



# datAcron

## Big Data Analytics for Time Critical Mobility Forecasting

Waldo Kleynhans: IMIS Global

## Presenter



Dr. Waldo Kleynhans received a Ph.D. (Electronic Engineering) from the University of Pretoria, South Africa and an MBA from Heriot-Watt University, Scotland.

He is focused on statistical signal processing and machine learning in the areas of telecommunication and maritime domain awareness.

He is a past chair of the IEEE Aerospace & Electronics Systems and IEEE Geoscience and Remote sensing local chapters.

Dr. Kleynhans is the Research and Development Manager at IMIS Global Limited.



## What is **datAcron**:

1. Project funded by the European Union's Horizon 2020 Program (<http://datacron-project.eu/>)
2. Research and innovation collaborative project introducing novel methods to detect threats and abnormal activity of large numbers of moving entities in vast geographic areas.
3. Aims to advance the management and integrated exploitation of voluminous and heterogeneous data-at-rest (archival data) and data-in-motion (streaming data) sources.
4. Advance the capacities of systems to promote safety and effectiveness of critical operations for large numbers of moving entities in large geographical areas.
5. Covers both the Air-traffic management (ATM) and maritime domains.

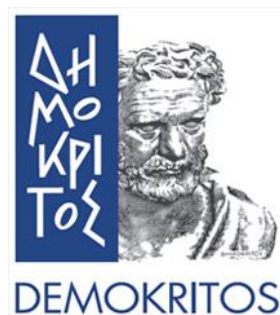
## The datAcron consortium



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
UNIVERSITY OF PIRAEUS



UNIVERSITY OF PIRAEUS  
RESEARCH CENTER



Fraunhofer





## Maritime data in **datAcron**:

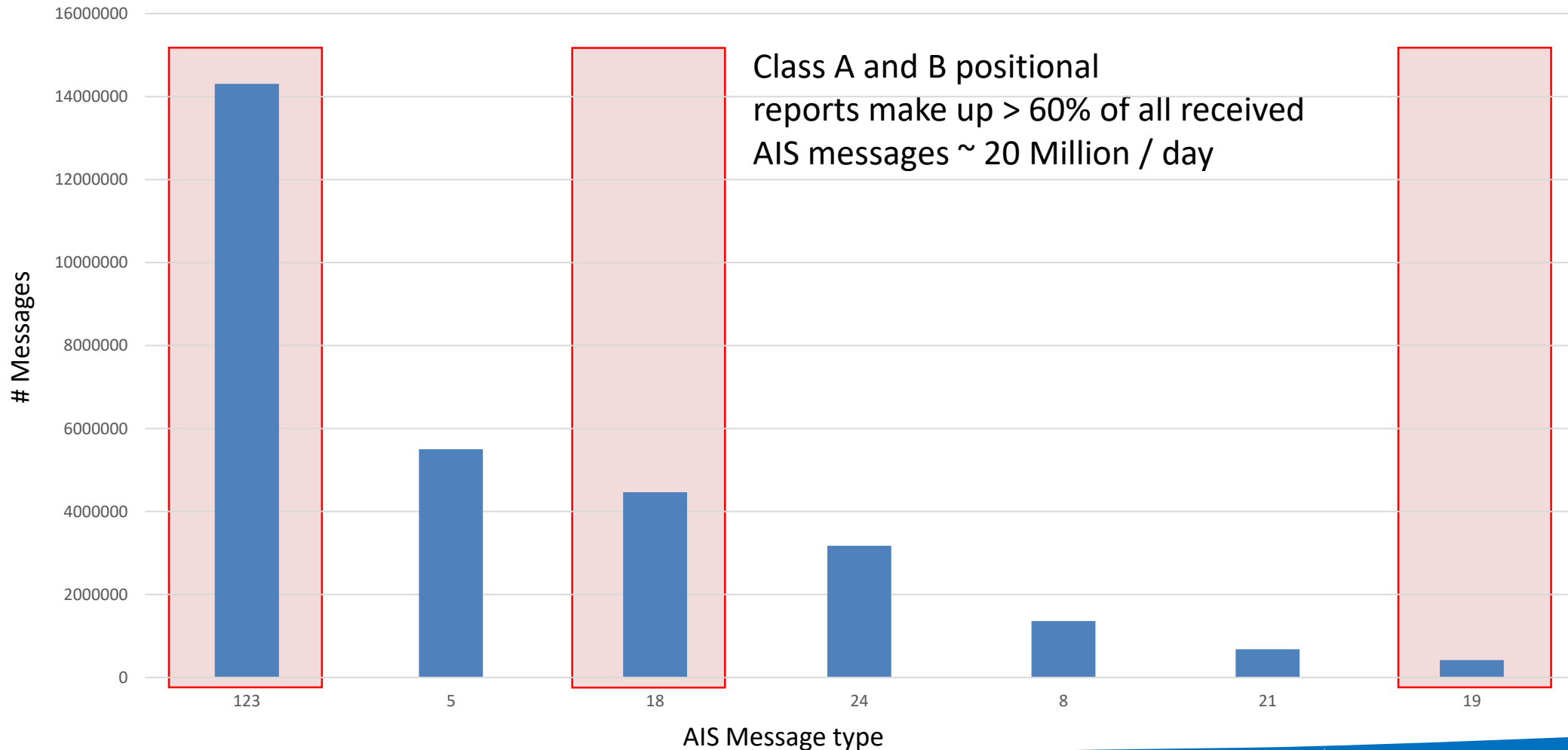
1. Maritime data in datAcron is mostly focused on AIS data.
2. AIS data is increasing! This is due to the global increase in vessels fitted with AIS transmitters as well as the proliferation of satellite and terrestrial AIS receiving stations.
3. While increased AIS data volumes are beneficial, the processing and storage of these large data volumes can become challenging.





