The datAcron project: A computing framework for the integration of maritime trajectories

C. Claramunt[1], C. Ray[1], L. Salmon[1], E. Camossi[2], M. Hadzagic[2], A.-L. Jousselme[2], G. Andrienko[3], N. Andrienko[3], Y. Theodoridis[4], G. A. Vouros[4]

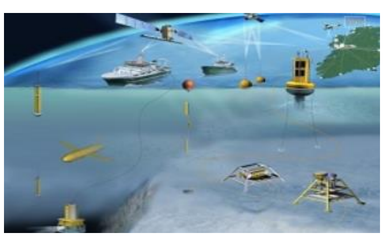
Naval Academy Research Institute, France[1]

CMRE, Italy[2]

Franhaufer Institute, Germany[3]

University of Piraeus, Greece[4]

datAcron



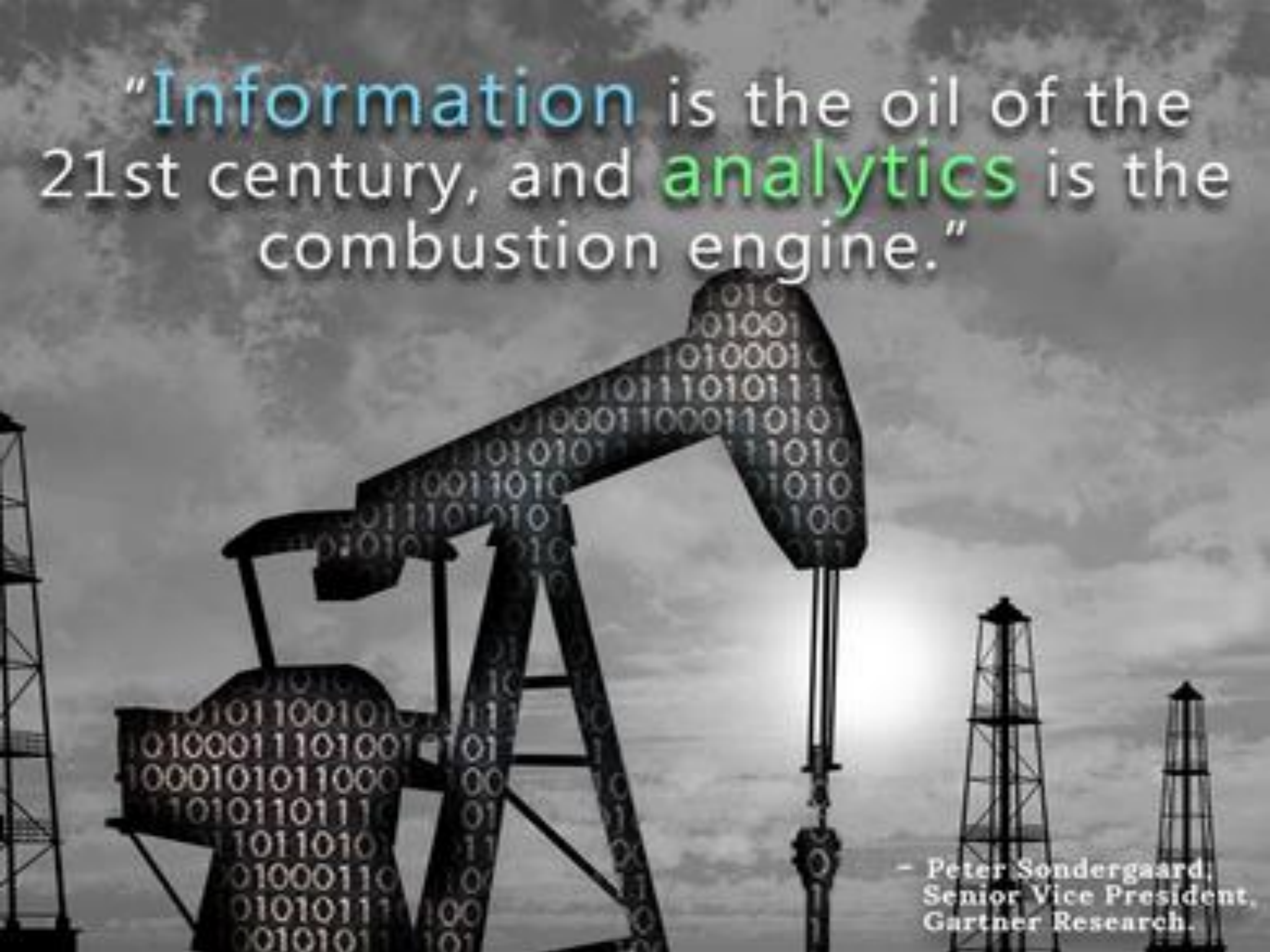
Journées Scientifiques du Projet NOUMEA

Brest, 31 mai & 1er juin 2017

datAcron project* - our vision

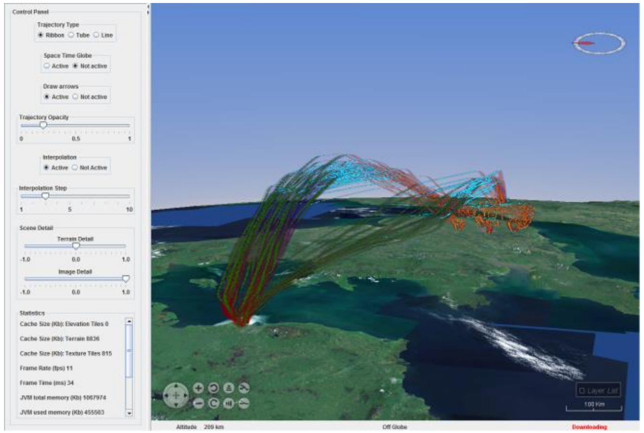
- advance the **management** and **integrated** exploitation of voluminous and heterogeneous **data-at-rest** (archival data) and **data-in-motion** (streaming data) sources, so as to …
- advance the capacities of systems to promote safety and effectiveness of critical operations for **large numbers** of **moving entities** in large **geographical areas**
  - Vessels
  - Aircrafts

*** EU H2020 project under call ICT-16-2015 "Big Data Research"**
(duration: 1.1.2016 – 31.12.2018) Big Data Analytics for Time Critical Mobility Forecasting

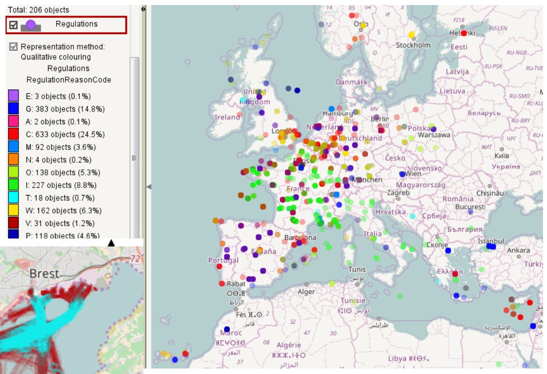

"Information is the oil of the 21st century, and analytics is the combustion engine."
– Peter Sondergaard, Senior Vice President, Gartner Research.

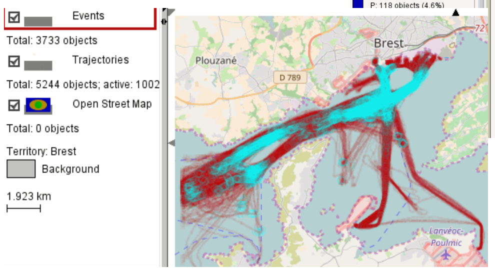
Our objectives

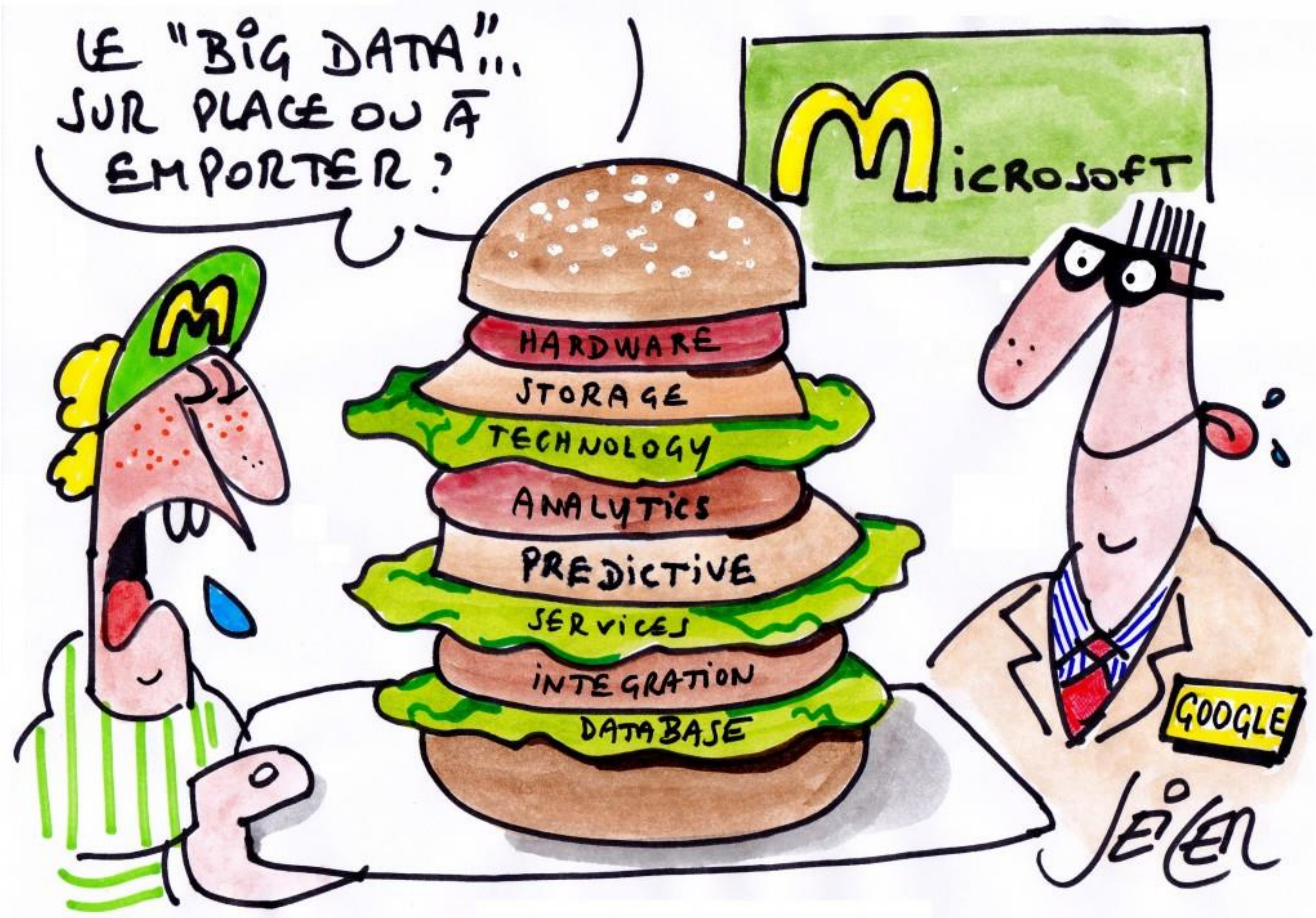
- **Big Data management**
  - Scalable, fault-tolerant cross-streaming in-situ data processing
  - Data integration, automatic link discovery
  - Distributed management and querying of integrated spatio-temporal data



- **Big Data analytics**
  - Analytics for trajectories detection, short- and long-term prediction
  - Analytics for complex event recognition
  - Visual Analytics



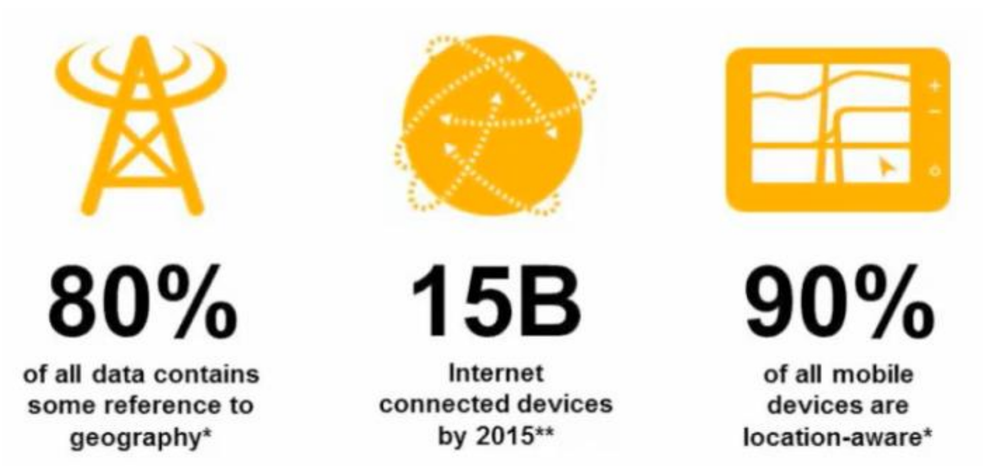



LE "BIG DATA"... SUR PLACE OU À EMPORTER ?
MICROSOFT
HARDWARE
STORAGE
TECHNOLOGY
ANALYTICS
PREDICTIVE
SERVICES
INTEGRATION
DATABASE
GOOGLE
SEILER

From big data to big spatial data…

- **Big Data**

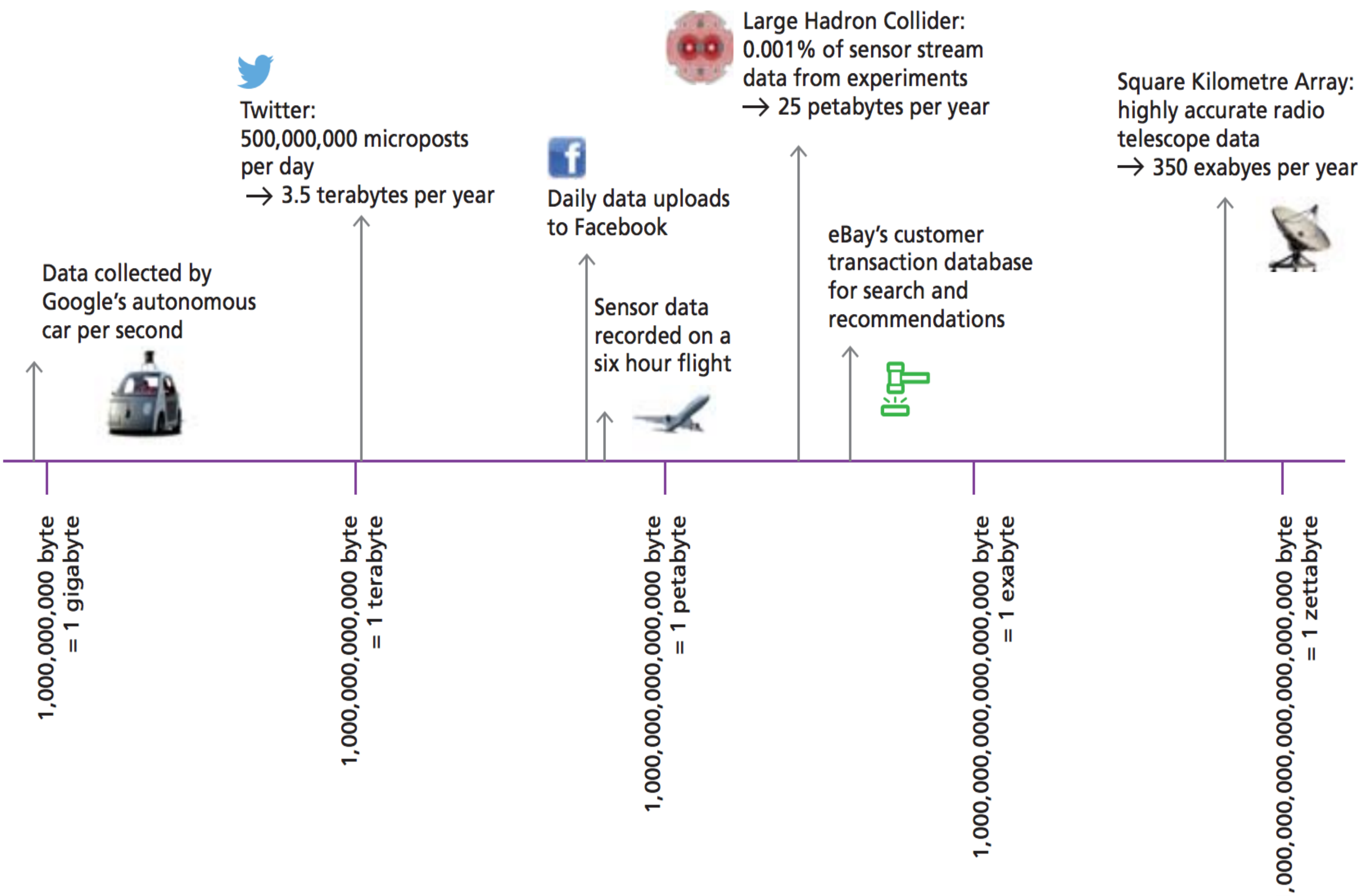
“is high-volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making”



