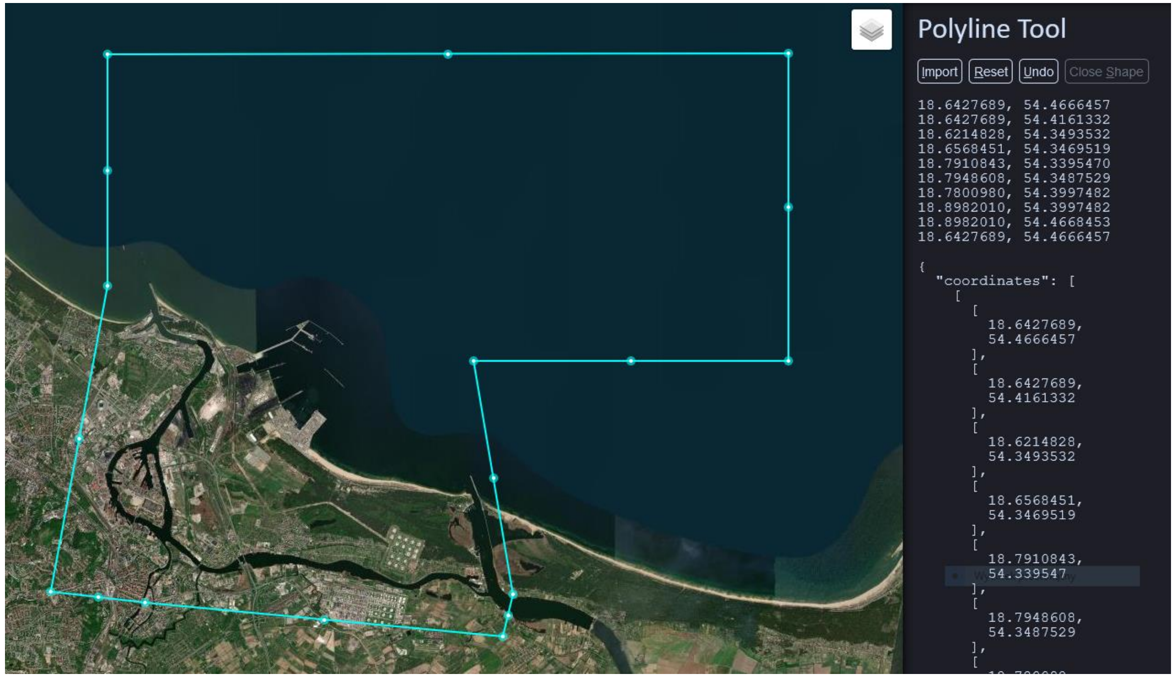
Static data

- Information on ship characteristic
- Manually transmitted
- Every 6 minutes

- International Maritime Organisation number (IMO)
- Call sign
- Name
- Type
- Dimensions
- Location of the positioning antenna on the vessel
- Type of positioning system
- Draught
- Destination
- ETA (estimated time of arrival)