# Data Compression

AIS data stream example: Total number of AIS messages per day ~ 30 Million / day [11 Billion Annually]





# Data Compression

1. datAcron uses an on-line critical point extraction methodology to filter positional AIS data
2. Spatio-temporal information contained in each newly presented AIS message is compared to the current vessel track history.
3. A decision is then made to include the new AIS position as a critical point in the compressed dataset based on a distance heuristic
4. Typical compression of >80% can be achieved when considering terrestrial AIS data.

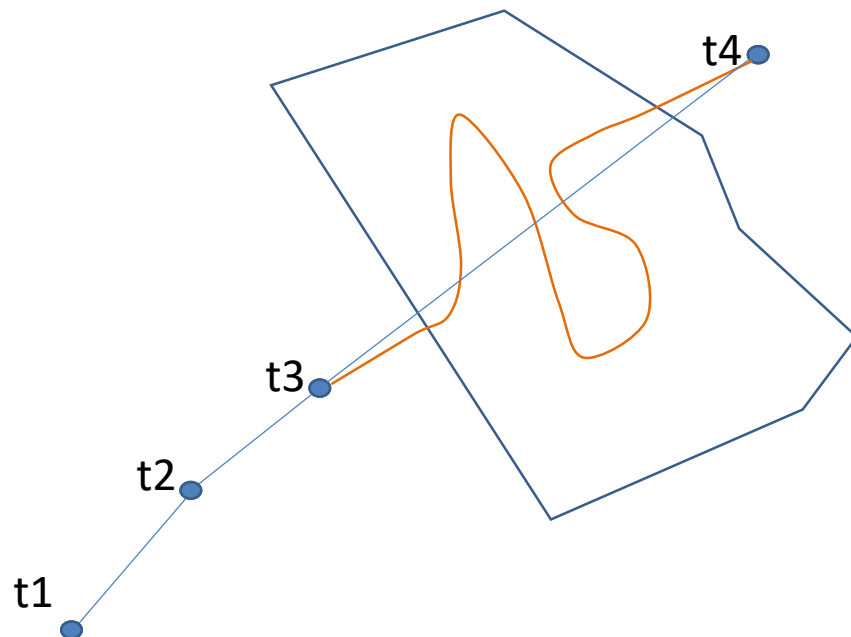


## Moving from points to tracks

1. Saving only critical points reduces data significantly.
2. These critical points now need to translate into a coherent track to increase its utility.
3. **Example:** Detecting the transition of a vessel into a Marine Protected Area (MPA)

# Moving from points to tracks 1

Marine Protected Area



1. **Trajectory data analytics:** Did the vessel enter an MPA?
2. **Future location predictor:** Given points t1, t2 and t3 what is the probability of this vessel entering the MPA?
3. **Complex event recognition:** Given the trajectory with its spatio-temporal characteristics and the MPA, what is the probability that the vessel was fishing while inside the MPA?



## Moving from points to tracks 2

- Working only with critical data points reduces resource requirements for streaming analytics significantly.
- There is still a requirement to save both critical and complete datasets for off-line analysis
- These data-points are in the order of billions
- datAcron uses a Resource Description Framework (RDF) for efficient storage of these data in the form of triples (subject, property, object)



## Data storage

By design, the datAcron distributed RDF engine targets the following objectives:

1. **Scalable storage and processing for vast volumes of RDF data**, in the order of Billions of RDF triples (targeting the Volume dimension of Big Data)
2. **Support for spatio-temporal RDF data**, i.e., RDF data which are mostly associated with spatio-temporal information
3. **Efficient processing strategies** that employ optimization techniques, in order to prune significant subsets of data and reduce execution time

## Data storage

### Example:

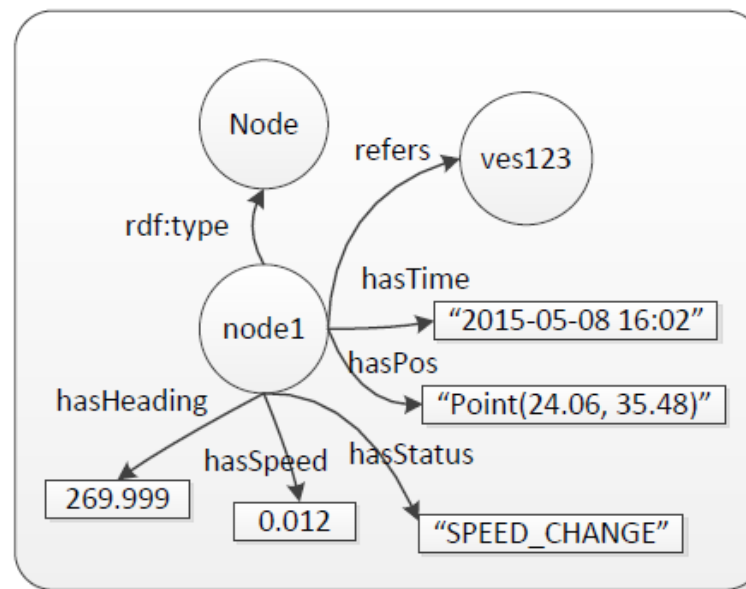
Vessel (ves123) reporting its spatio-temporal position as (x=24.06, y=35.48) at time t=2015-05-08 16:02 as well as its speed, heading and status.