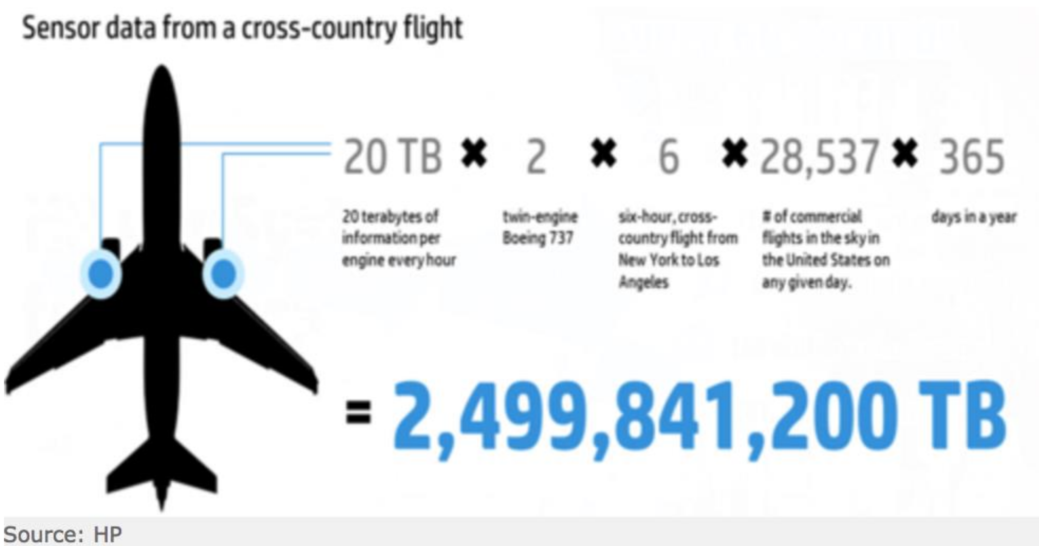
So most of big databases & social networks are likely to be geographical !

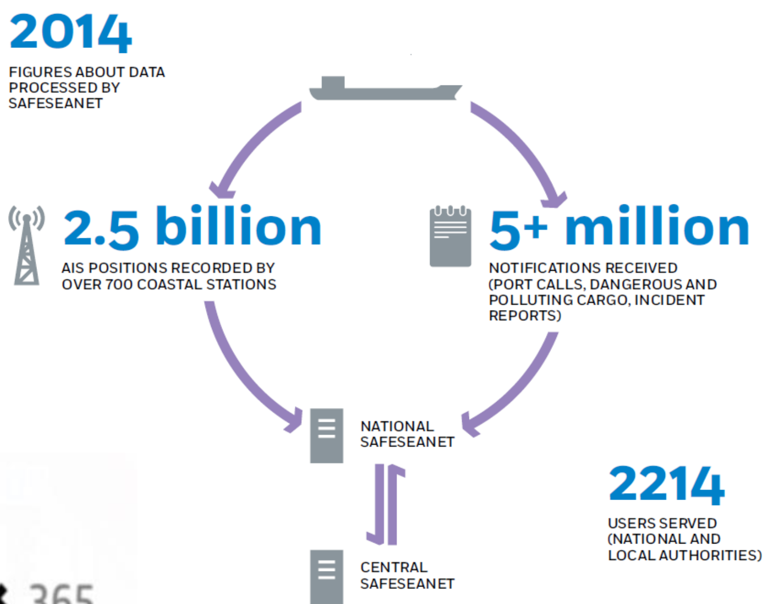

BIG DATA
VOLUME
DATA SIZE
VELOCITY
SPEED OF CHANGE
VARIETY
DIFFERENT FORMS
OF DATA SOURCES
VERACITY
UNCERTAINTY OF
DATA

A few words on the big data context



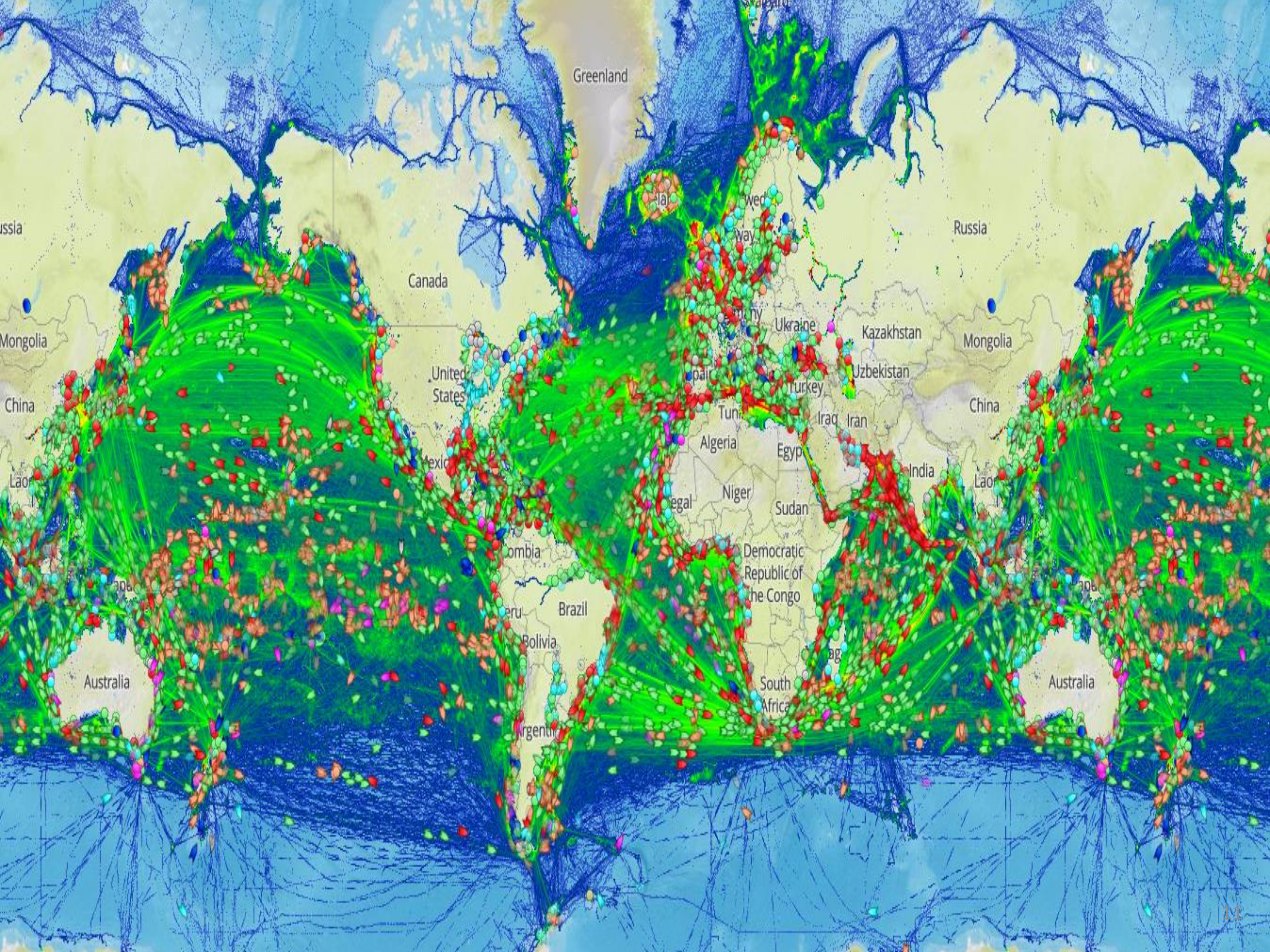
A few words on the big data context

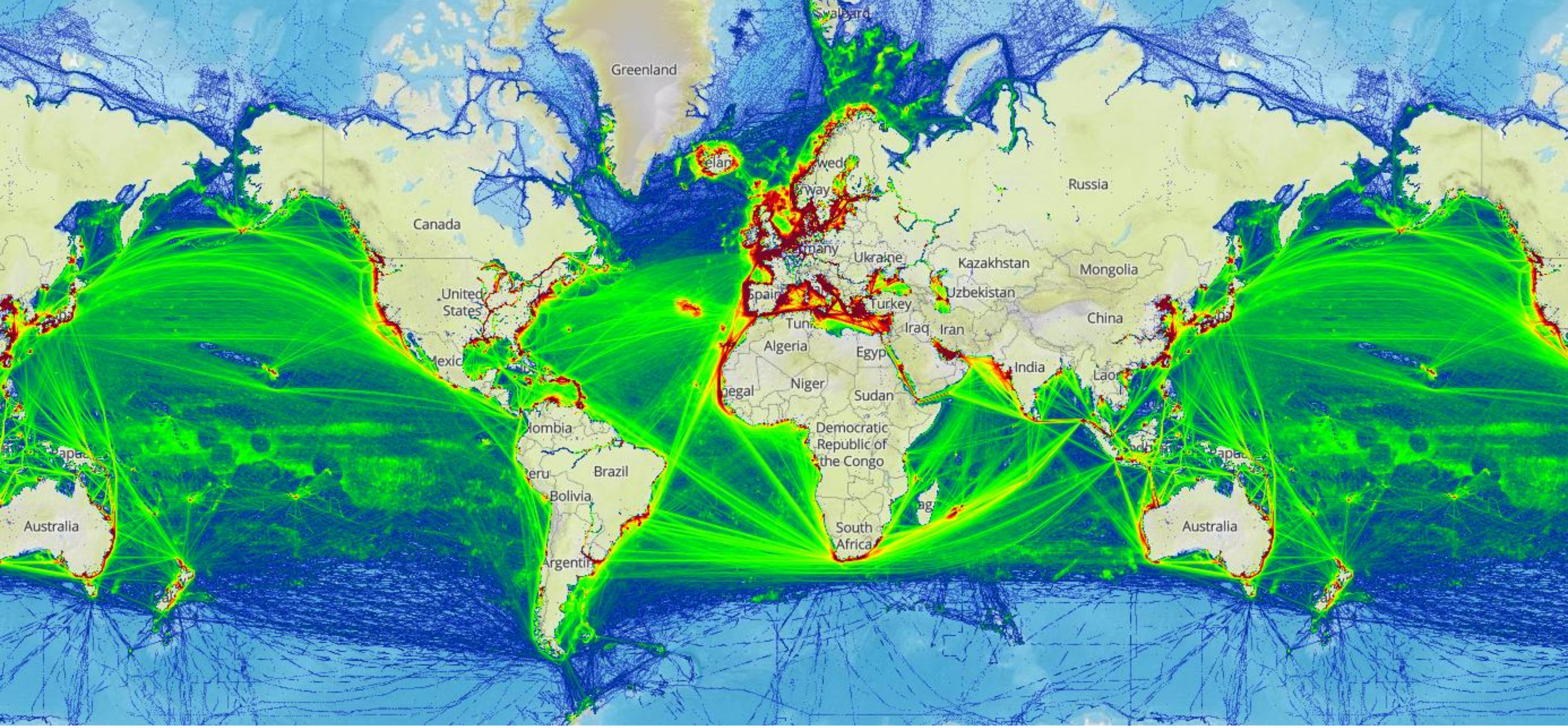




(EMSA, 2014), volume multiplié par 1.5 en 2016

A few words on our big data context …


Greenland
Canada
Russia
Mongolia
Ukraine
Kazakhstan
Mongolia
Uzbekistan
United States
Turkey
China
China
Iraq
Iran
Algeria
Niger
Sudan
India
Brazil
Bolivia
Australia
Australia
South



Worldwide vessels positions acquired by satellites

Online tracking, early recognition of events, and real-time forecast of vessels trajectories are crucial to safety and operations at sea !

Our objective is to review research challenges tied to **the integration, management, analysis, and visualization of objects moving at sea** as well as to provide a few suggestions **for a successful development of maritime forecasting and decision-support systems**

The Maritime domain

Maritime navigation faces several crucial issues due to worldwide traffic increase, staff reduction, piracy & terrorist risks

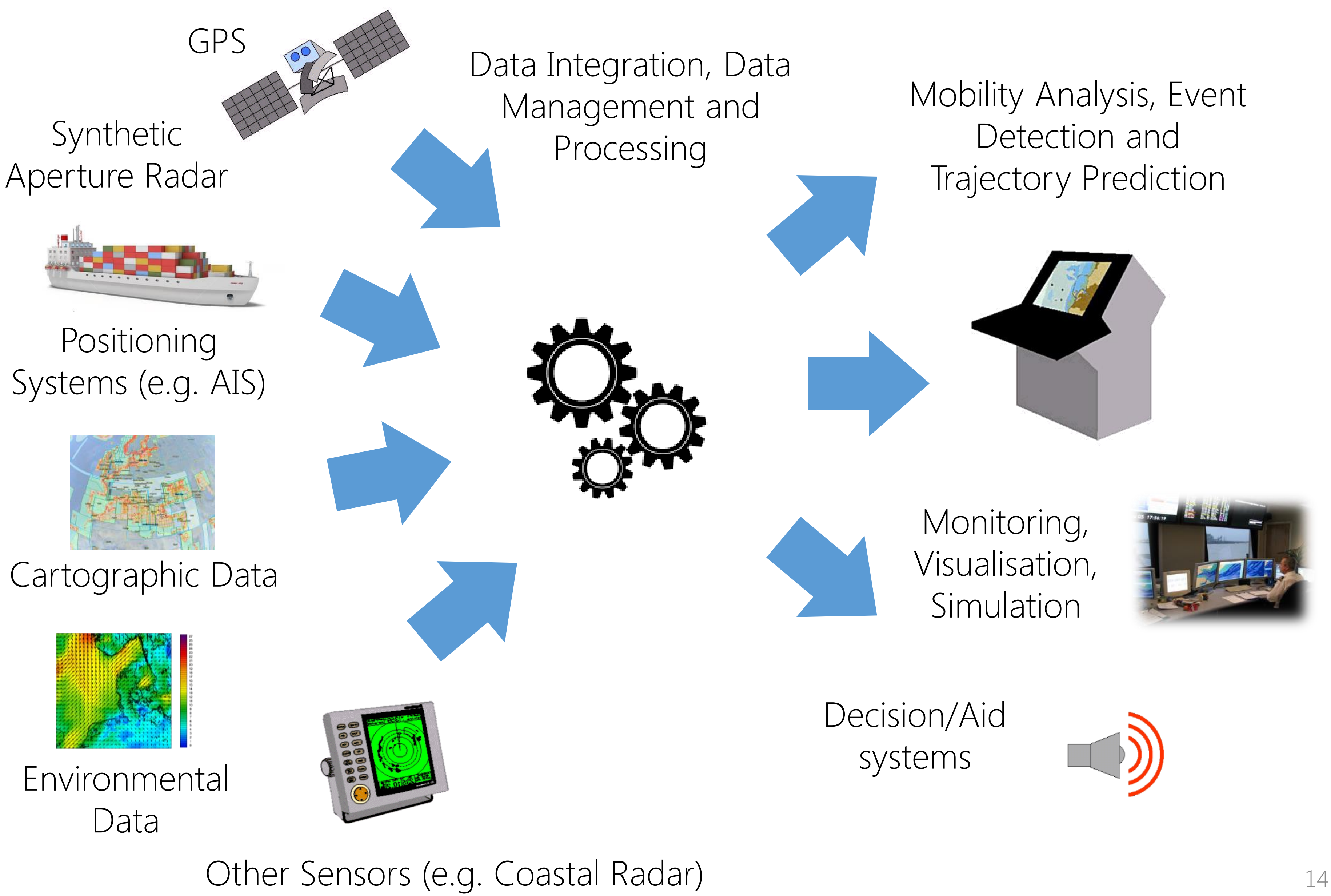
Maersk Containership Grounds in Southern Italy
January 10, 2017

Five Dead, Six Missing After MSC Containership Collides with Fishing Vessel Off Ecuador
December 19, 2016

Up to 80,000 Trout Escape After Cargo Ship Crashes Into Fish Farm in Denmark
October 11, 2016

Source: gcaptain.com

Heterogeneous data integration challenges

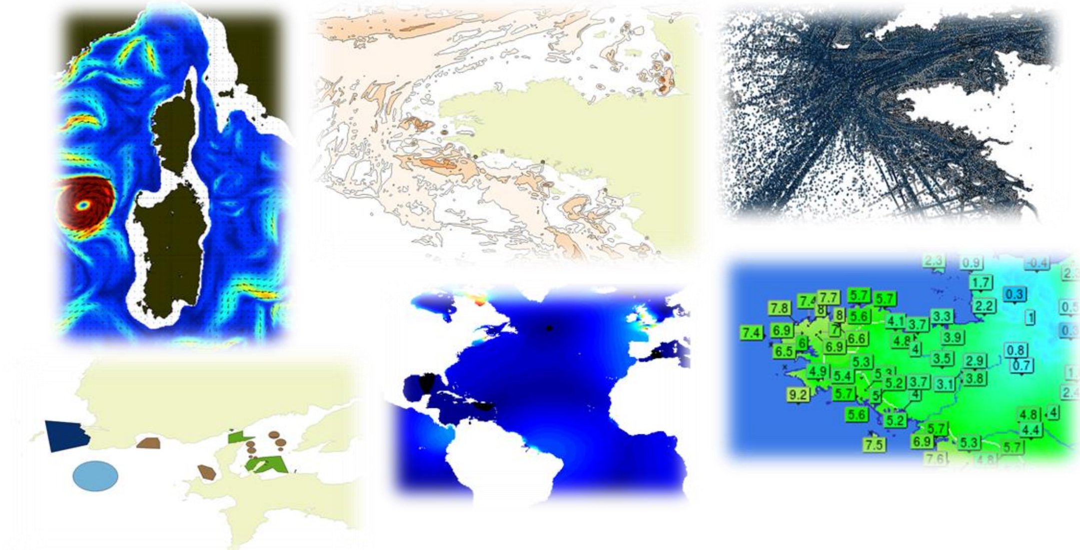


ST data processing: big data challenges !