Statistics of traffic intensity in maritime transport

The port of Gdańsk



The port of Gdynia

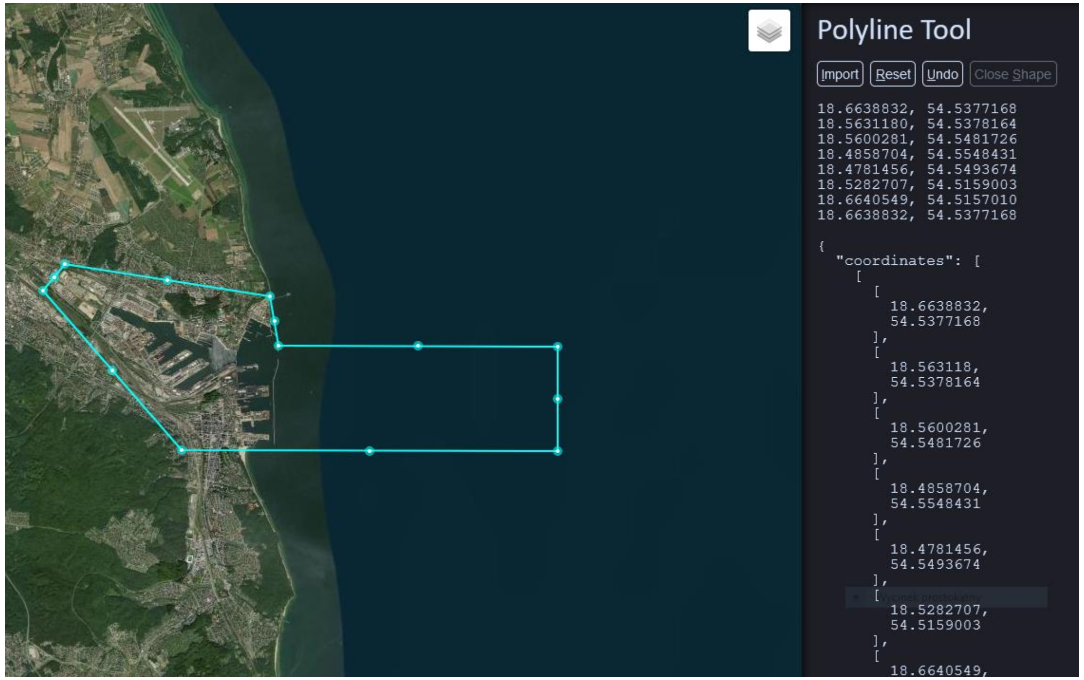
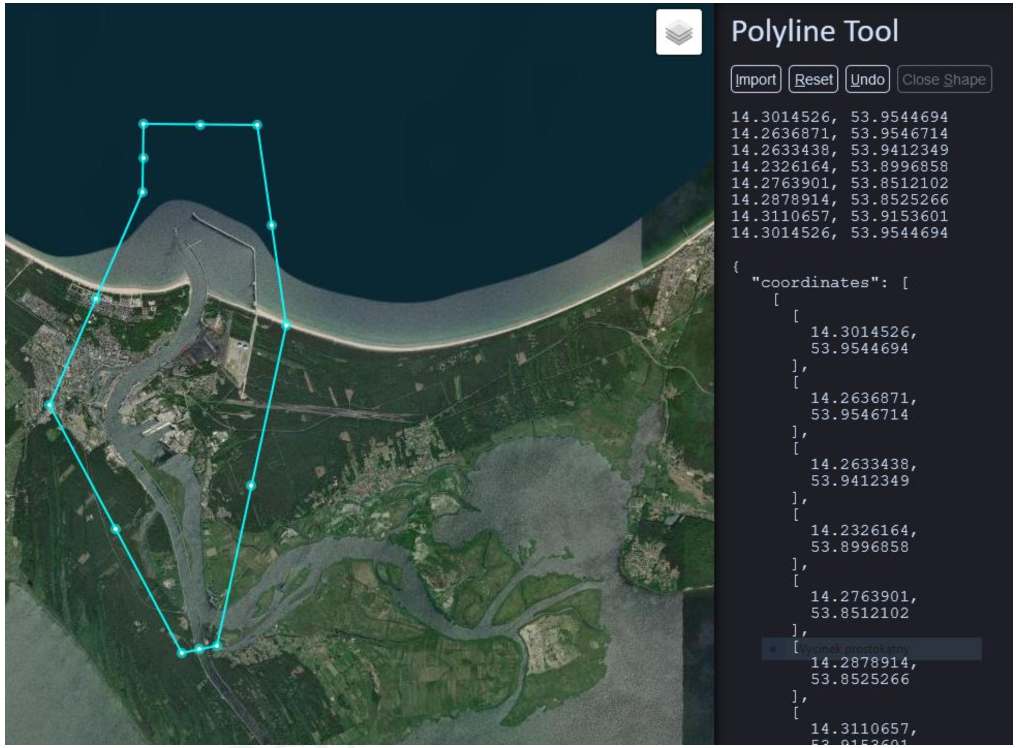


Figure 3. The ports of Gdańsk, Gdynia

Source: own study, generated on the basis of the tool:
https://www.keene.edu/campus/maps/tool/