Subject	Property	Object
:node1	:refers	:ves123
:node1	rdf:type	:Node
:node1	:hasTime	"2015-05-08 16:02"
:node1	:hasPos	"Point(24.06, 35.48)"
:node1	:hasStatus	"SPEED_CHANGE"^^xsd:string
:node1	:hasSpeed	0.012^^xsd:double
:node1	:hasHeading	269.999^^xsd:double

(a) RDF Triples

Subject	Property	Object
100	-1	-2
100	-3	-4
100	-5	-6
100	-7	-8
100	-9	-10
100	-11	-12
100	-13	-14

(c) Integer-encoded Triples

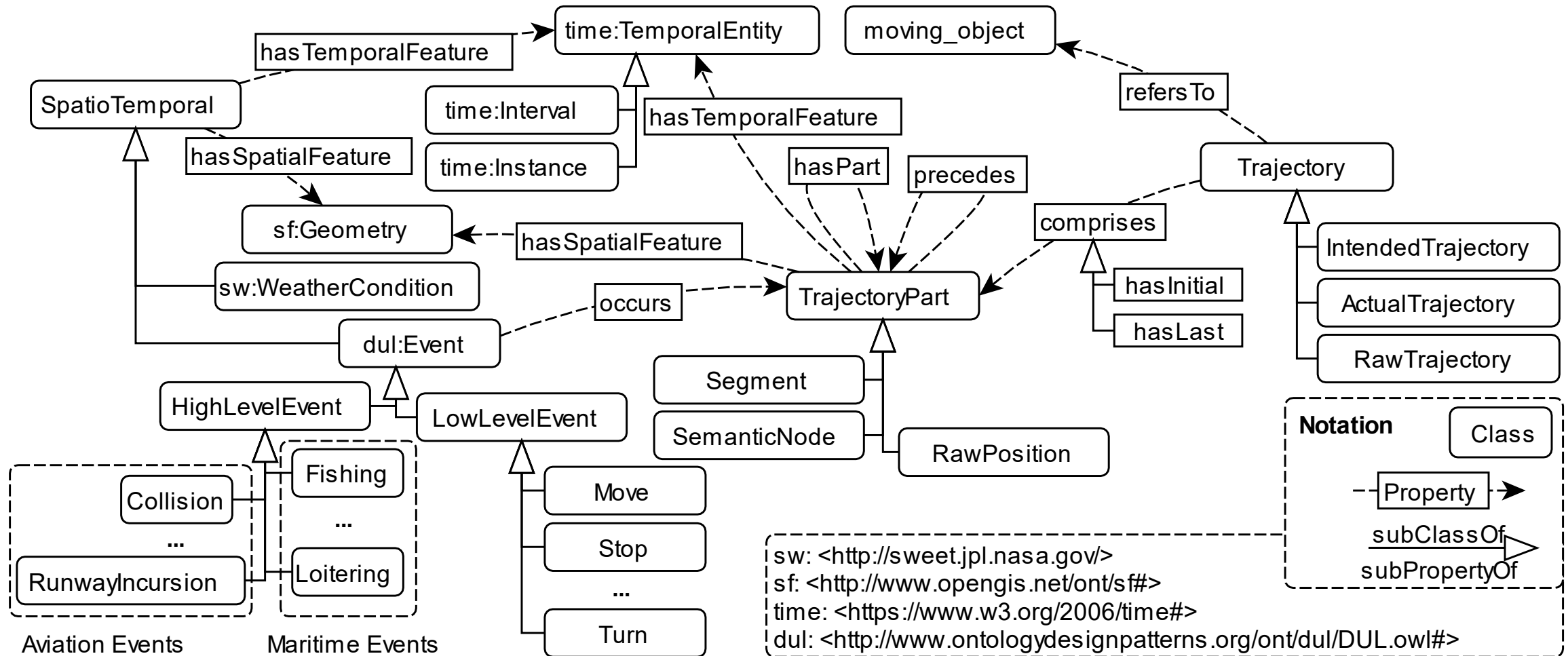


(b) RDF Graph

ID	URI/Literal
100	:node1
-1	:refers
-2	:ves123
-3	rdf:Type
-4	:Node
-5	:hasTime
-6	"2015-05-08 16:02"
-7	:hasPos
-8	"Point(24.06, 35.48)"
-9	:hasStatus
-10	"SPEED_CHANGE"
-11	:hasSpeed
-12	0.012
-13	:hasHeading
-14	269.999