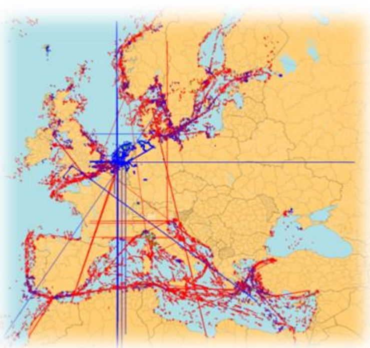
Variety

Volume and Velocity

15 million messages/day (14 800 000 AIS, 54 000 VMS, 31 000 LRIT), from 79 000 vessels [EMSA, 2015]

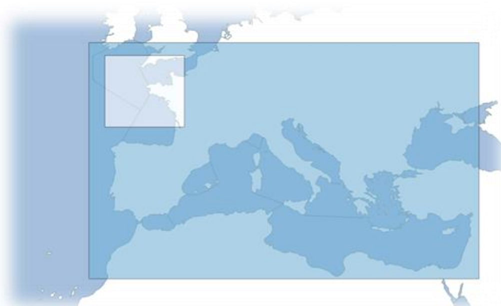


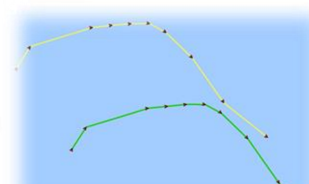
Historical & aggregated data, geographical & environmental data, contextual data, meta information



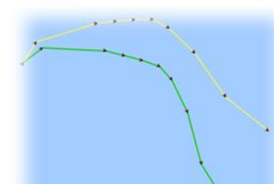
Noisy and error-prone date due to receivers limited coverage, positioning devices switch-off

Veracity Issues





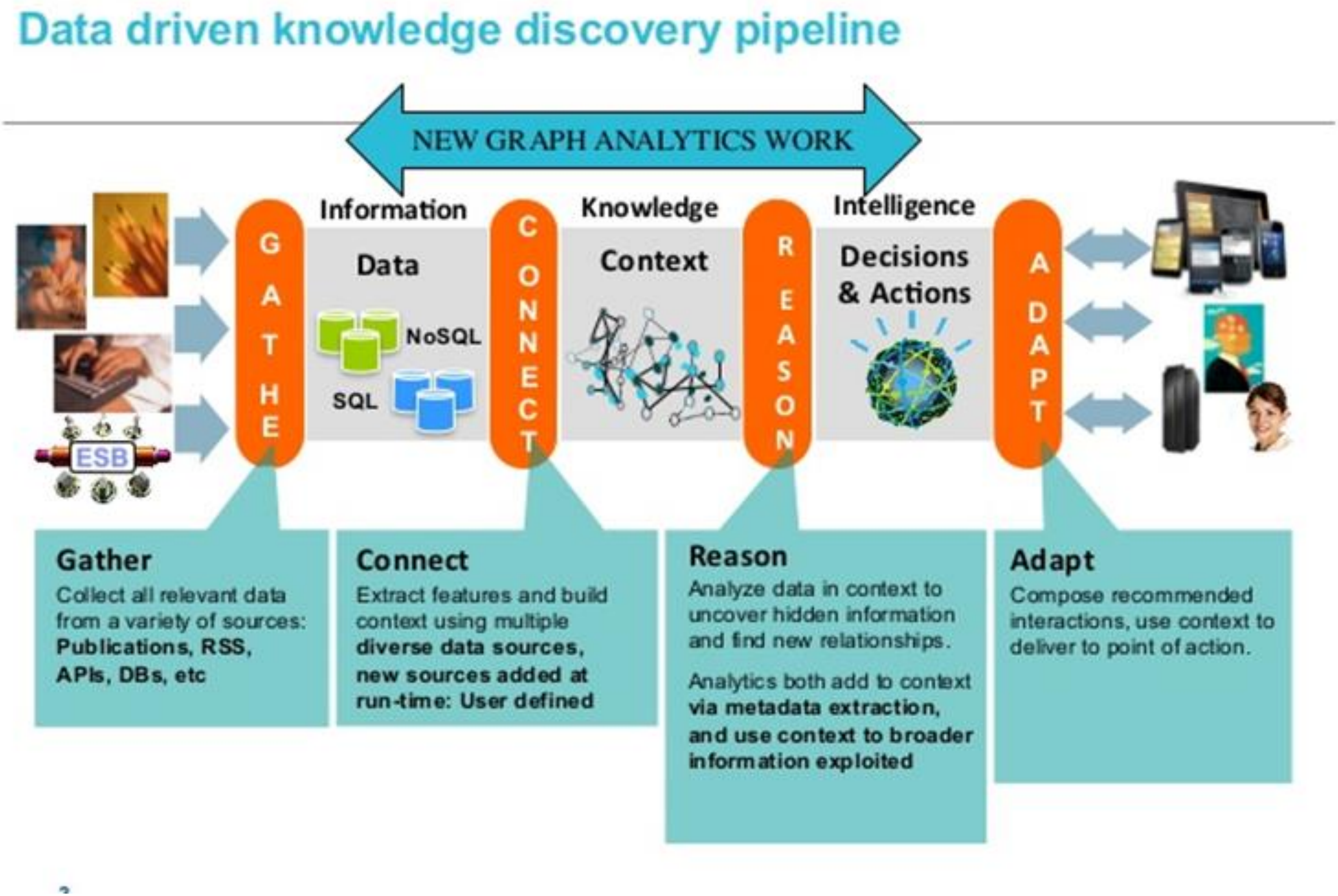
Trajectory translation



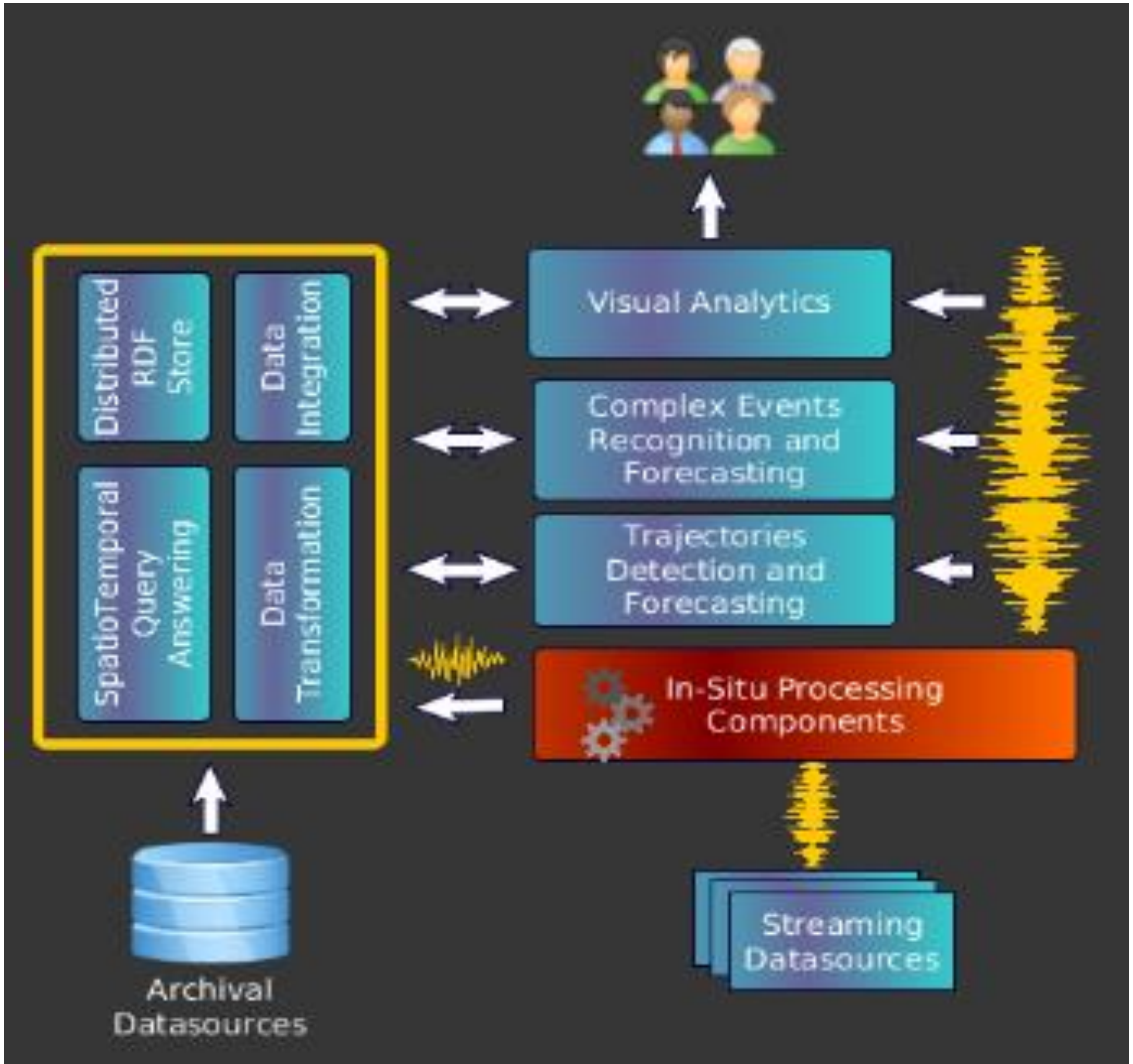
Trajectory rotation

Multi-scale assessment with pseudo-synthetic labelled data

datAcron project "philosophy"



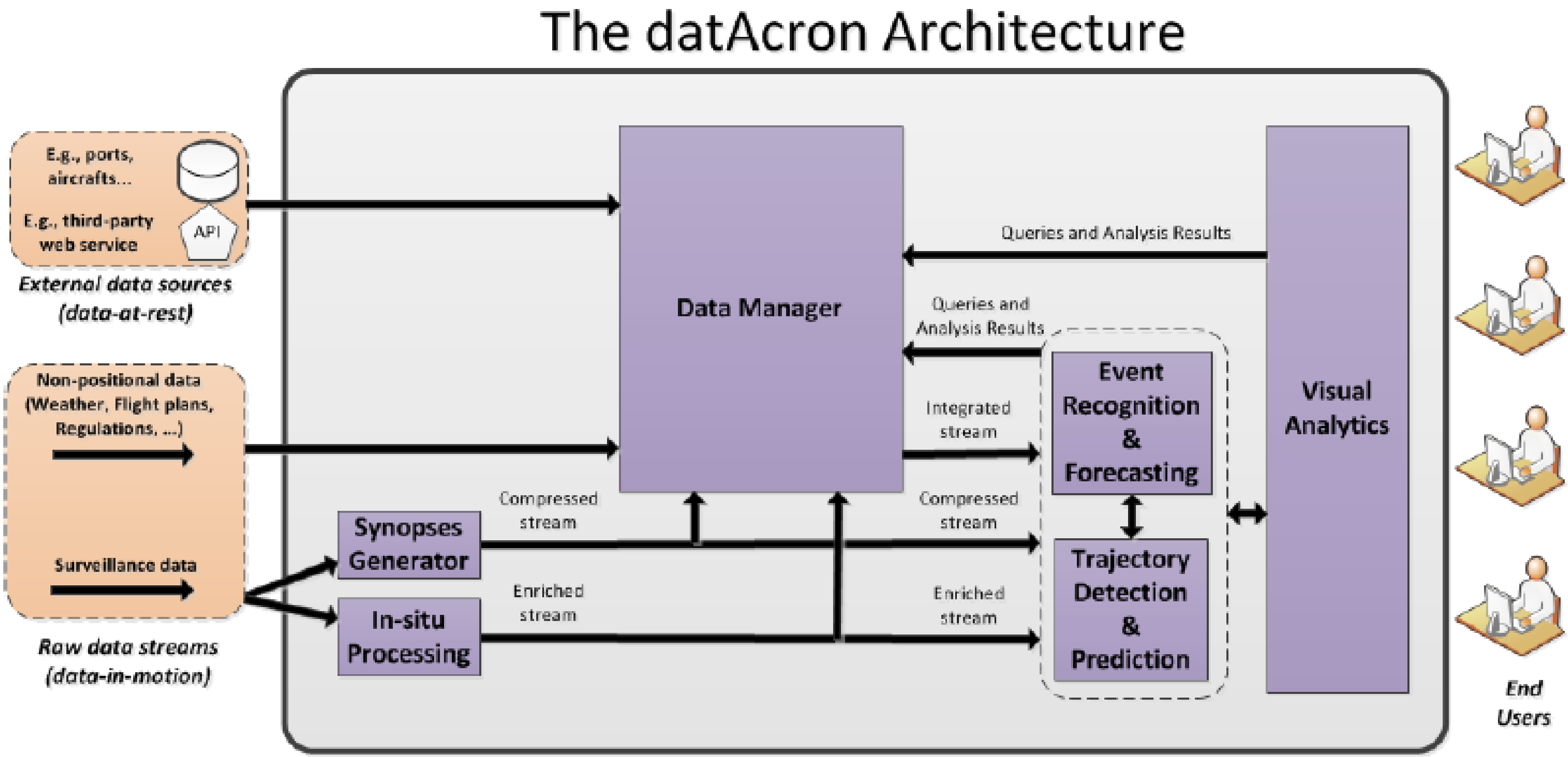
Towards an integration of very large maritime trajectory data



Challenges ahead:

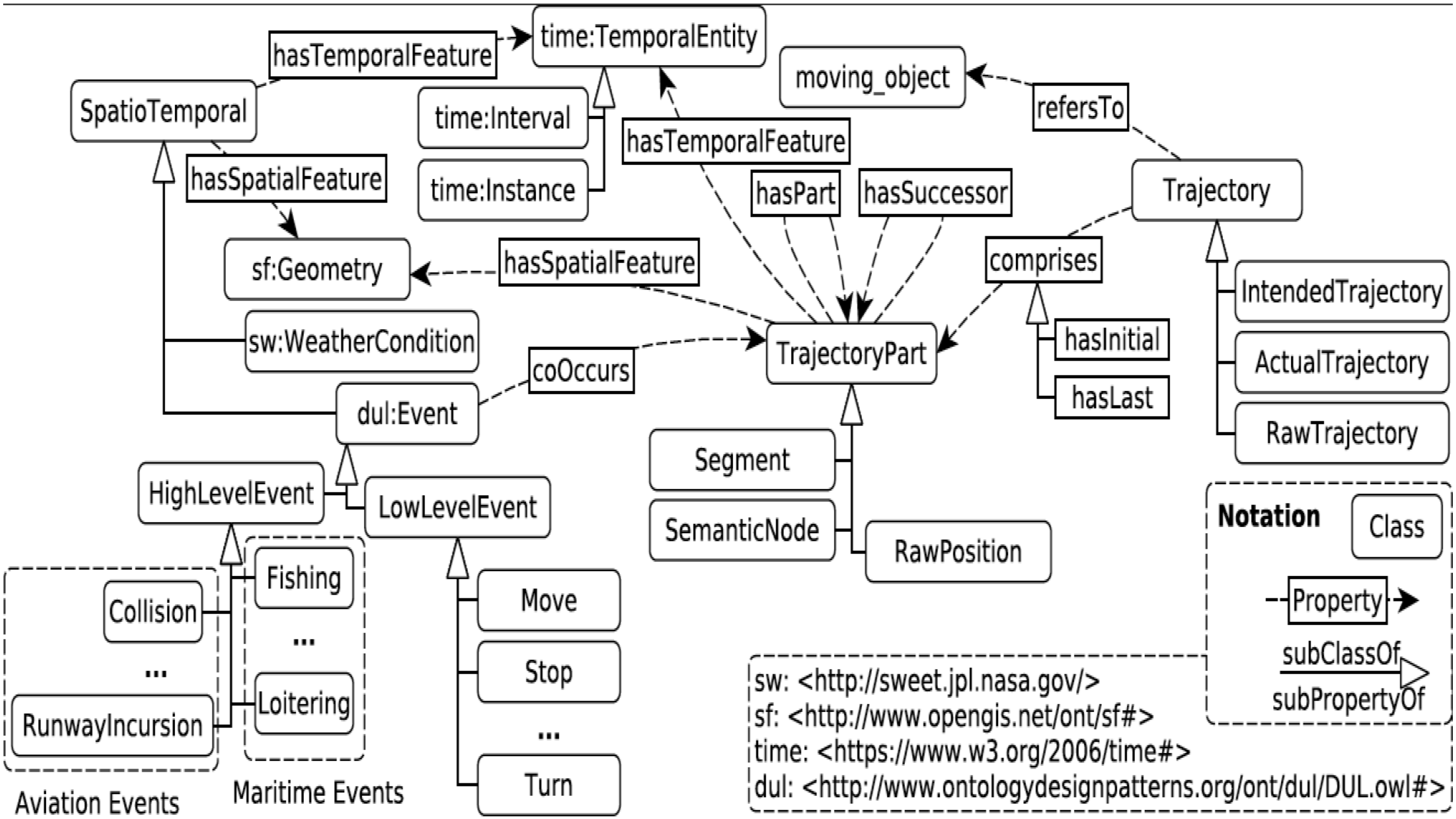
- **Integration of in-situ streaming data**
- **Trajectories detection and forecasting**
- **Recognition and identification of complex events**
- **Development of visual analytics interfaces for maritime experts and decision-makers.**

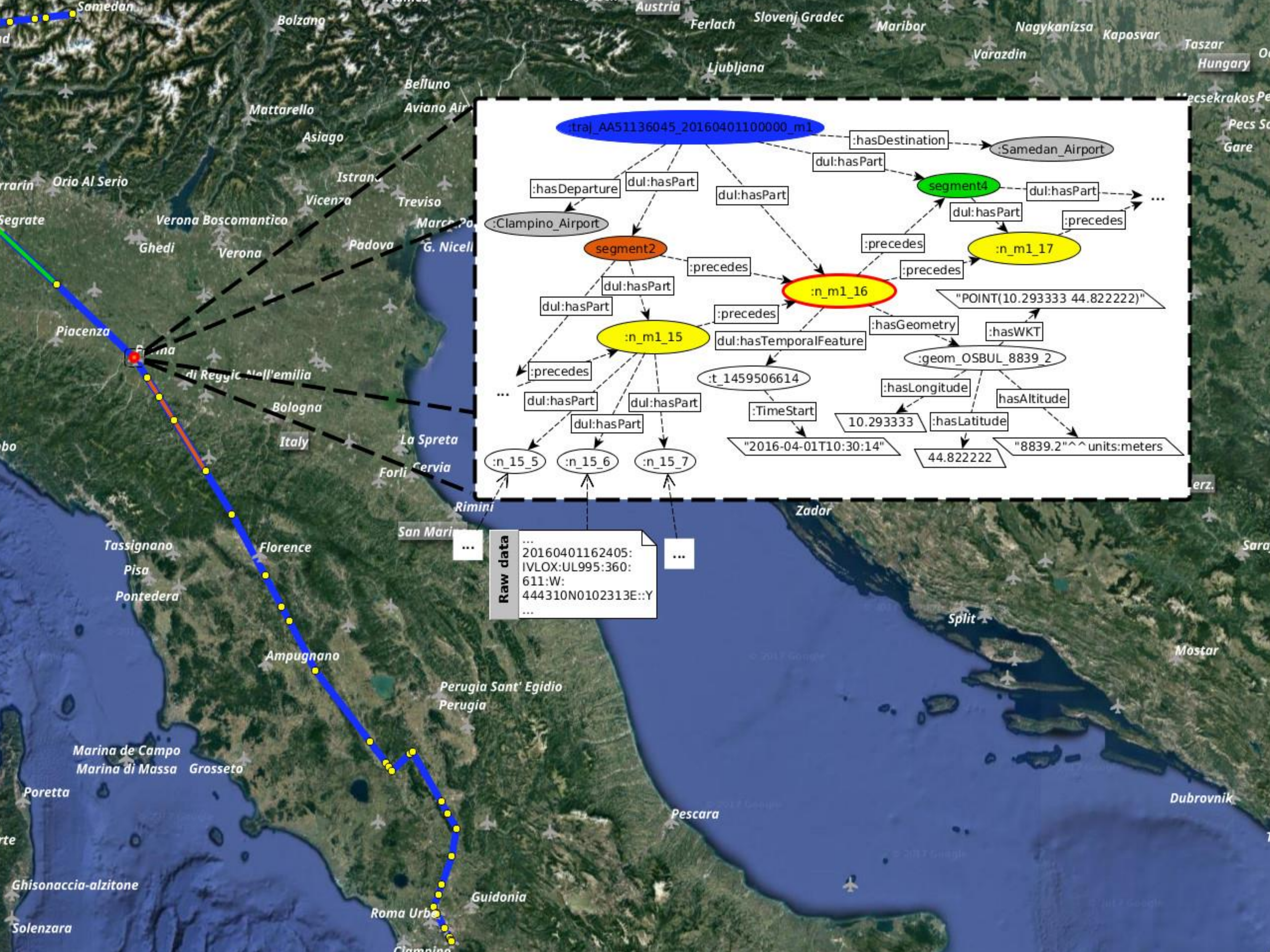
Towards an integration of very large maritime trajectory data

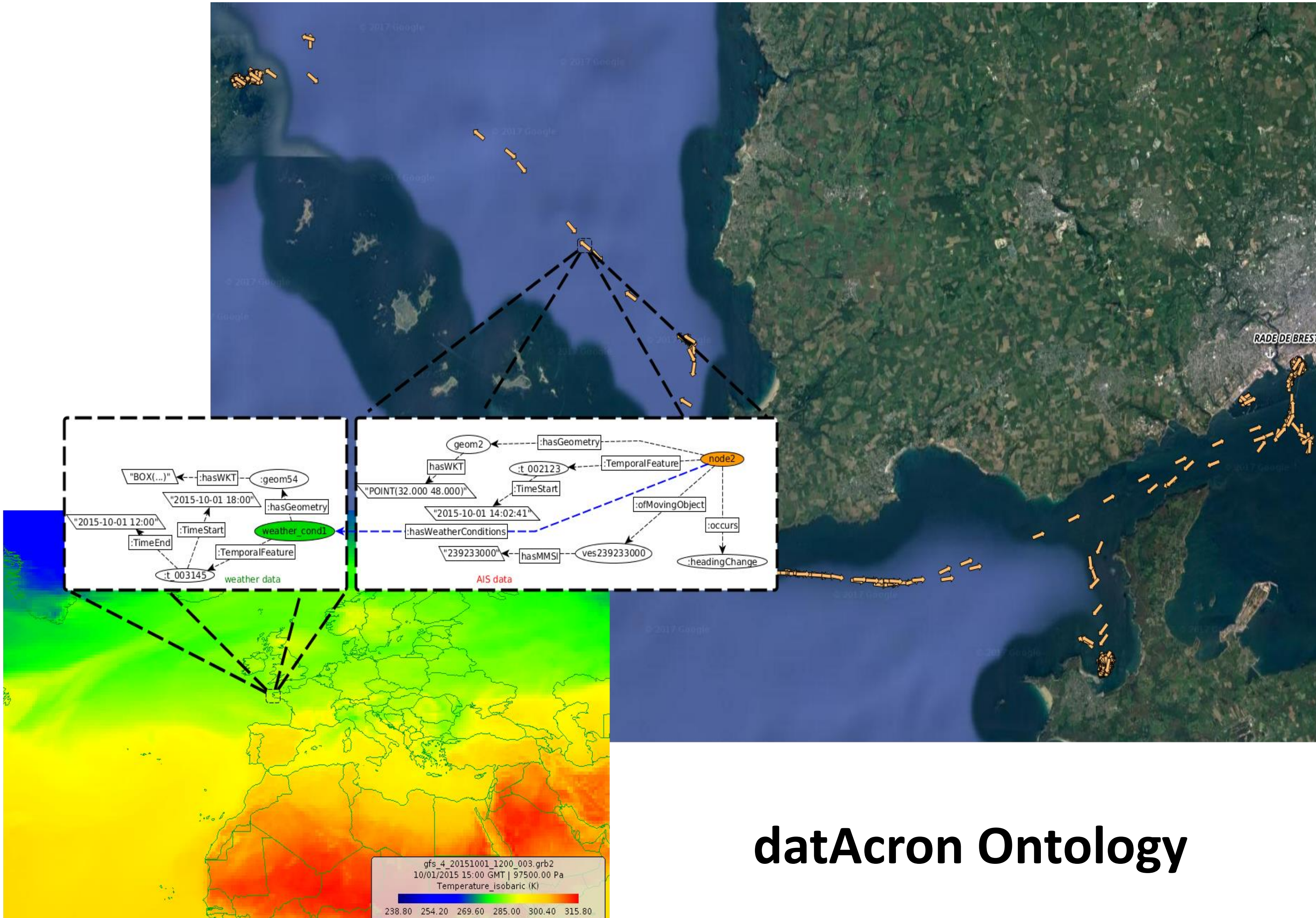


Four main data streams:
- Raw streams of surveillance data
- Compressed stream (=trajectory synopses)
- Enriched stream (low level events)
- Integrated stream (synopses linked with other data)

datAcron Ontology



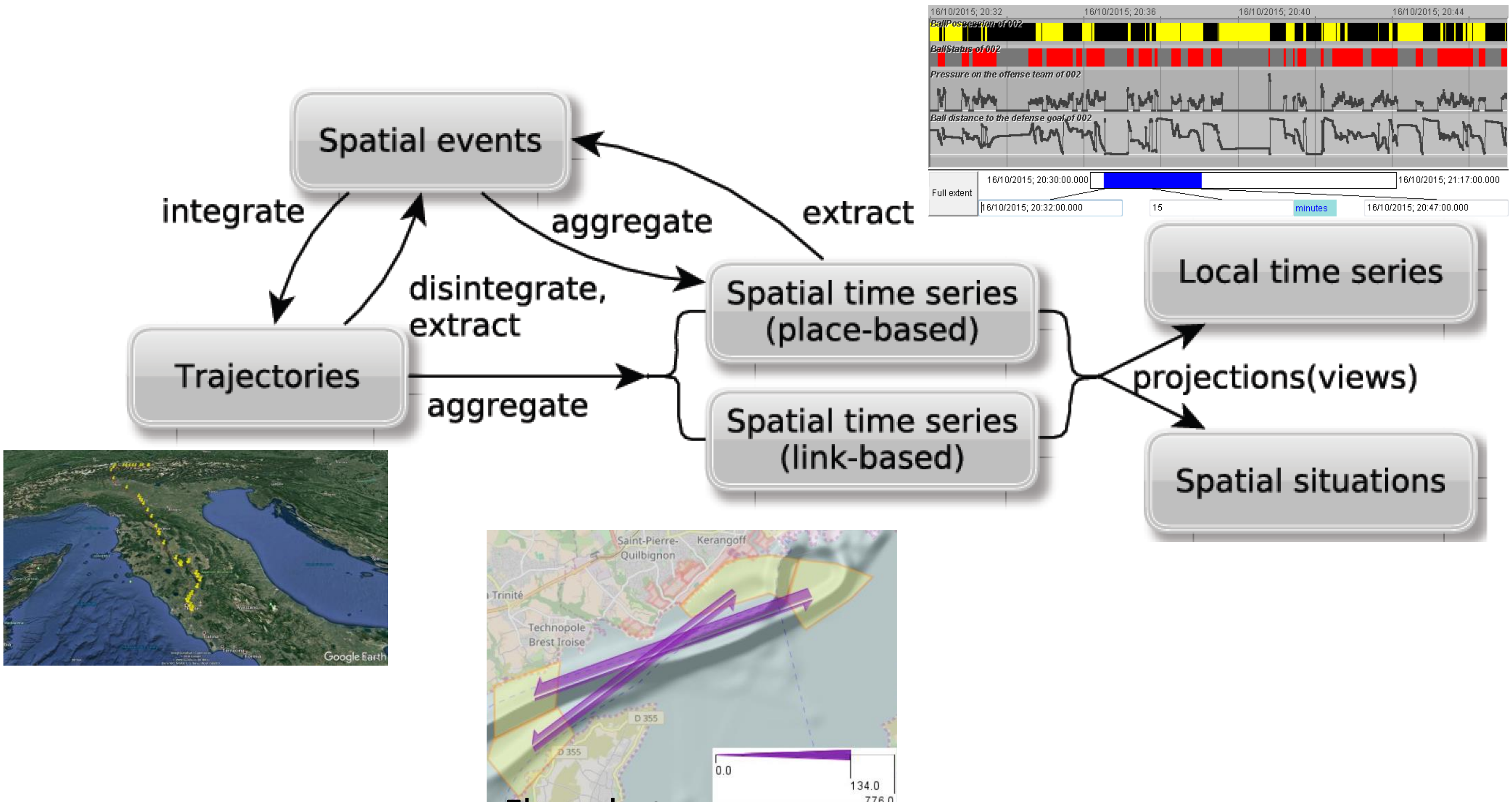

:traj_AA51136045_20160401100000_m1
:hasDestination
:Samedan_Airport
dul:hasPart
:hasDeparture
:Clampino_Airport
segment2
segment4
:precedes
:n_m1_16
:n_m1_17
:n_m1_15
"POINT(10.293333 44.822222)"
:hasGeometry
:hasWKT
dul:hasTemporalFeature
:geom_OSBUL_8839_2
:t_1459506614
:hasLongitude
hasAltitude
:TimeStart
10.293333
:hasLatitude
"2016-04-01T10:30:14"
44.822222
"8839.2"^^units:meters
:n_15_5
:n_15_6
:n_15_7
...
Raw data
...
20160401162405:
IVLOX:UL995:360:
611:W:
444310N0102313E::Y
...
Samedan
Bolzano
Austria
Ferlach
Slovenj Gradec
Maribor
Nagykanizsa
Kaposvar
Taszar
Hungary
Varazdin
Ljubljana
Belluno
Mattarello
Asiago
Orio Al Serio
Istrana
Vicenza
Treviso
Verona Boscomantico
Ghedi
Verona
Padova
Piacenza
Bologna
Italy
La Spreta
Cervia
Forli
Rimini
San Marino
Zadar
Tassignano
Florence
Pisa
Pontedera
Ampugnano
Perugia Sant' Egidio
Perugia
Split
Mostar
Marina de Campo
Marina di Massa
Grosseto
Poretta
Pescara
Dubrovnik
Ghisonaccia-alzitone
Guidonia
Roma Urbe
Solenzara