Statistics of traffic intensity in maritime transport

The port of Świnoujście



The port of Szczecin

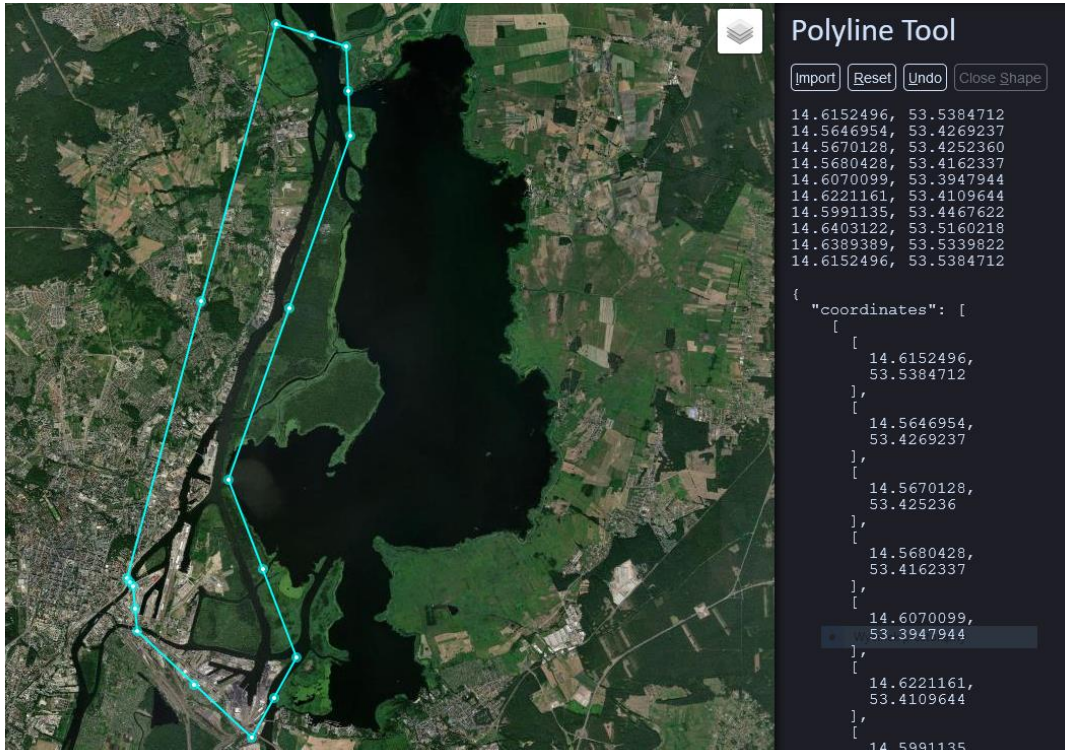


Figure 4. The ports of Świnoujście, Szczecin

Source: own study, generated on the basis of the tool:
https://www.keene.edu/campus/maps/tool/

Statistics of traffic intensity in maritime transport - assumptions

Traffic intensity is understood as the intensity of the stream, defined as the number of transport units passing through the line delimiting a given area in a certain period of time.

Implementation in TranStat application:

Location: ports of Gdańsk, Gdynia, Szczecin, Świnoujście.

Data source: Automatic Identification System (AIS)

Statistics of traffic intensity in maritime transport - assumptions

As a result of the developed algorithms for the traffic intensity, the following variables and breakdowns are obtained, among others:

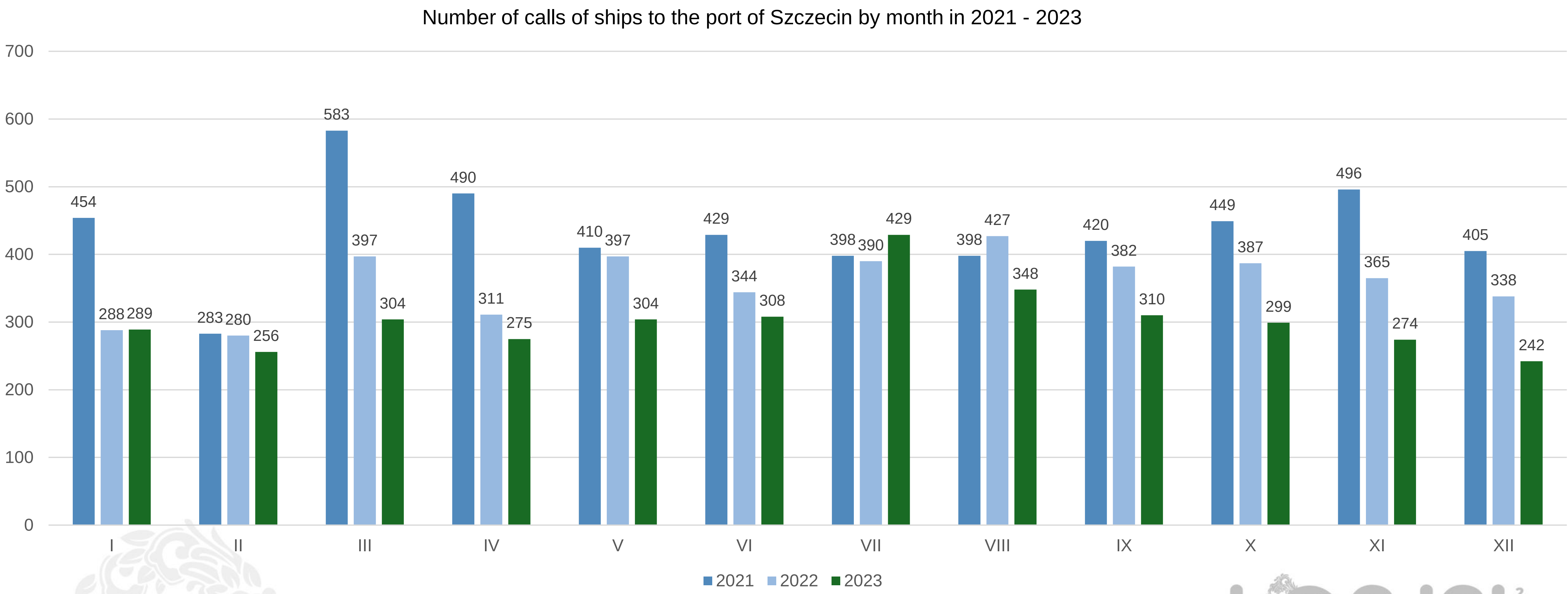
variables:

- number of ships at the port
- number of calls by ships;

breakdowns:

- time: day, month, quarter, year
- location: ports of Gdańsk, Gdynia, Szczecin, Świnoujście
- means of maritime transportation: by type of ships, by country of flag