(d) Dictionary

# datAcron ontology

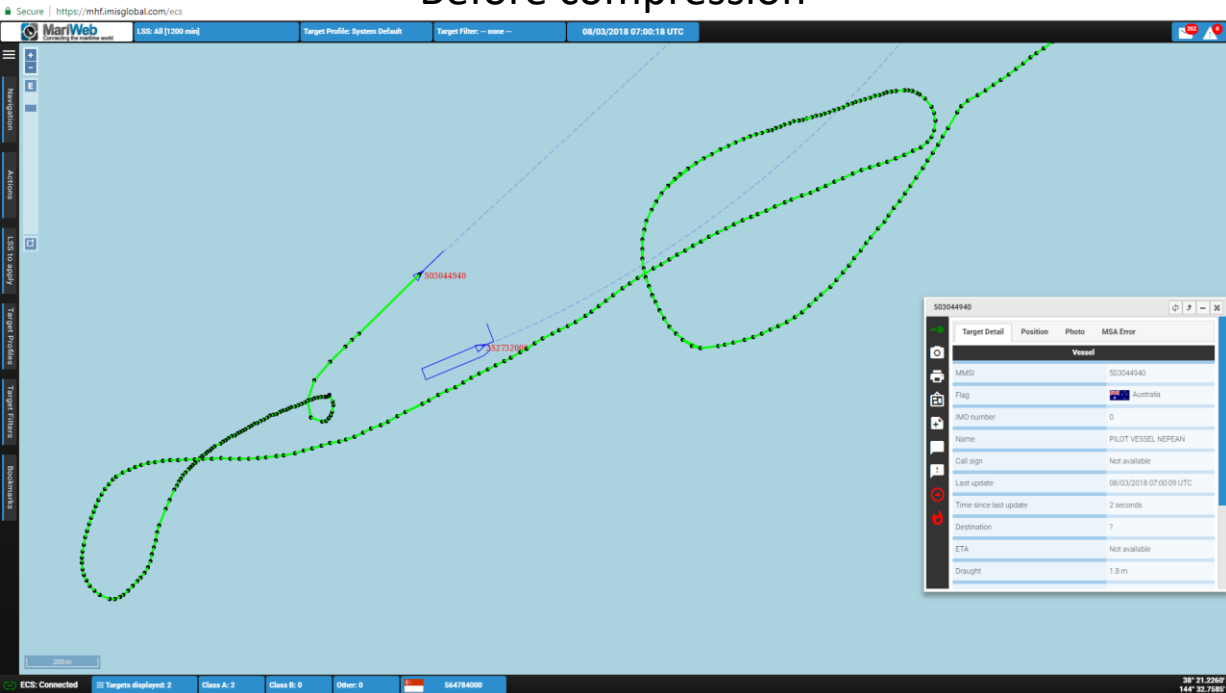


[http://ai-group.ds.unipi.gr/datacron\\_ontology/](http://ai-group.ds.unipi.gr/datacron_ontology/)

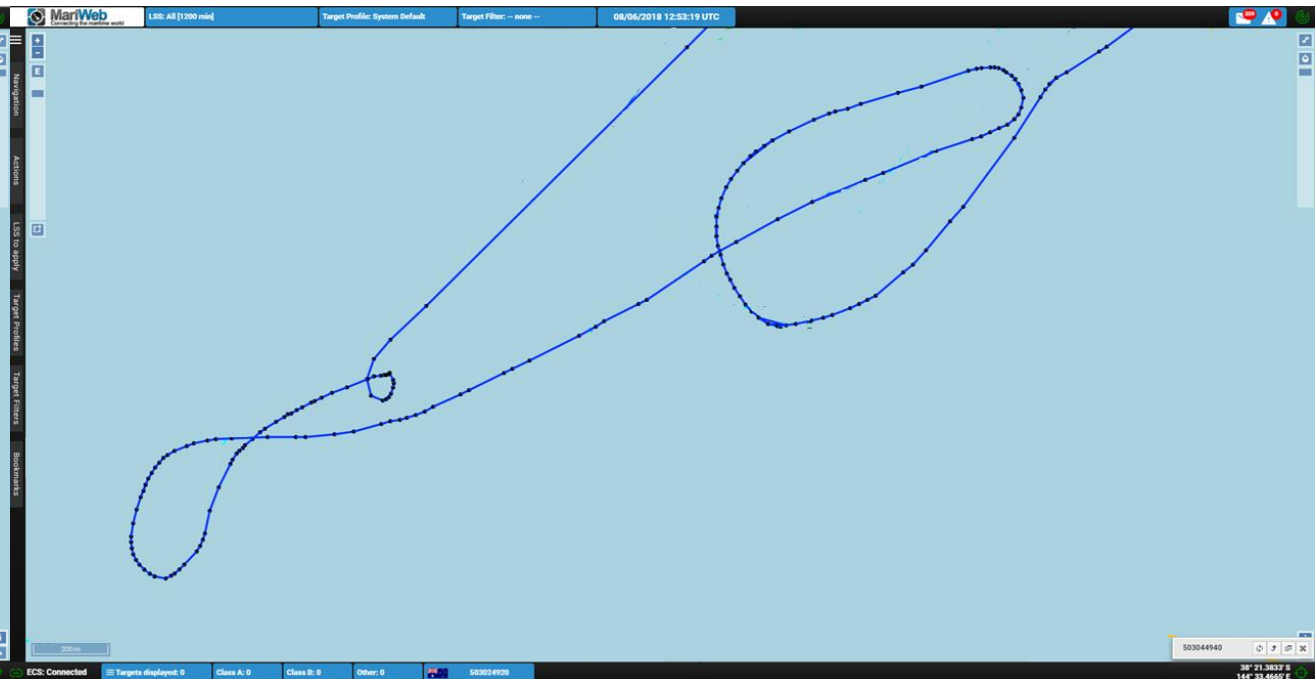
# Practical implementation of the datAcron concepts