datAcron Ontology

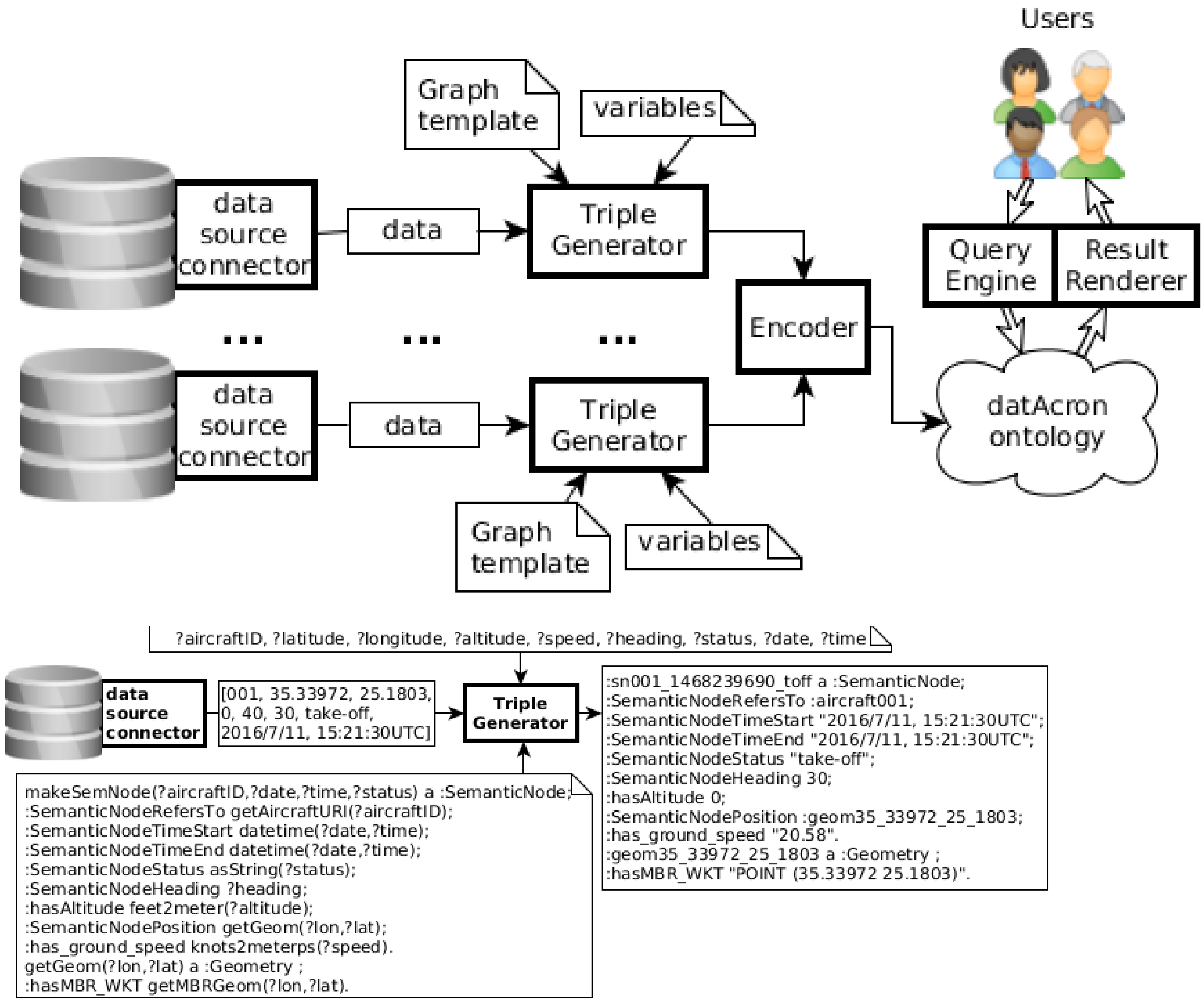
Visual analytics: Data transformation

Events per time interval, and counts

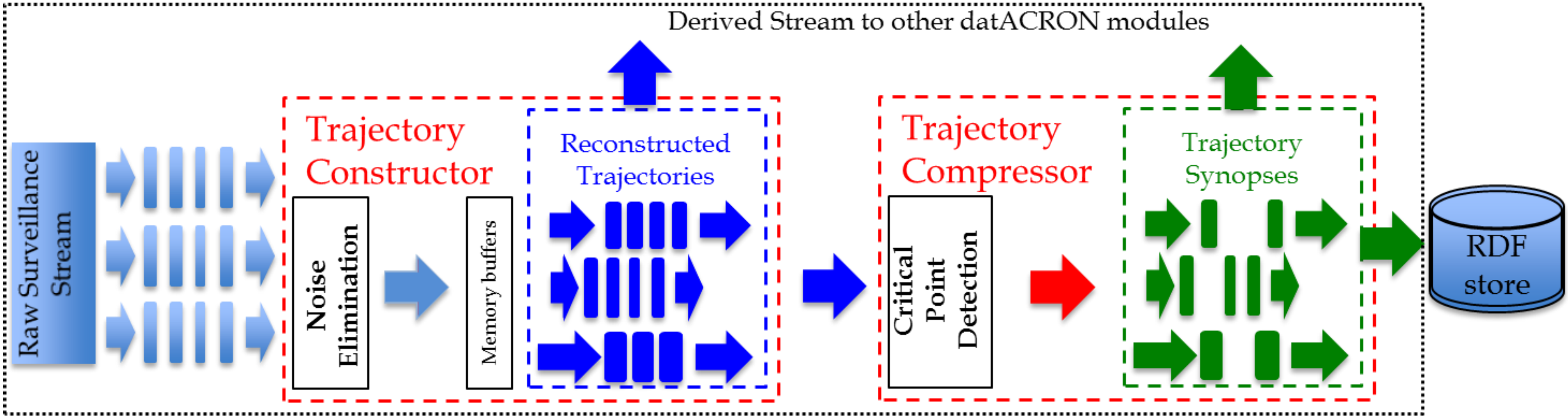


Flows between areas

Populating the datAcron ontology



In-Situ data processing: Towards data synopses (1)

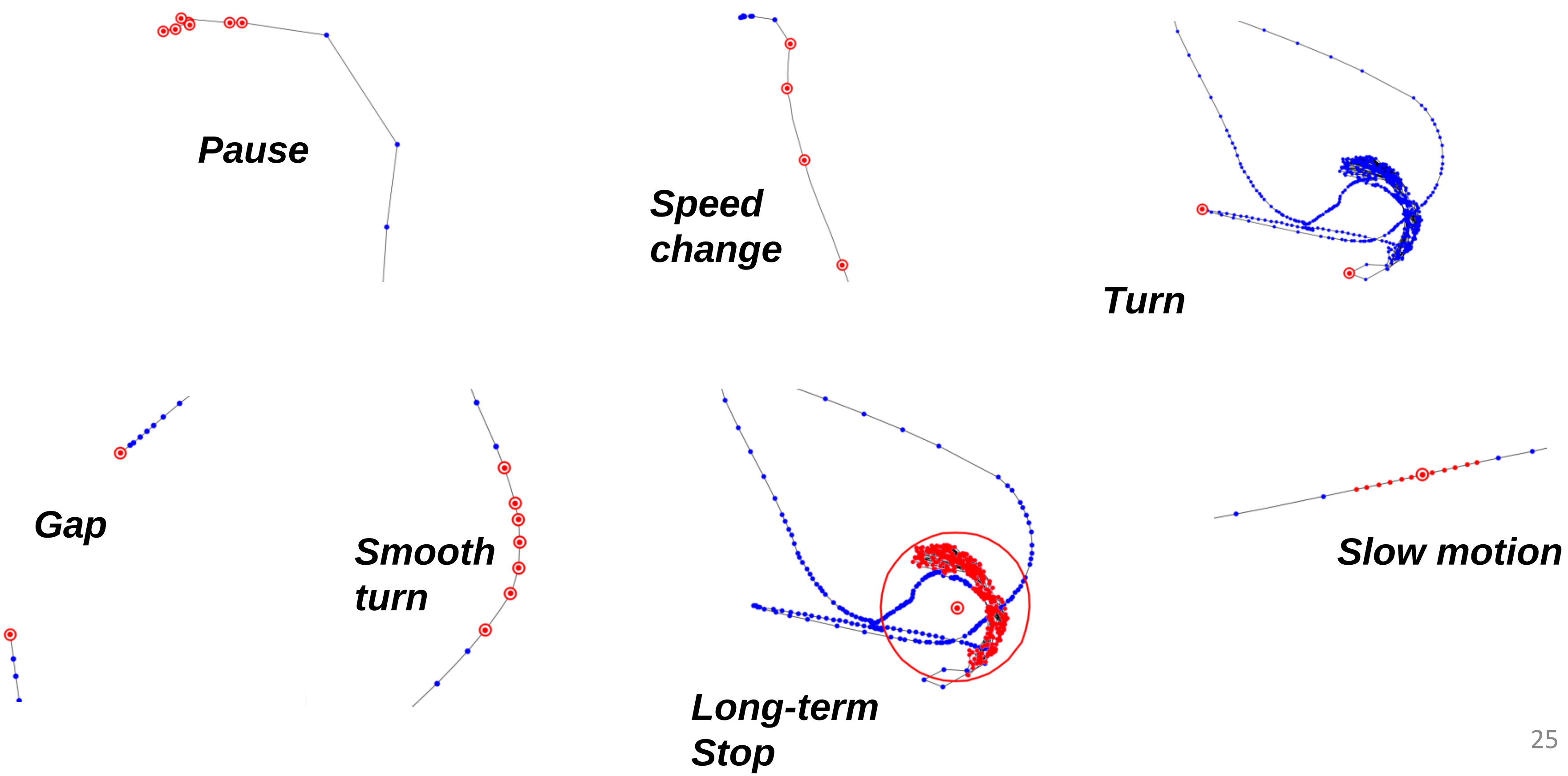


Cross-streaming data integration with contextual information (e.g., weather data) enables further effectiveness in detection and predictive analytics.

The challenge is to address high levels of data compression without compromising the accuracy of the prediction / detection components.

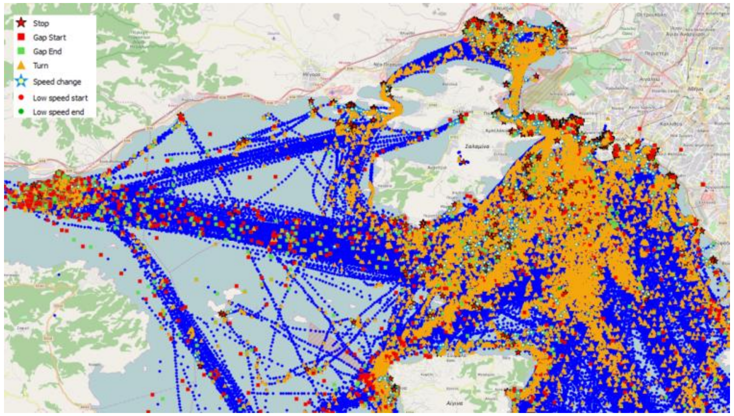
In-Situ data processing: Towards data synopses (2)

- **Shortening the time needed for detecting patterns of interest within a single- or cross-streaming process and for instance using distributed stream processing architectures.**
- **Integration of streaming-based and contextual data**
- **Search for data synopses**



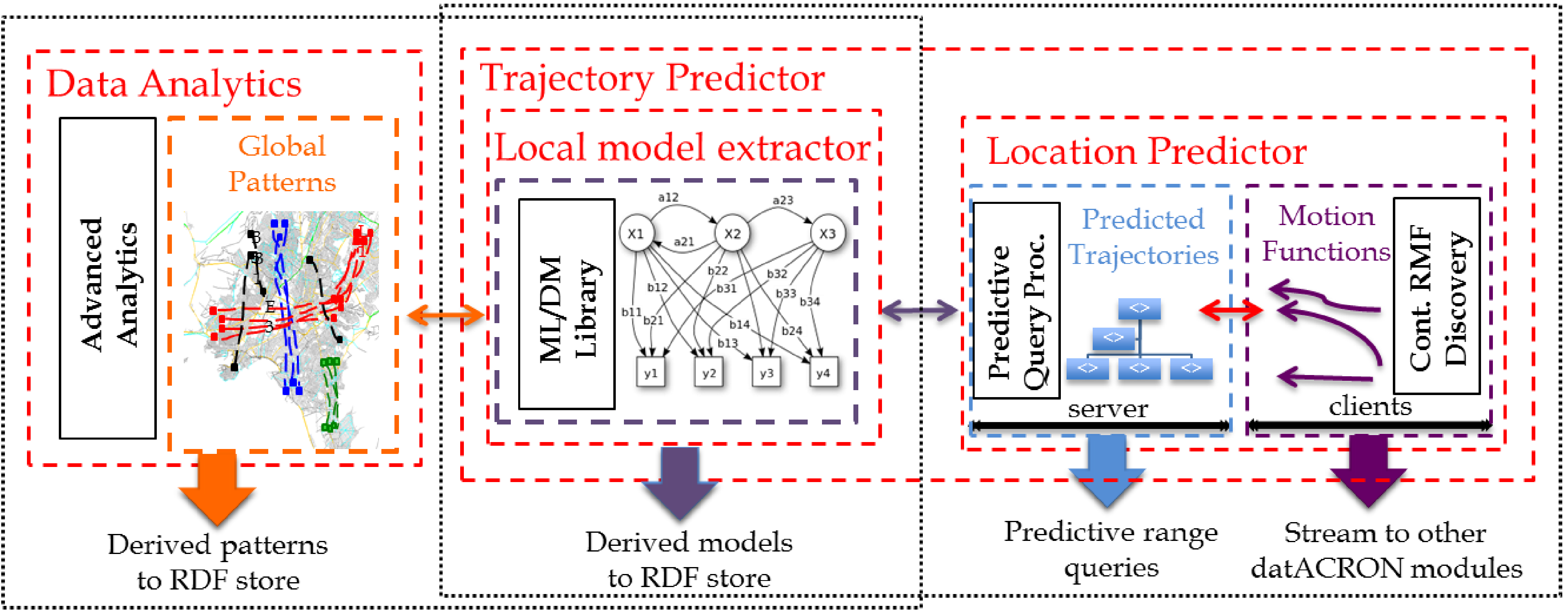
In-Situ data processing: Towards data synopses (3)

- Empirical study over a real AIS dataset – *courtesy IMIS Hellas*
  - Period: *June-August 2009* for *6425 vessels*
  - Size: ~23 GB – 168,240,595 timestamped positions
  - On average, each vessel reports *once every two minutes*
  - *Mean arrival rate*: 50 positions/sec
    → very low for a data stream system!
    + *artificially increased rates*
  - Calibrated empirical parameterization for mobility tracking



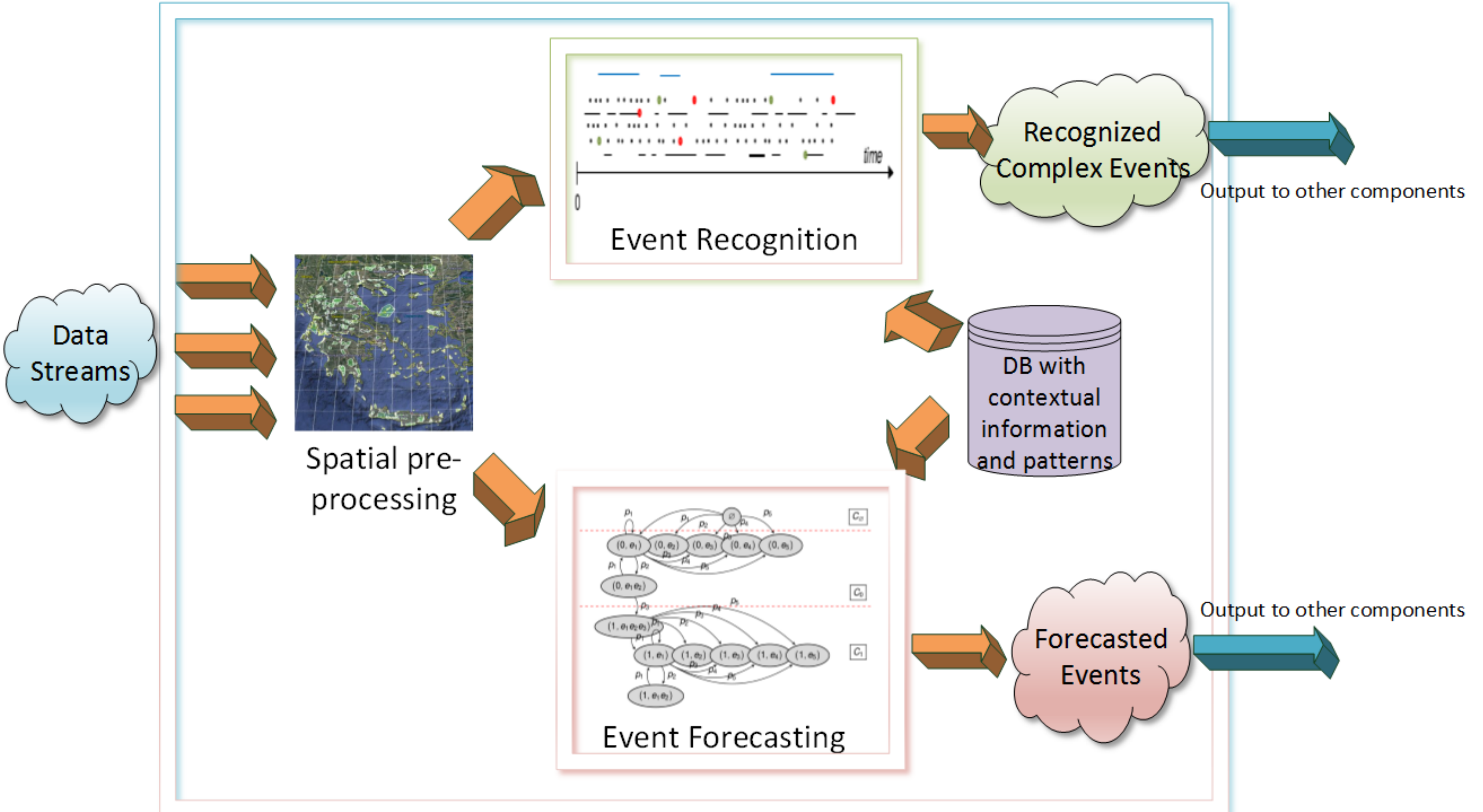
- Lessons learned:
  - *Timeliness*: critical points issued within msec for various window sizes
  - *Compression*: up to 98% reduction compared with raw AIS data
  - *Quality*: tolerable approximation error (RMSE)
  - *Scalability*: increased data volumes at varying arrival rates

Trajectory detection and prediction



- **Taxonomies and ontologies provide alternatives to bridge the gap between low level data from maritime sensors and maritime domain semantics,**
- **For example to enhance the integration of maritime information, to model ships' behavior, for patterns identification, abnormal behavior detection, and prediction.**

Event pattern detection



The range of possible events of interest is very large, from detecting vessels in distress and collisions at sea to discovering illegal fishing and any other illicit activities occurring at sea such as contrabands and smuggling.