Statistics of traffic intensity in maritime transport - results



Source: own study based on the results from the TranStat system

Statistics of traffic intensity in maritime transport

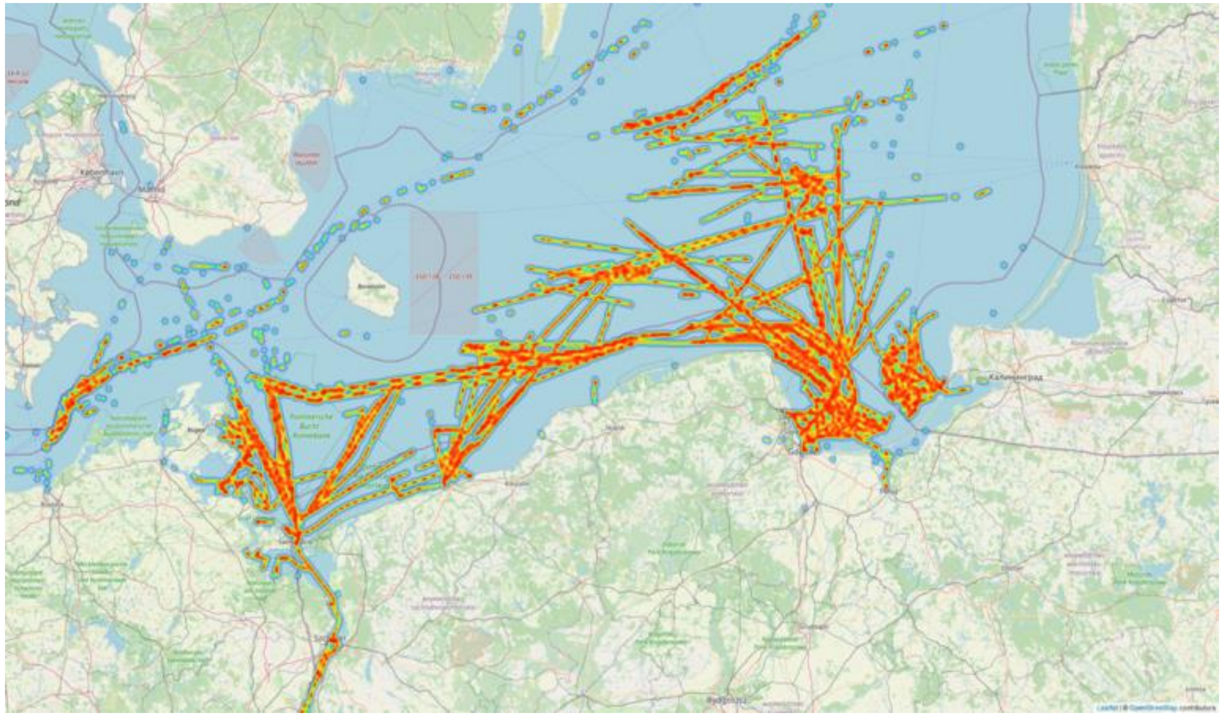


Figure 5. Traffic intensity of vessels for Poland - as of March 1, 2024

Figure 6. Traffic intensity of vessels for port of Świnoujście - as of March 1, 2024

Source: own study based on the results from the TranStat system

Statistics of transportation volume in maritime transport

Tonne-kilometre (tkm) is the unit of measure representing the transport of one tonne of cargo in a ship over one kilometre.

Passenger-kilometre (pkm) is the unit of measure representing the transport of one passenger in a ship over one kilometre.

We need to know about : the amount of cargo (loaded/unloaded) and the ship's route.

Implementation in TranStat application:

- **Location:** ports of Gdańsk, Gdynia, Szczecin, Świnoujście.
- **Data source:** Automatic Identification System (AIS), Maritime transport data set based on Directive 2009/42/EC of the European Parliament and of the Council of 6 May 2009 on statistical returns in respect of carriage of goods and passenger by sea.

Statistics of transportation volume in maritime transport

The transportation volume estimation model implements the presentation of possible ship routes in the form of a directed (weighted) graph, where the vertices of the graph are navigation points and the edges are straight sections between them.

Each edge contains the coordinates of the start and end points, and the weight is the distance between individual nodes, calculated by the Haversine formula.

- The graph consists of 9 859 vertices covering the entire globe.
- There are 10 731 connections between the vertices.
- Ports are vertices that have been described with UNLOCODE.
- There are 3 564 ports included in the graph.
- The sum of the weights of the edges of the graph is 1 088 864 km.