Data compression: Within harbor critical point extraction example

Before compression



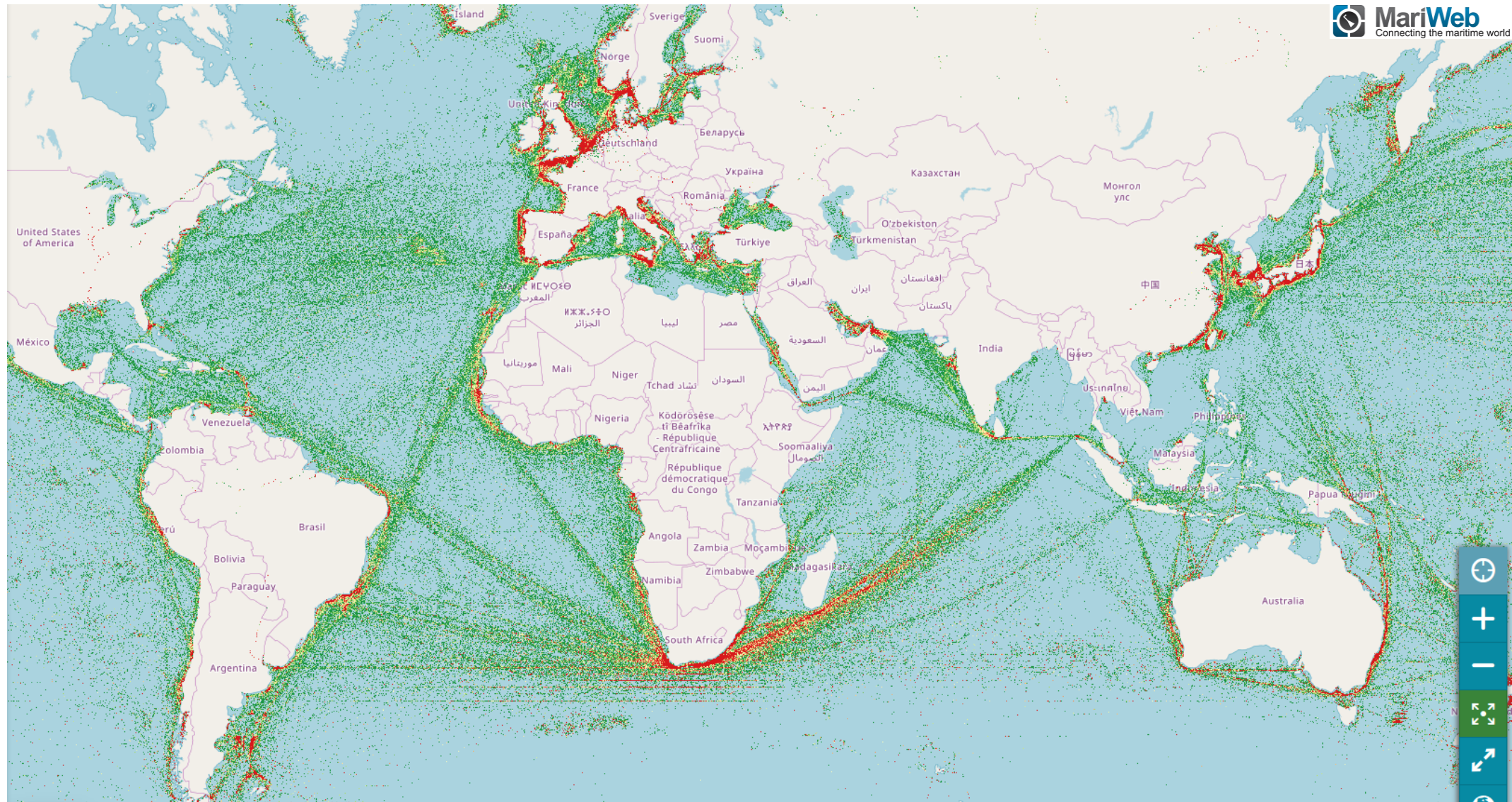
After compression



Overall compression achieved > 80%

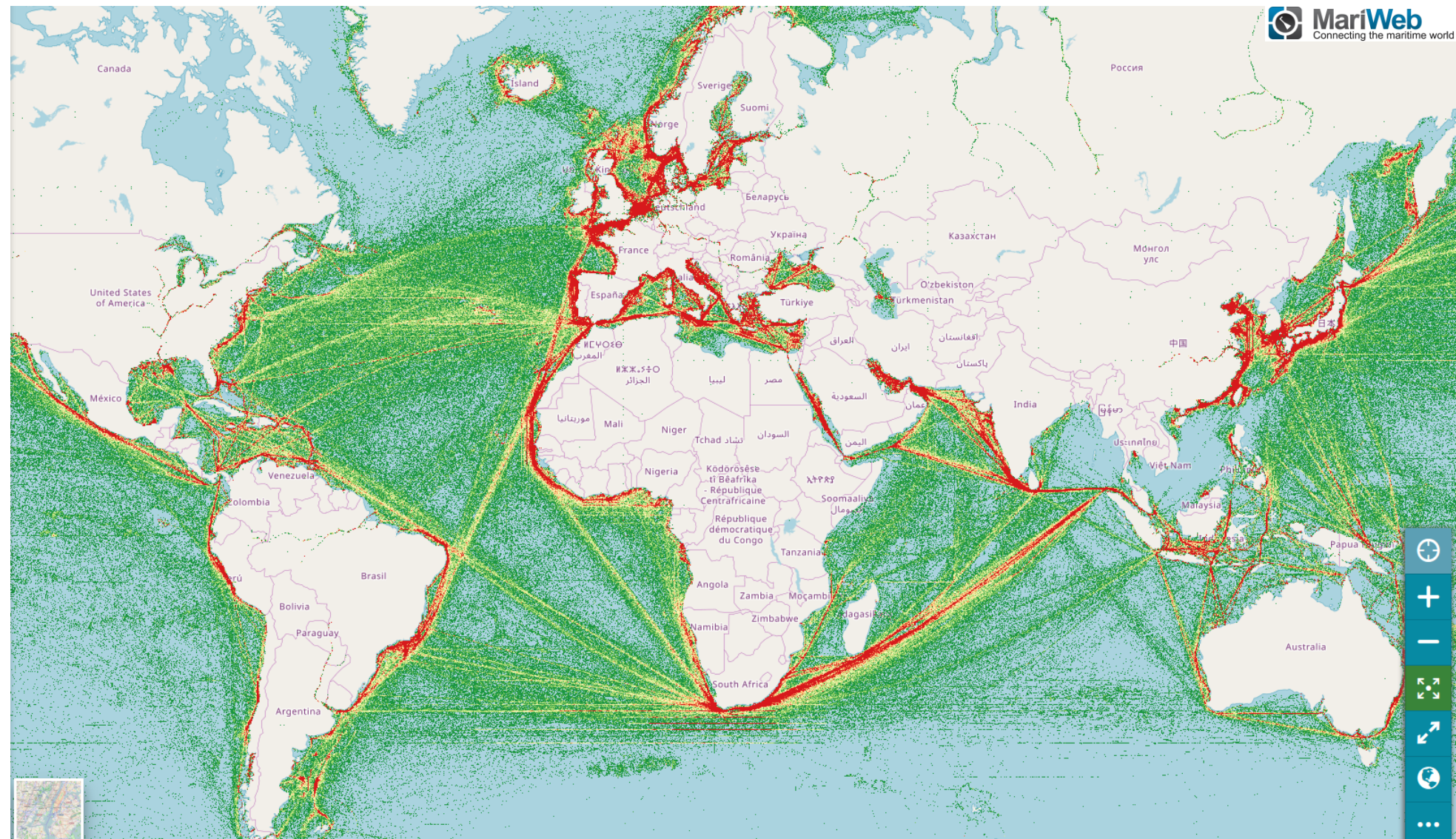