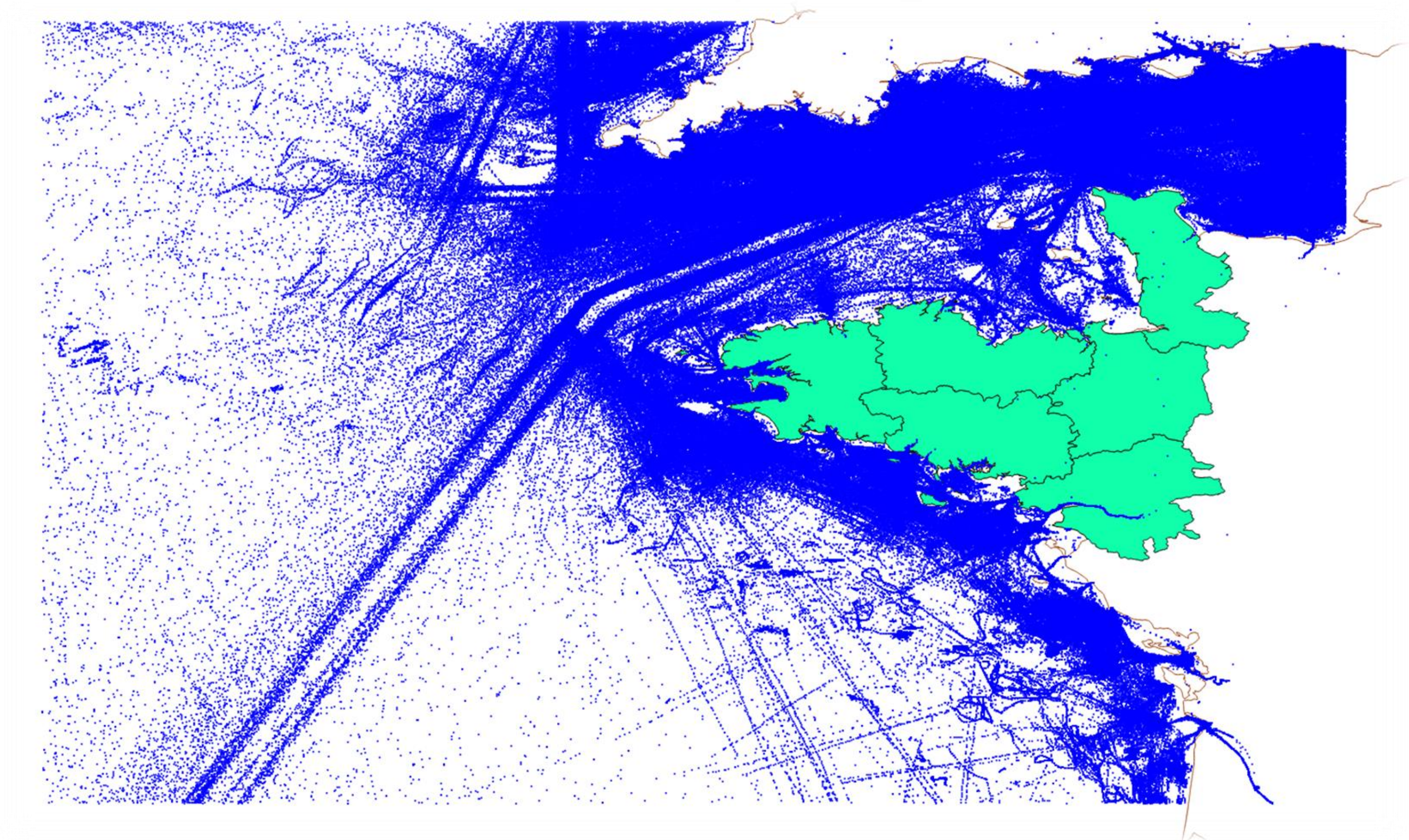
dataCron sample maritime data

Training Dataset Characteristics							
NARI				IMISG			
Number of positions	Time period	Number of different ships	Average velocity	Number of positions	Time period	Number of different ships	Average velocity
19,152,196	October 2015 - March 2016 (6 months)	4802	77 messages per minutes	3,779,626	January 2016 (one month)	4799	104 messages per minutes

+ MMSI country codes, navigational status, detailed list of types (csv)

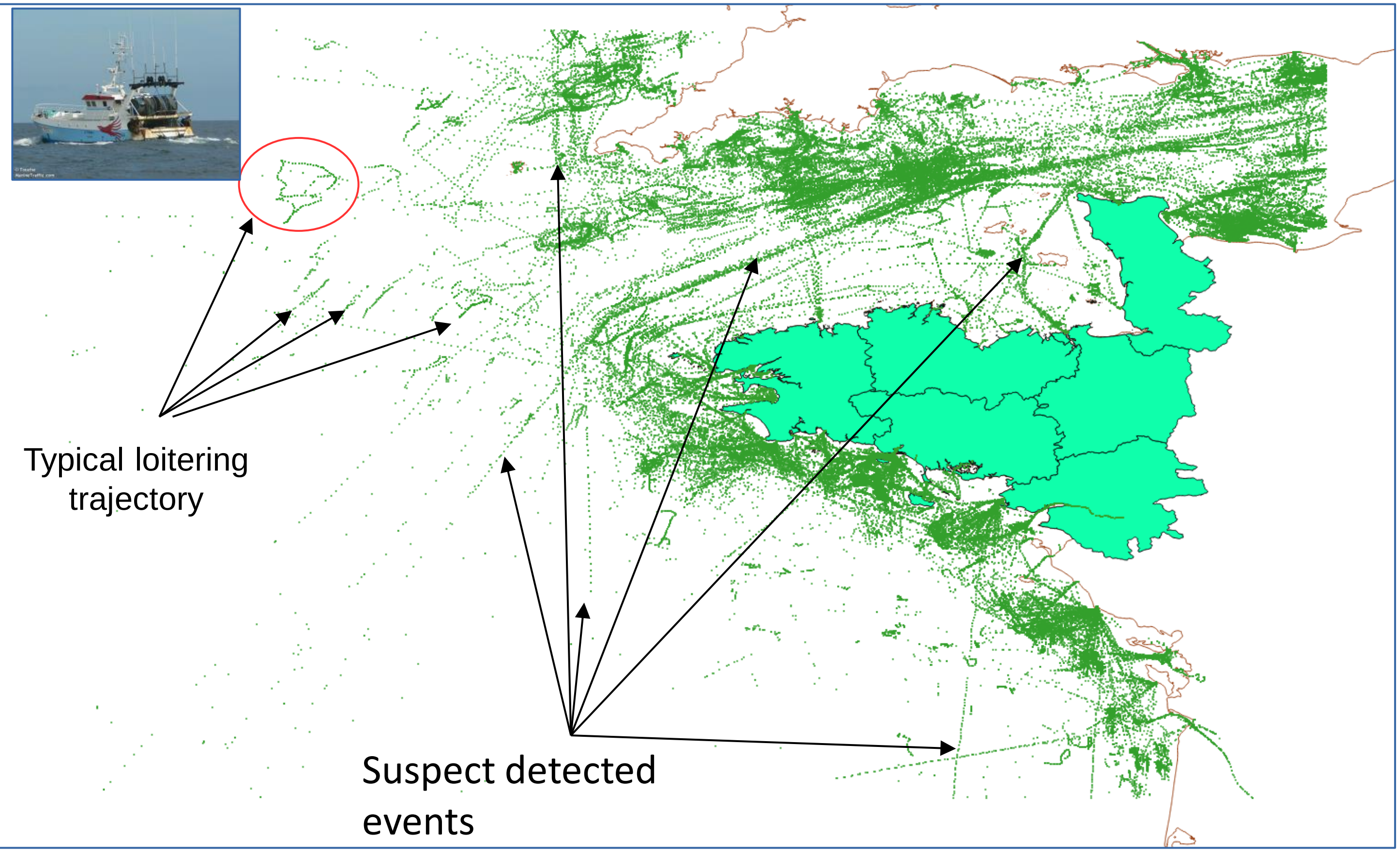
Detecting, predicting, planning… warning

Interactive detection of events



Sample dataset from IMISG, January 2016

Detecting events: Loitering … ?



Detected events based on synopses

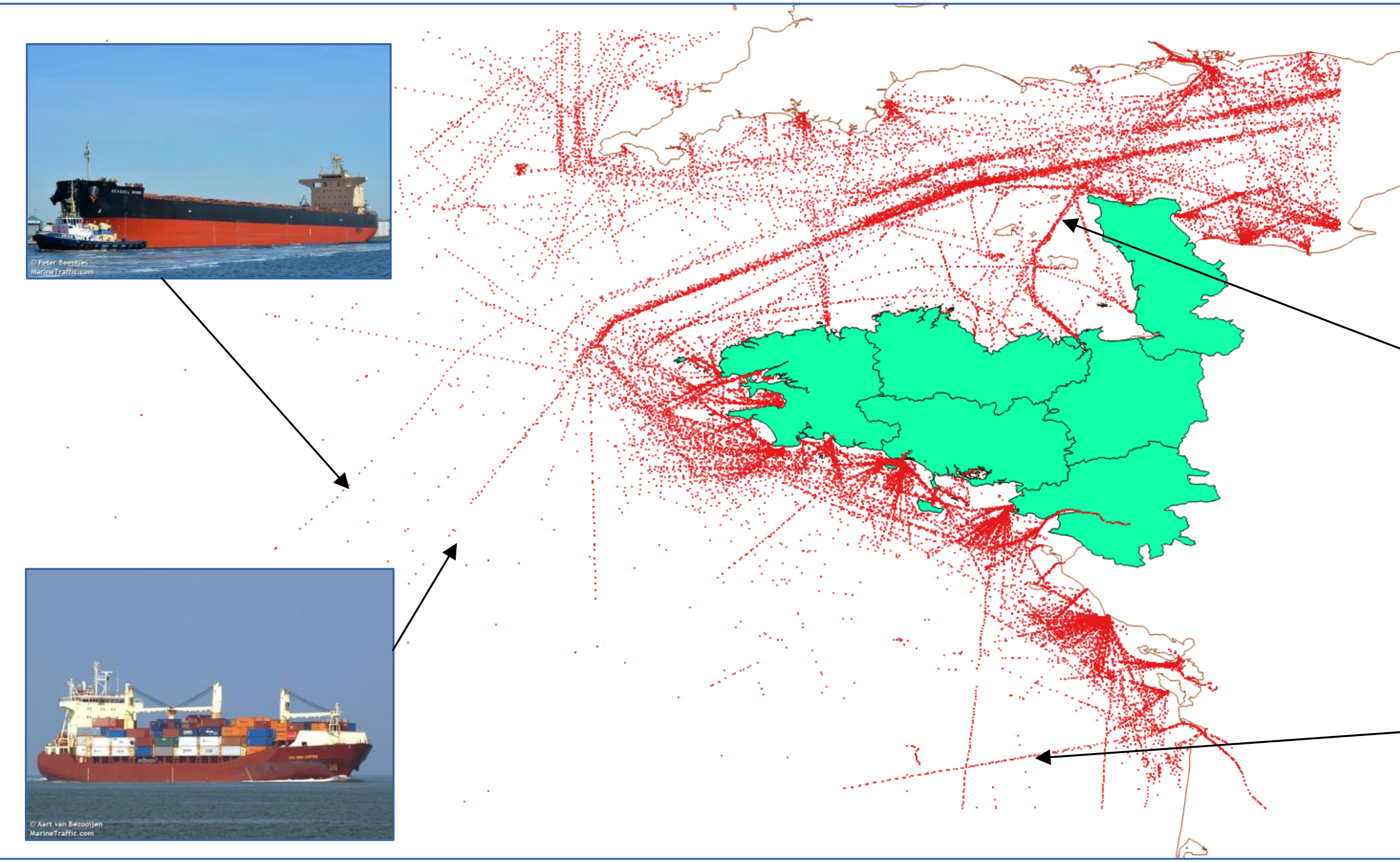
Long range straight trajectories = suspect events

Looks like « commercial ship », not fishing vessels

Raw data from « IMISG data set » based on mmsi / startTime / endTime of detected events

- All detected events based on synopses

Detecting events: Loitering (incoherent speed: > 6 knots)



Loitering speed is about 3-4 knots

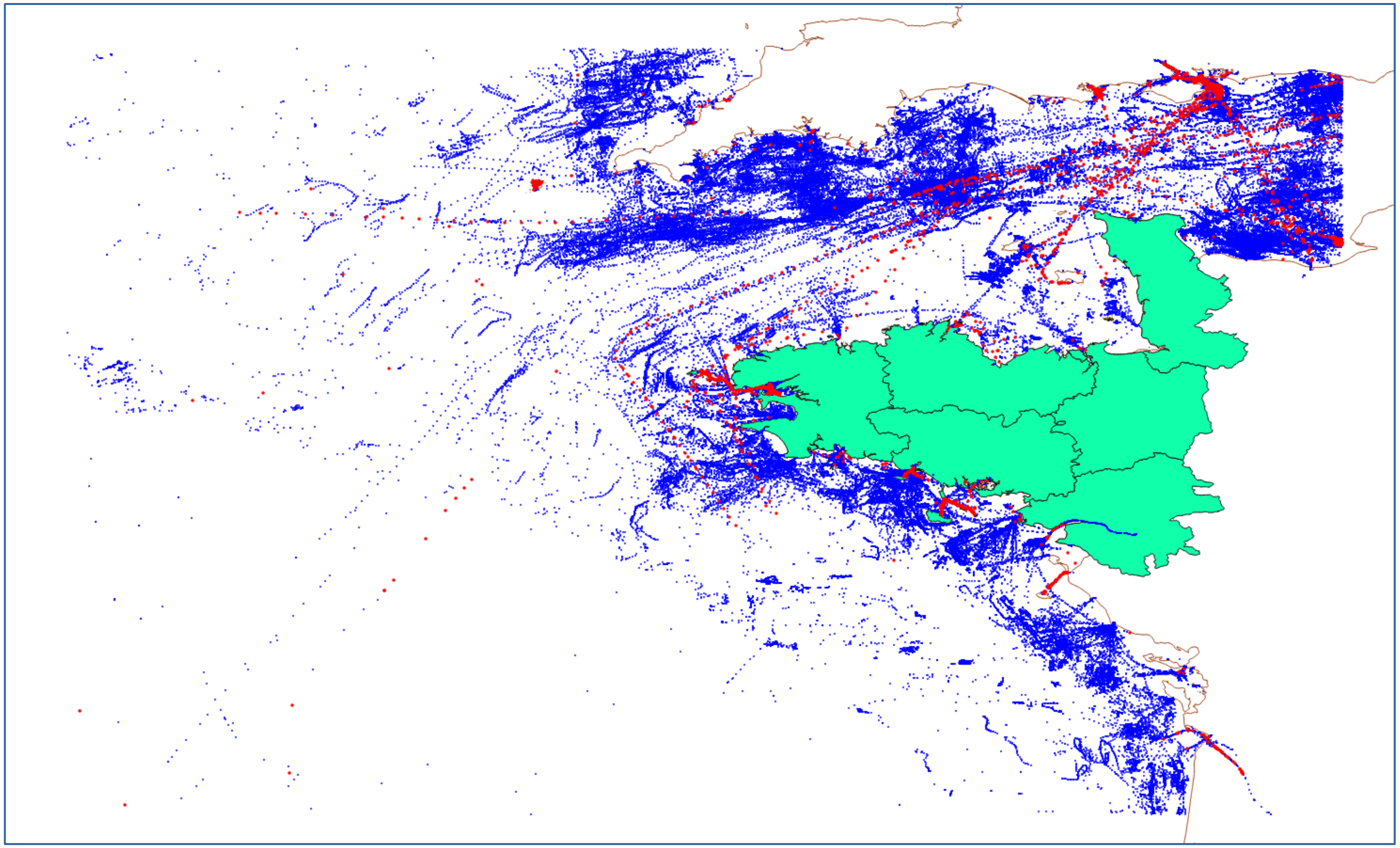
Raw data from « imis data set » based on mmsi / startTime / endTime of detected event