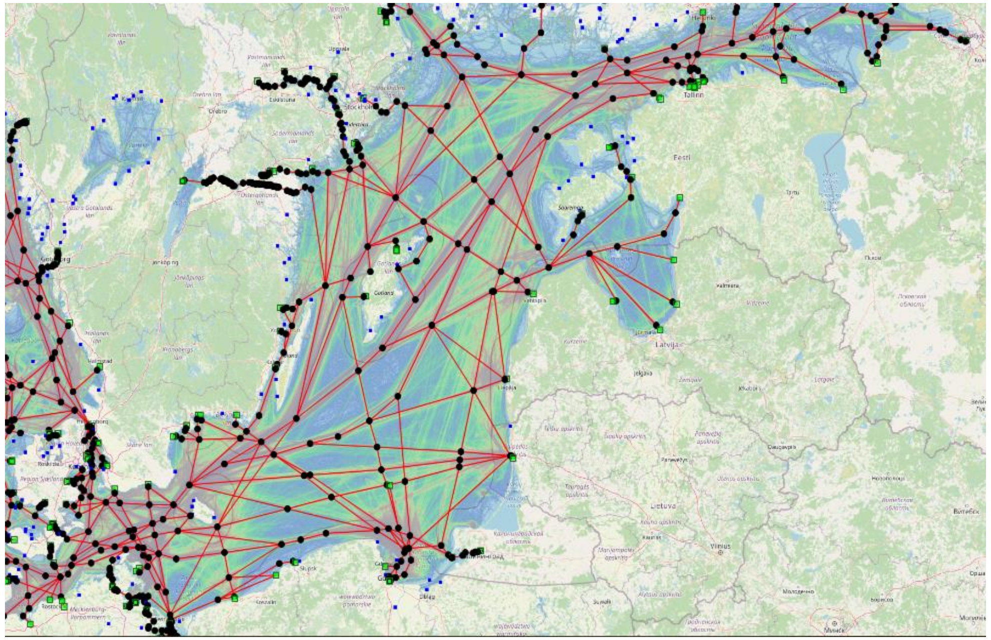


Figure 7. *Graph visualization for the Baltic Sea*

Source: Maritime University of Szczecin

Statistics of transportation volume in maritime transport

Implementation of port distance estimation based on directed graph:

- determining the weights of the edges of a graph - the Haversine formula.
- finding the shortest path in a graph- the Dijkstra's algorithm

A - start vertex , C - end vertex

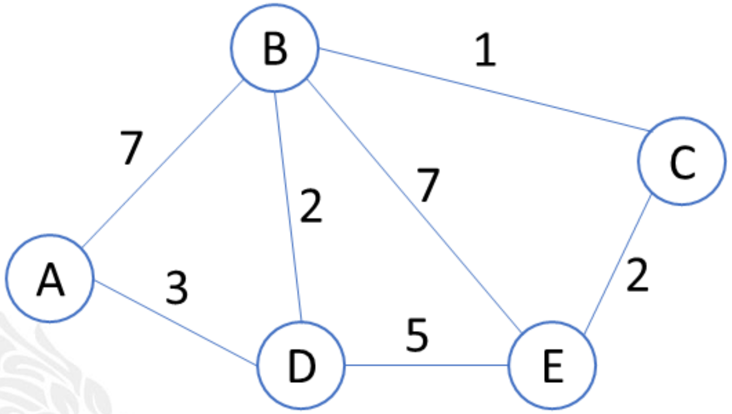


Figure 8. Graph with the representation of weights

Calculated shortest path from A to C is 6
and goes through vertices A, D, B and C

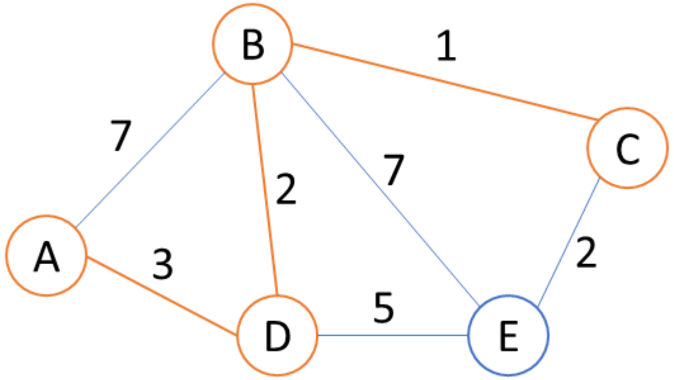


Figure 9. A graph with a representation of the shortest path

Statistics of transportation volume in maritime transport

As a result of the developed algorithms for the transportation volume, the following variables and breakdowns are obtained, among others:

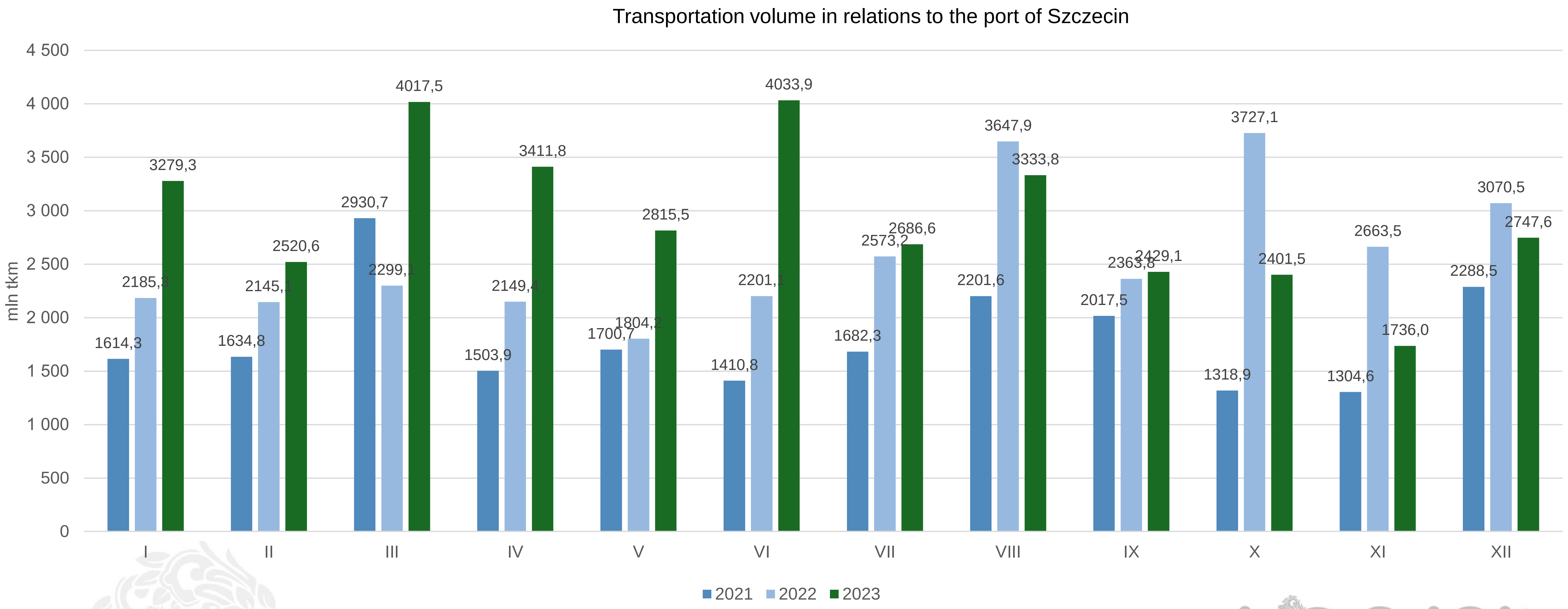
variables:

- transportation volume for cargo and passengers,
- avarage transport distance for 1 tonne of cargo in kilometers
- avarage transport distance for 1 passenger in kilometers

breakdowns:

- time: day, month, quarter, year,
- location: ports of Gdańsk, Gdynia, Szczecin, Świnoujście
- means of maritime transportation - by type, by flag, by gross tonnage,
- type of cargo – cargo group, commodity group.

Statistics of transportation volume in marine transport - results



Source: own study based on the results from the TransMor system

IOS-ISI 2024 MEXICO CONFERENCE

Statistics Poland

e-TOLL – electronic toll collection system

eTOLL

Krajowa Administracja Skarbowa

e-TOLL is an advanced solution developed, implemented, maintained and monitored by the Head of the National Revenue Administration

It is based on the Global Navigation Satellite System for user position location with the use of virtual gates.

In the TranStat system, we only test heavy vehicles with a maximum permissible weight of more than 3,5 tons, including buse

The length of the paid sections is currently approximately 3,677 km

Figure 10. National road network including the e-TOLL system

source: own study

IOS-ISI 2024
MEXICO CONFERENCE

Statistics Poland

Statistics of traffic in road transport - assumptions

In total, there are 951 virtual gates on motorways, expressways and national roads covered by the e-TOLL system.

In order to create statistics on traffic volume, it was assumed that a vehicle made a trip under the e-TOLL system if it was registered in at least 2 transactions from the analyzed dataset.