# Practical implementation of the datAcron concepts

Points heatmap  
using a commercial  
analytics platform



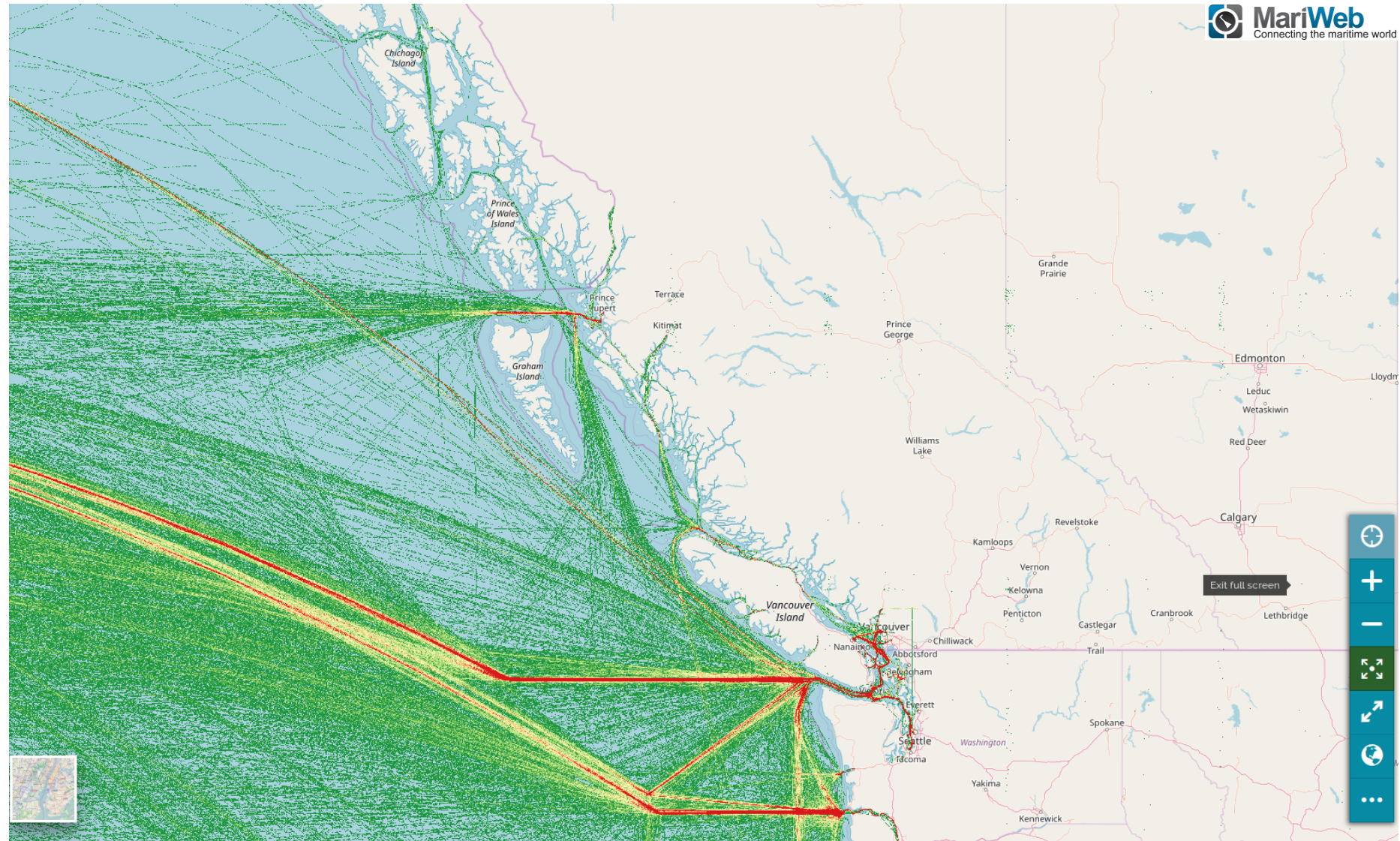
# Practical implementation of the datAcron concepts