- Filter : speed>6knots

Some detected events are over the « loitering speed »

Long range / high speed => typical cargo-tanker-commercial ships trajectories

Detecting events: speed incompatible with a given area



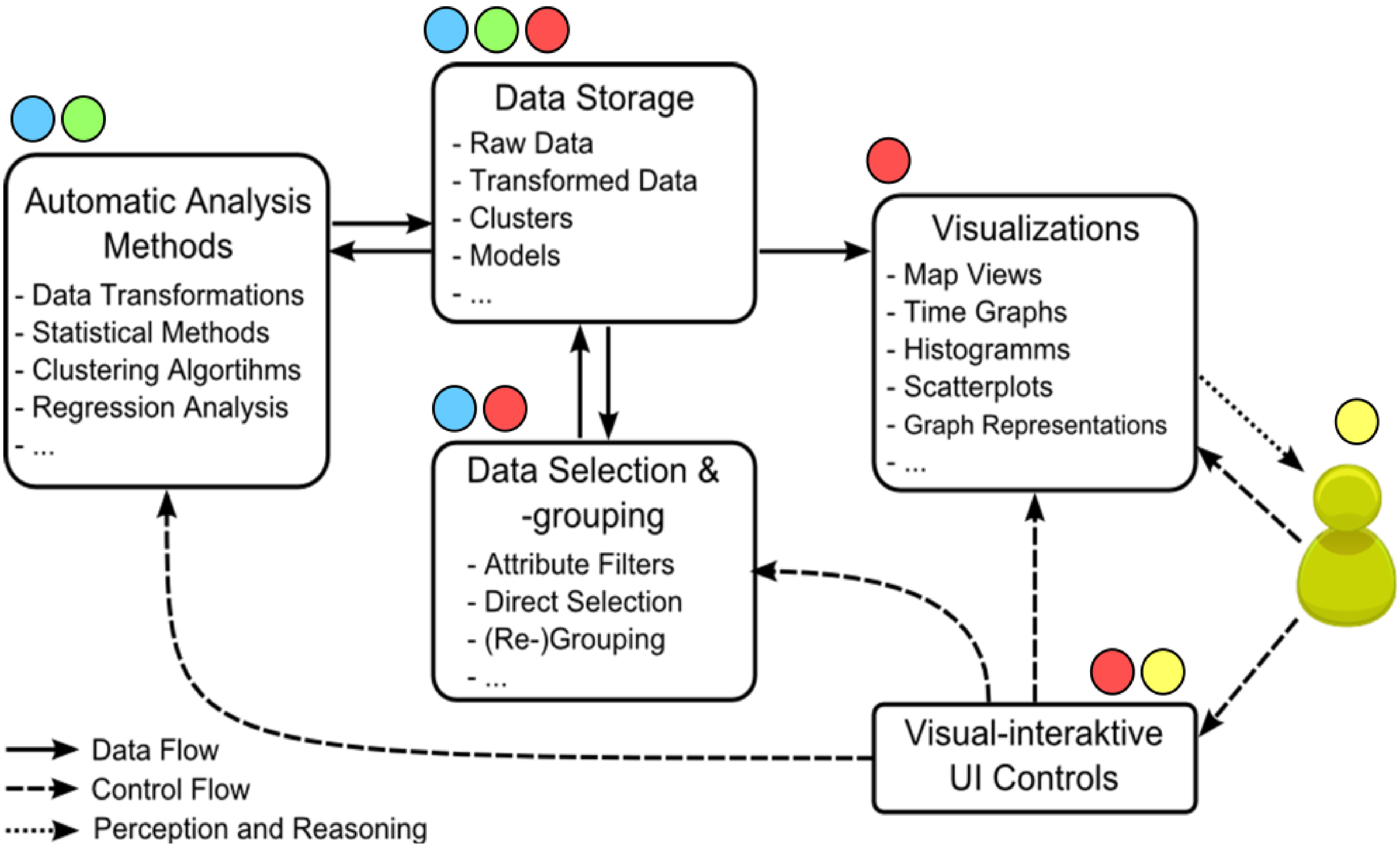
Detected events

Compared with speed transmited in AIS message

In red : speed over 15 knots

What is the criteria θ_{speed} in the detection event algorithm ?

Visual analytics



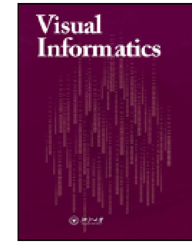
The objective is not only the reconstruction of vessel trajectories and computation of events and BUT also multi-scale visualizations of data and patterns via advanced analytics techniques.

Visual analytics

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

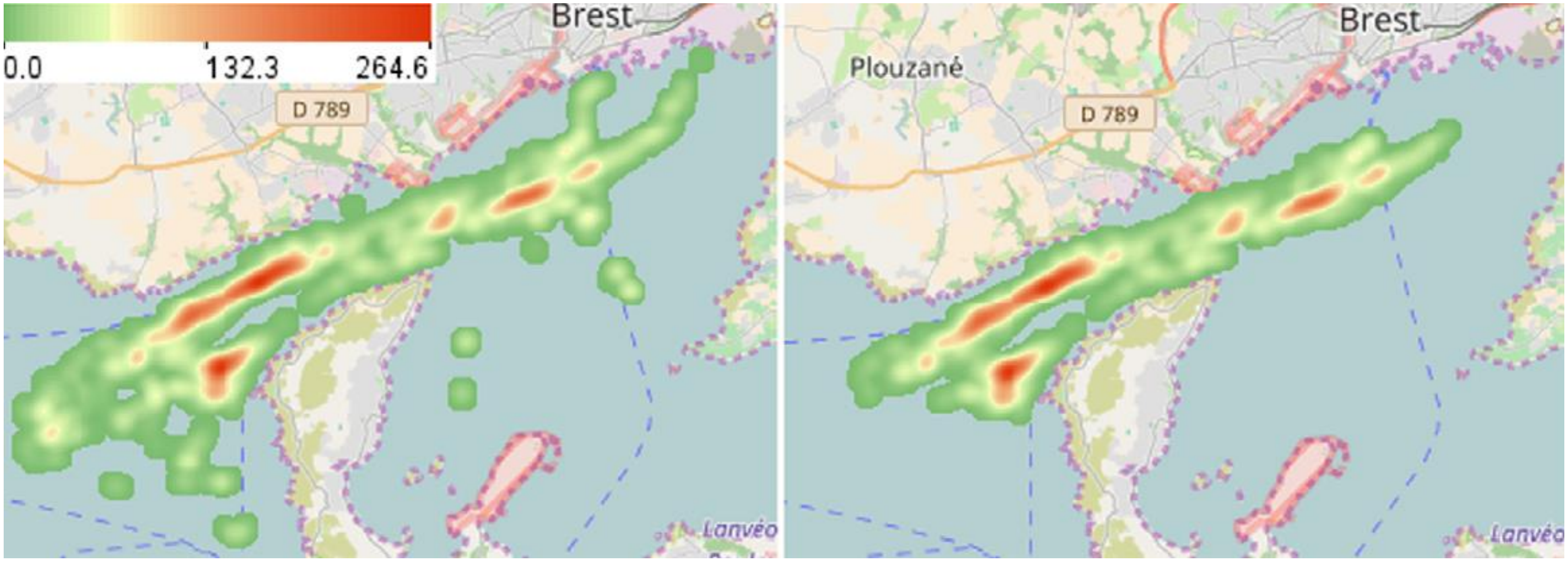


Visual exploration of movement and event data with interactive time masks

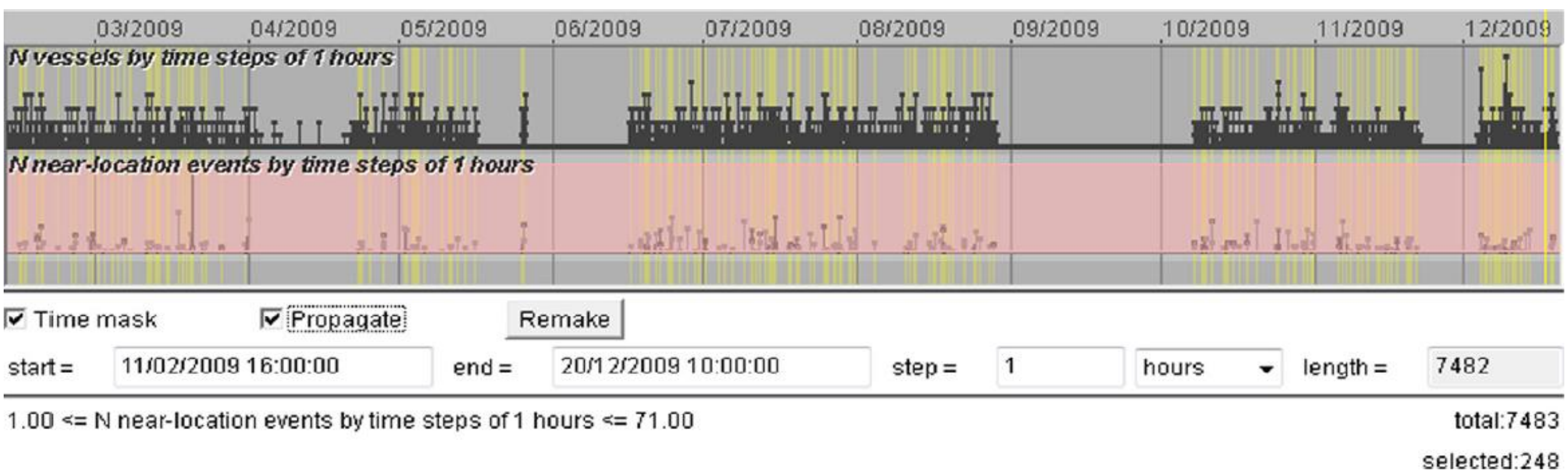
Natalia Andrienko [a,b], Gennady Andrienko [a,b,*], Elena Camossi [c], Christophe Claramunt [d], Jose Manuel Cordero Garcia [e], Georg Fuchs [a], Melita Hadzagic [c], Anne-Laure Jousselme [c], Cyril Ray [d], David Scarlatti [f], George Vouros [g]

[a] *Fraunhofer Institute IAIS, Sankt Augustin, Germany*
[b] *City University London, London, UK*
[c] *NATO Science and Technology Organization, Centre for Maritime Research and Experimentation, Italy*
[d] *Naval Academy Research Institute, France*
[e] *CRIDA - Reference Center for Research, Development and Innovation in ATM, Madrid, Spain*
[f] *Boeing Research & Technology Europe, Spain*
[g] *Department of Digital System, University of Piraeus, Greece*

Interactive detection of events (1)

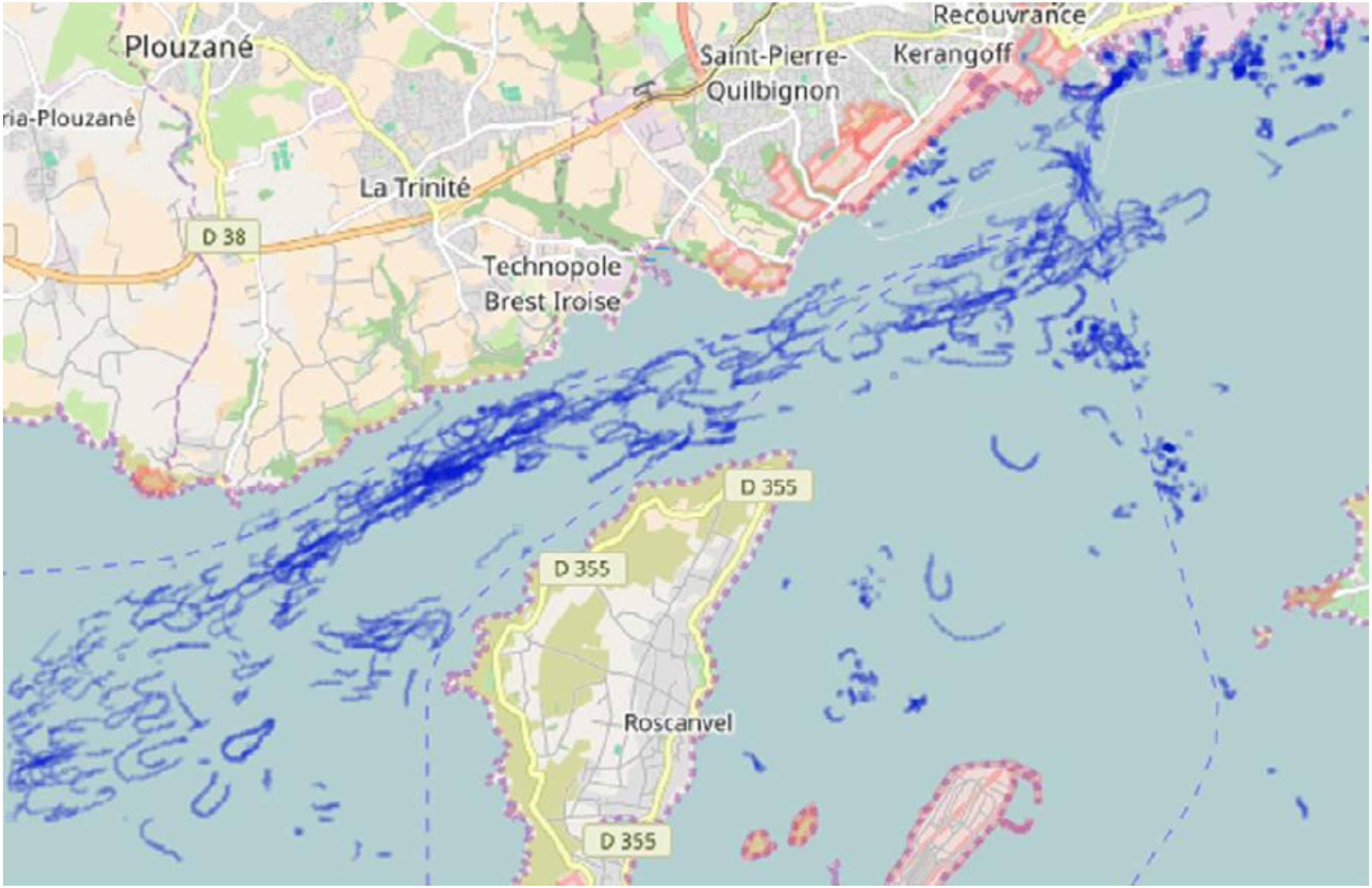


Density of the extracted near-location events (left: all events; right: the events that occurred in the main traffic lanes)