As a result of the developed algorithms for the traffic the following variables and breakdowns are obtained, among others:

variables:

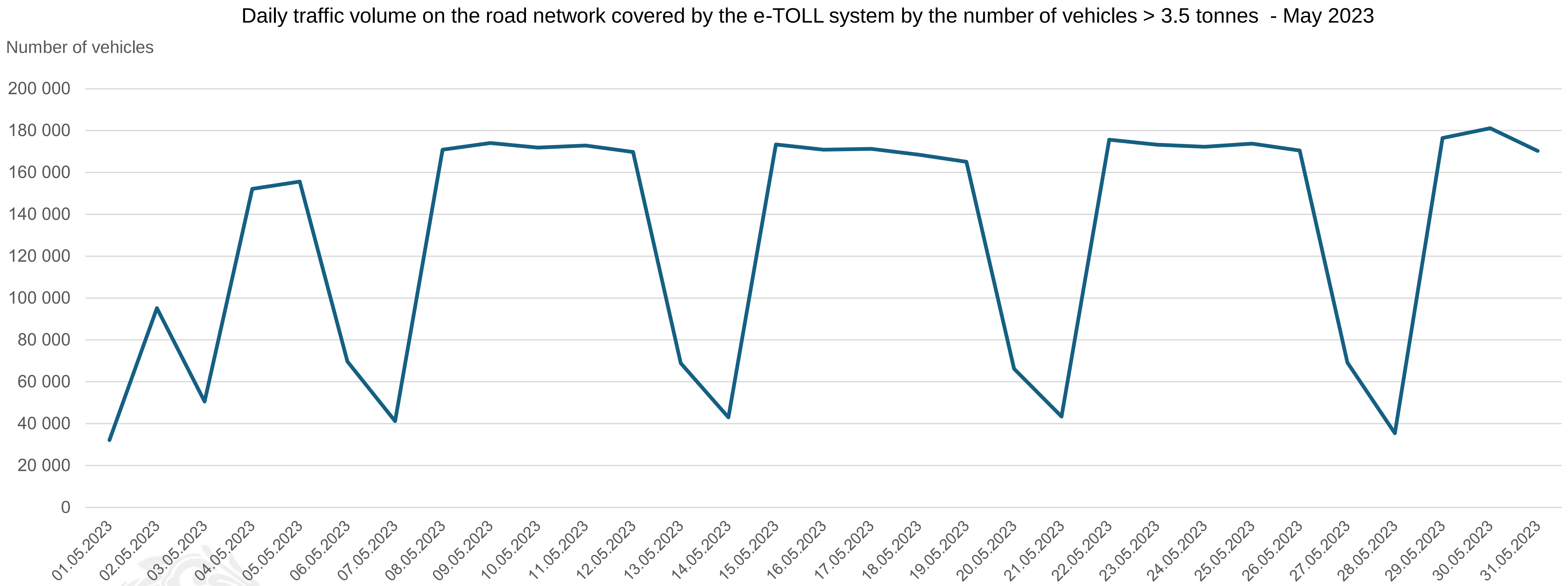
- number of transactions - the number of toll transactions for vehicles subjected to toll, registered on the toll section;
- number of vehicles - unique number of vehicle occurrences at a toll collection point or section.

Statistics of traffic in road transport - assumptions

breakdowns:

- time: day, week, month
- spatial: road number
- categories of vehicles according to payload groups (GVW)
- coaches, capacity group 30, with more than 9 seats (including the driver),
    - heavy duty vehicles:
    - load group 41 – heavy duty vehicles with a GVW above 3.5 tons and below 12 tons,
    - load group 42 - heavy duty vehicles with a GVW above 3.5 tonnes and below 12 tonnes with the physical ability to tow a trailer,
    - load group 50 – heavy duty vehicles with a GVW over 12 tons.
- categories of vehicles according to the Euro emission class (0 - 6) - European emission standard specifying the standards of permissible emissions in new vehicles sold in the EU and the European Economic Area.

Statistics of traffic in road transport - results



Source: own study based on the results from the TranStat system

Conclusions

The implementation of the TranStat project in the field of maritime statistics has enriched the current statistical production carried out by Statistics Poland through:

- access to streaming Big Data source related to maritime transport (AIS);
- implementation of the necessary Big Data technology for sensory data enabling an automatic process of data flow, validation and processing;
- development of traffic intensity, transportation volume and emissions models in maritime transport with the use of sensory data;
- development of algorithms enabling generation of new statistics and obtaining new knowledge in the field of maritime transport statistics by using the correlation of multiple data sources;
- reduction of research costs thanks to the use of modern technology in the collection and processing of non-statistical sources (AIS).

Thank you