Tracks heatmap  
using a commercial  
analytics platform



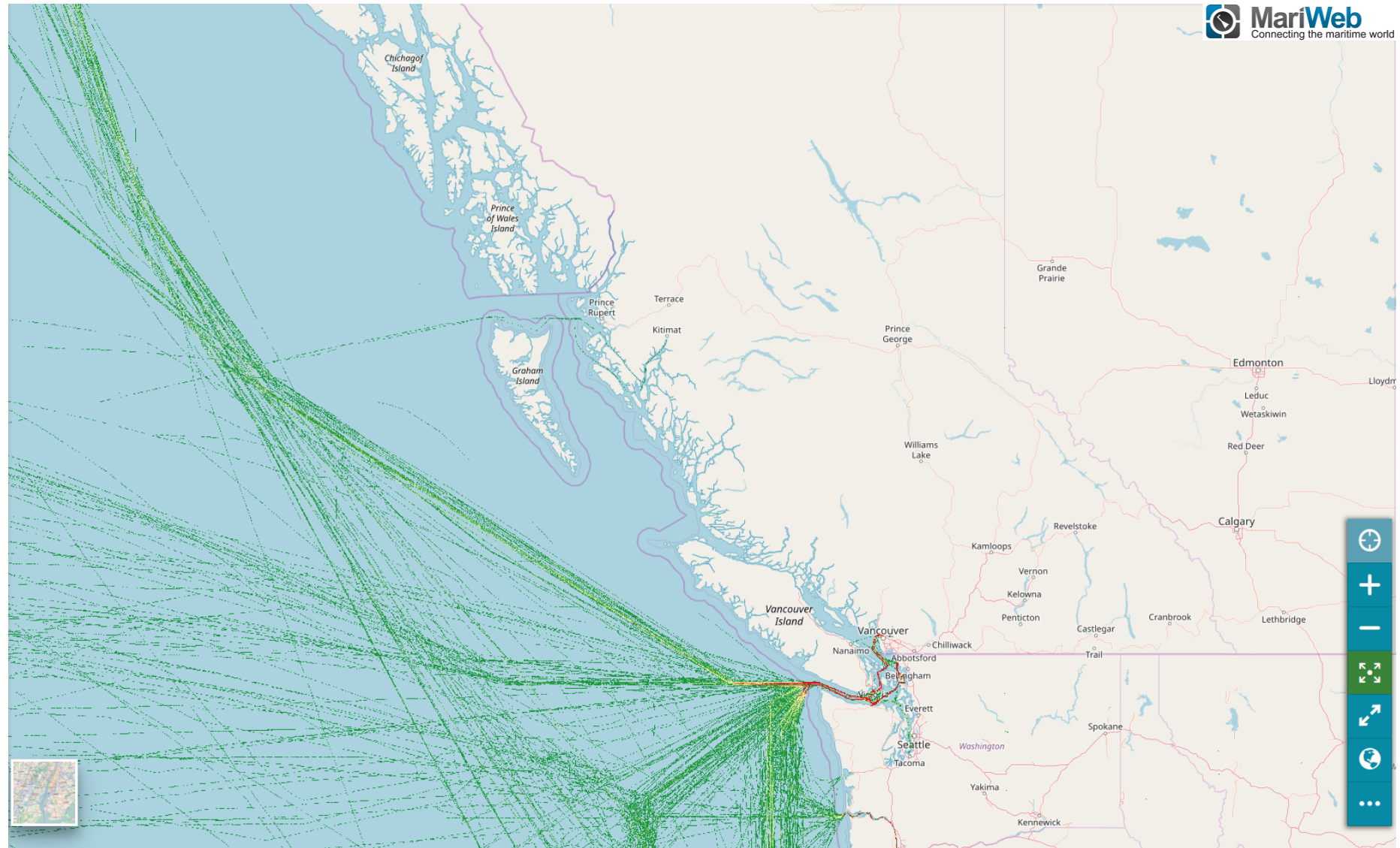
# Practical implementation of the datAcron concepts

Cargo tracks heatmap  
using a commercial  
analytics platform



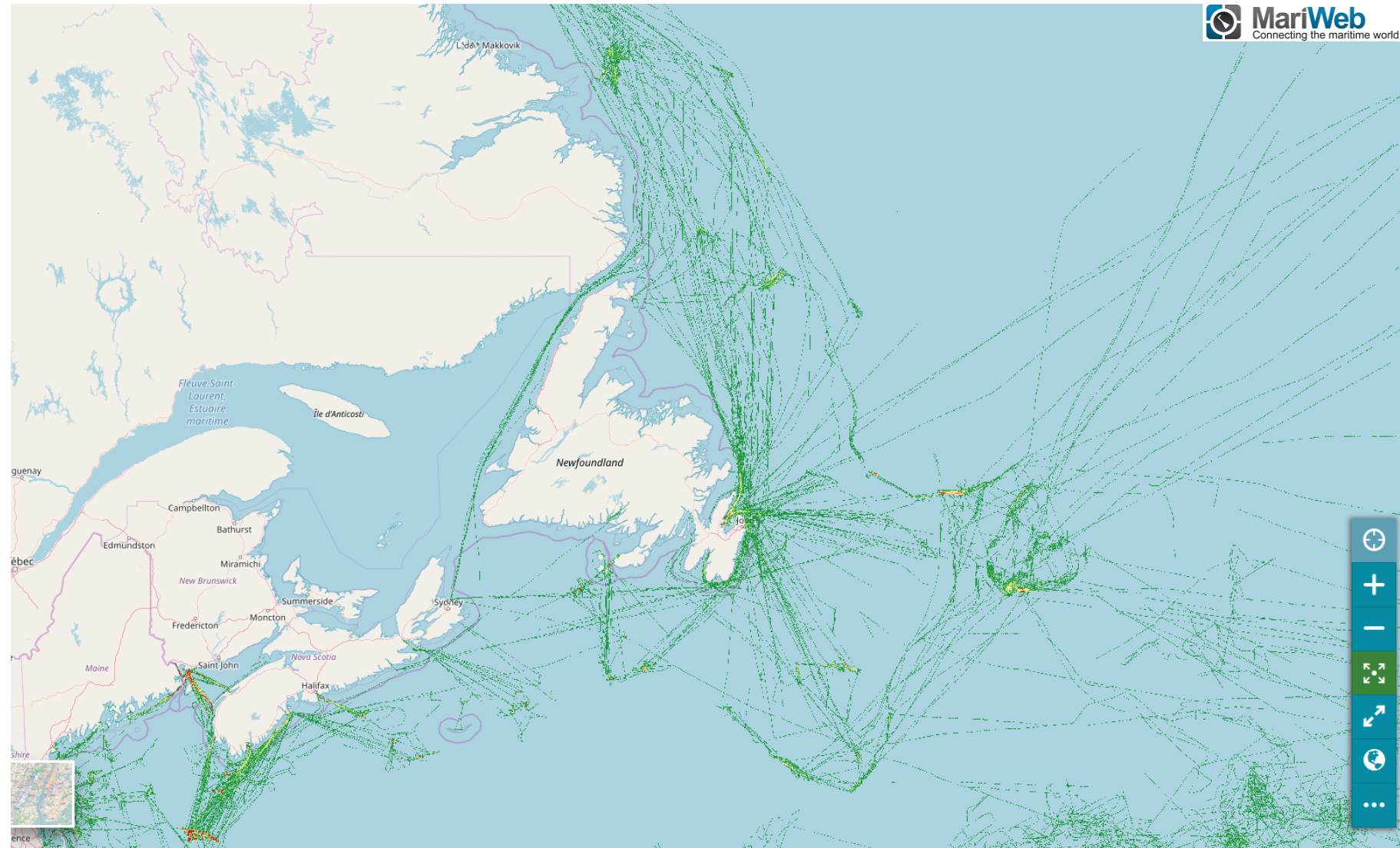
# Practical implementation of the datAcron concepts

Tanker tracks heatmap  
using a commercial  
analytics platform



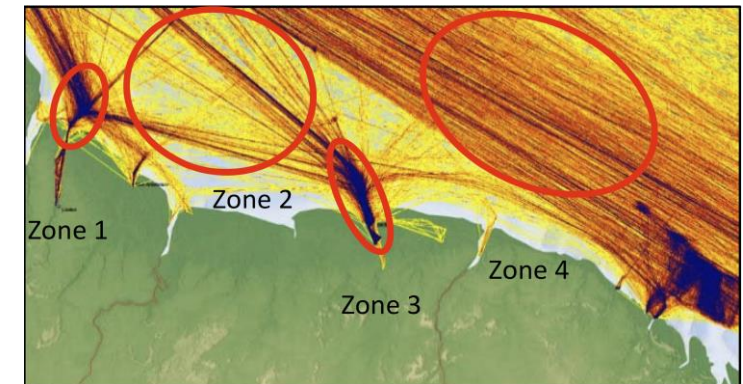
# Practical implementation of the datAcron concepts

Fishing tracks heatmap  
using a commercial  
analytics platform



# Opportunities

1. **Data compression** – Can be achieved giving faster database queries and smaller storage environments.
2. **Moving from geo-spatial points to tracks** - Trajectory data enables data analytics not possible with geo-point data giving new insights especially with regards to traffic related risk models often used in the maritime AtoN domain.  
The conversion of geo-point to trajectories can be standardised.
3. **Data storage** - Moving data into an Resource Description Framework (RDF) environment is efficient for storage and querying with an already developed and documented ontology.





# datAcron

## Big Data Analytics for Time Critical Mobility Forecasting

Waldo Kleynhans: IMIS Global