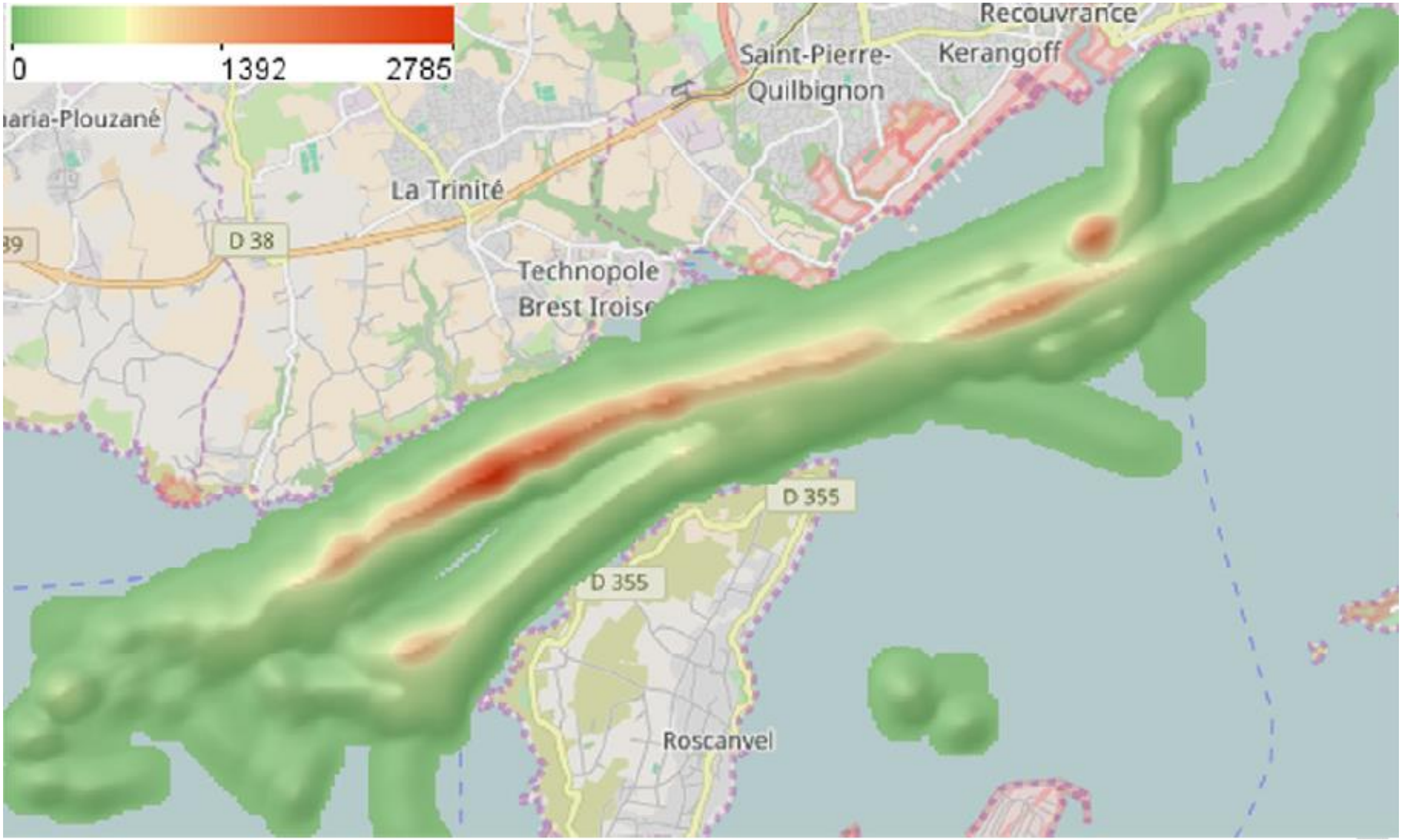
A time series display shows the counts of the vessels (upper row) and the near-location events (lower row) by 1-hour time steps. A query selects the intervals containing at least one event

Interactive detection of events (2)



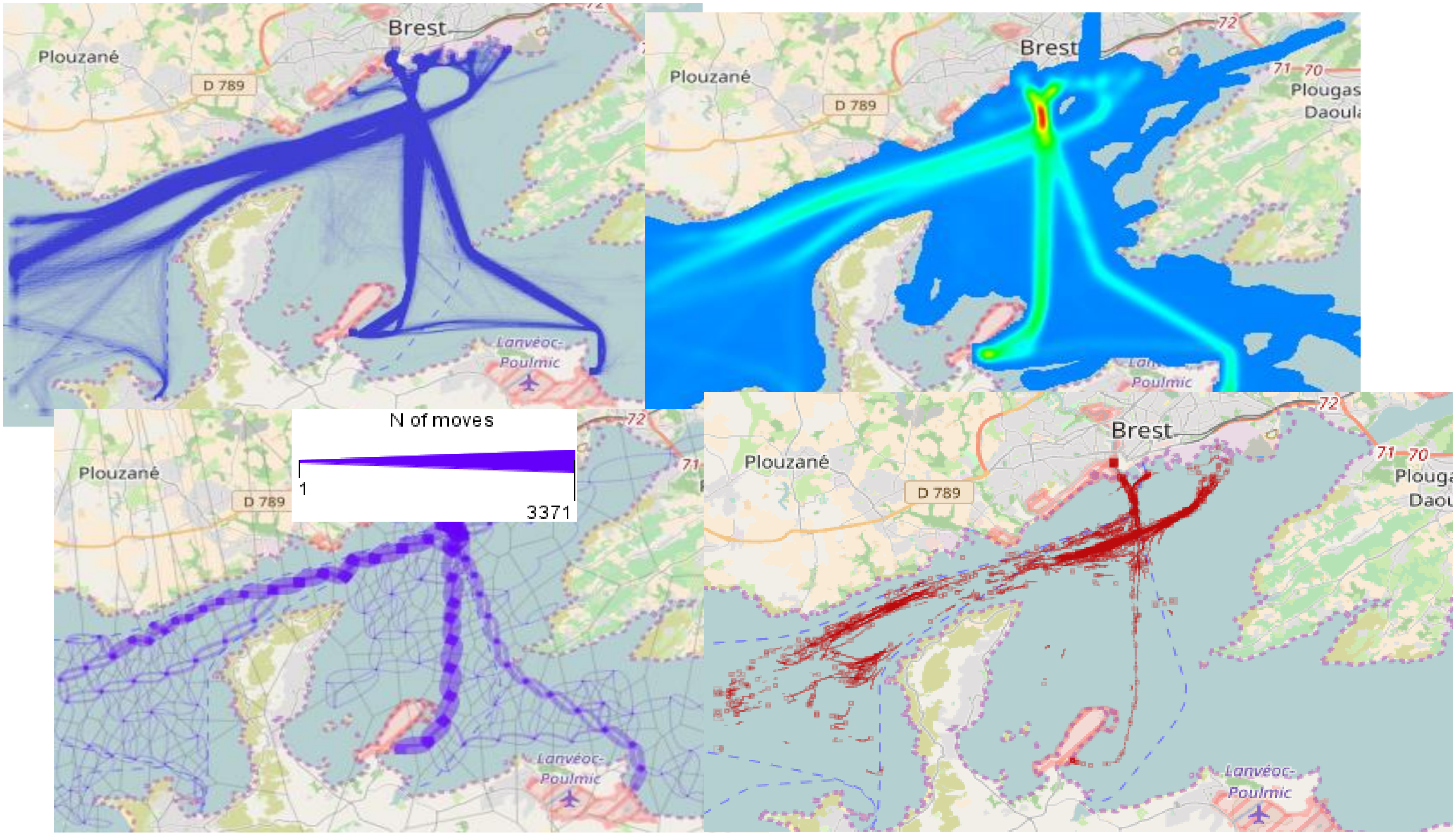
Trajectory segments corresponding to curvy movement

Interactive detection of events (3)



Trajectory segments corresponding to curvy movement

Visual analytics: Data exploration



The main visual analytics challenges are :

- **to develop interactive and scalable data exploration and**
- **patterns extraction of both archival (data-at-rest) and streaming (data-in-motion) spatio-temporal data at varying levels of resolution.**

Towards a maritime decision support and forecasting system

Raw AIS data

Contact-based clustering

Stopping/stationary areas

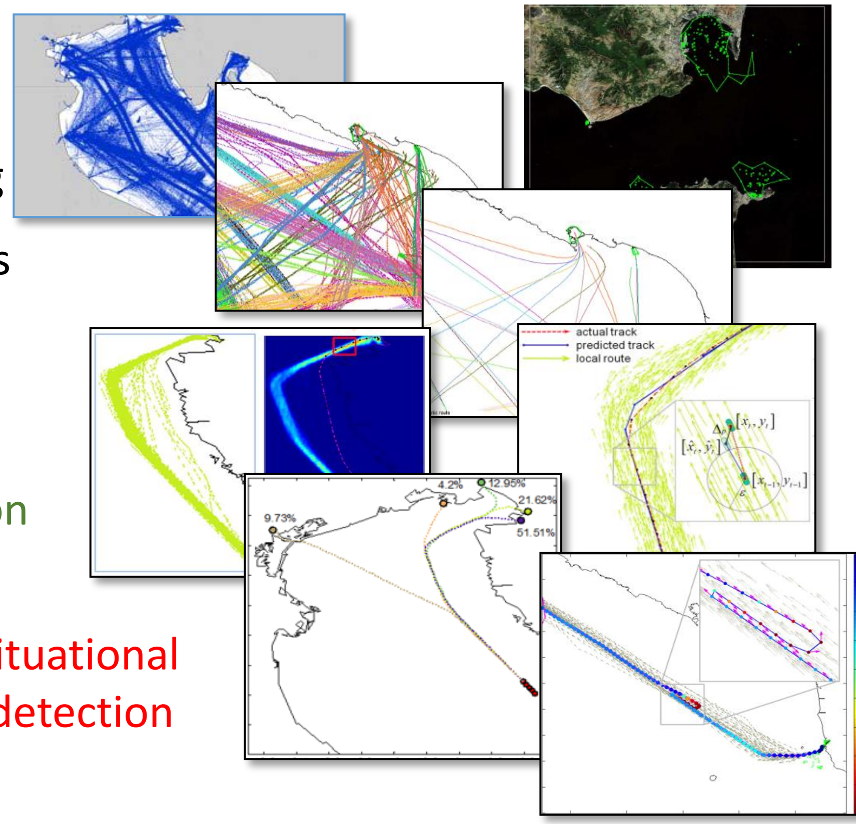
Route Extraction

Synthetic Route

Vessel-to-route association

Position, Route and Destination prediction

Maritime Situational Indicators detection

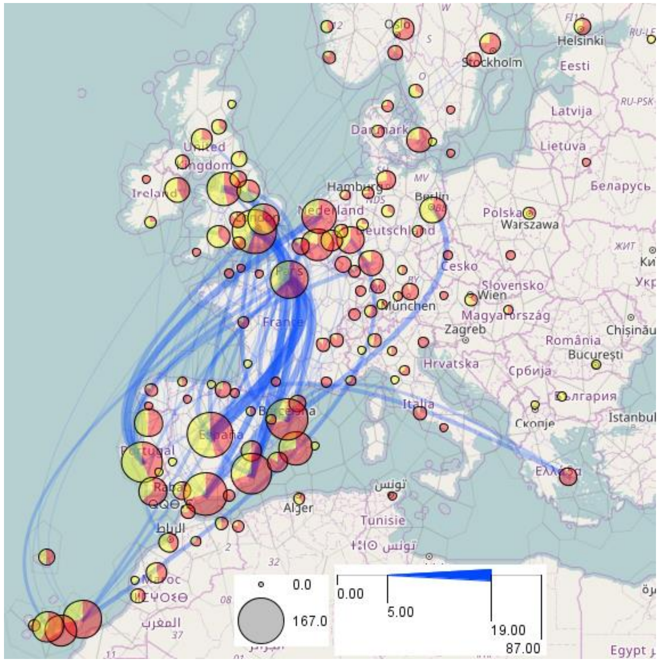


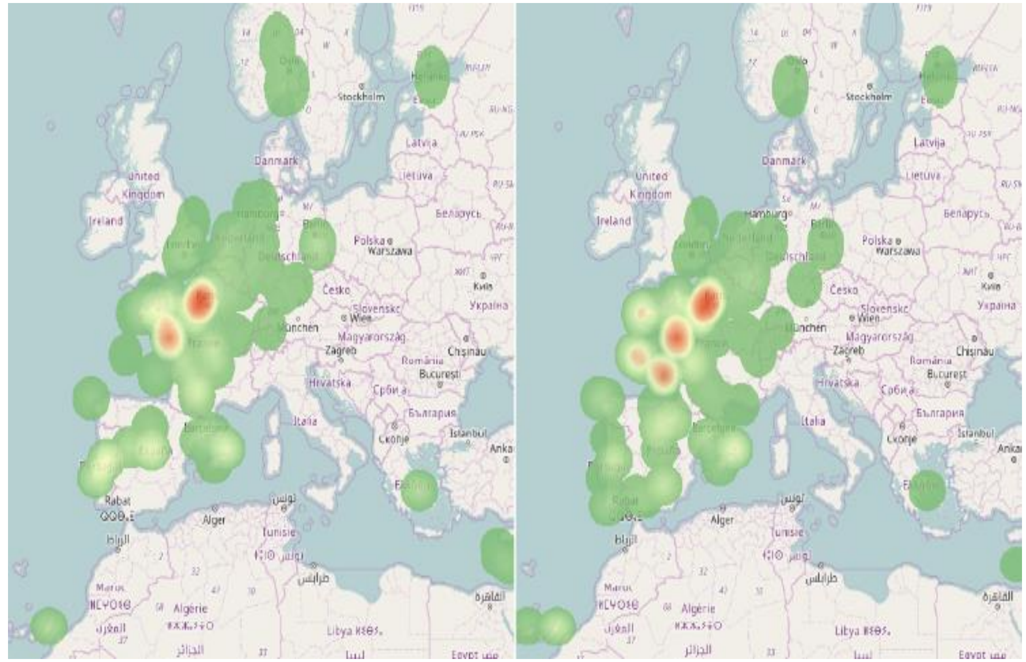
The underlying challenges are :

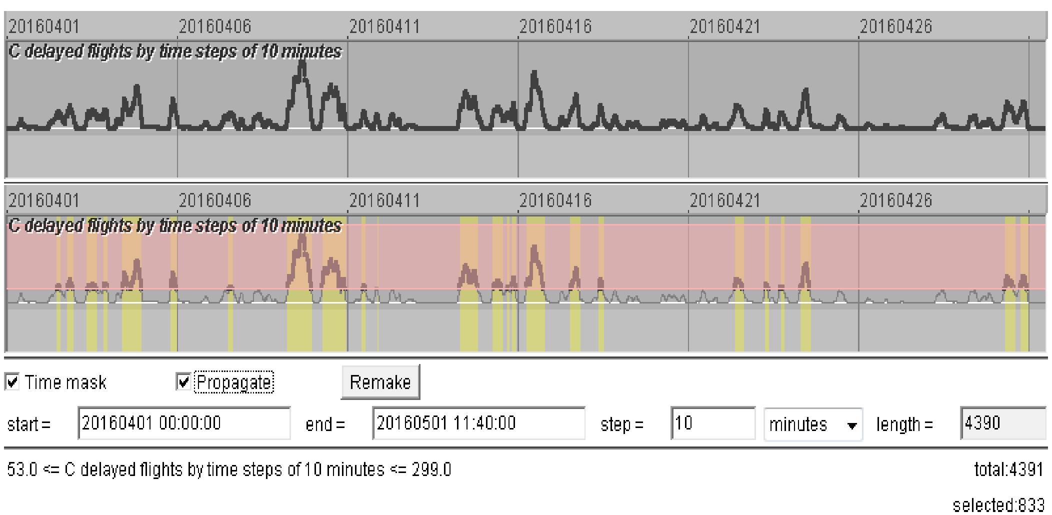
- to properly capture the human generated information including the associated uncertainty assessment so it can be meaningfully aggregated with other information from physical sensors or databases,

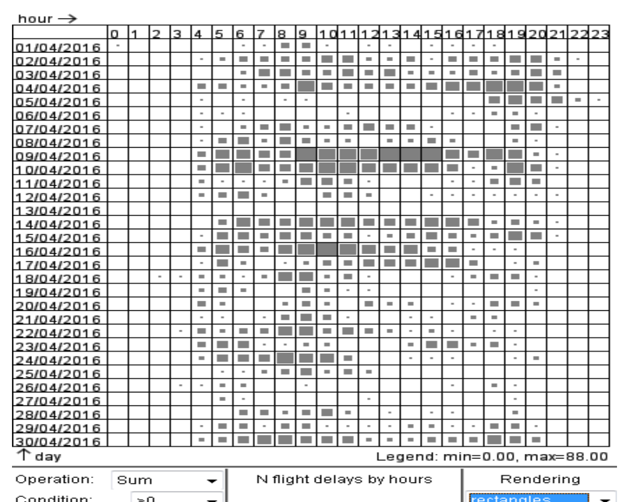
- to ensure that the system outputs meaningful, interpretable and unambiguous results on which the user can take appropriate decisions.

Back to the air flight domain…

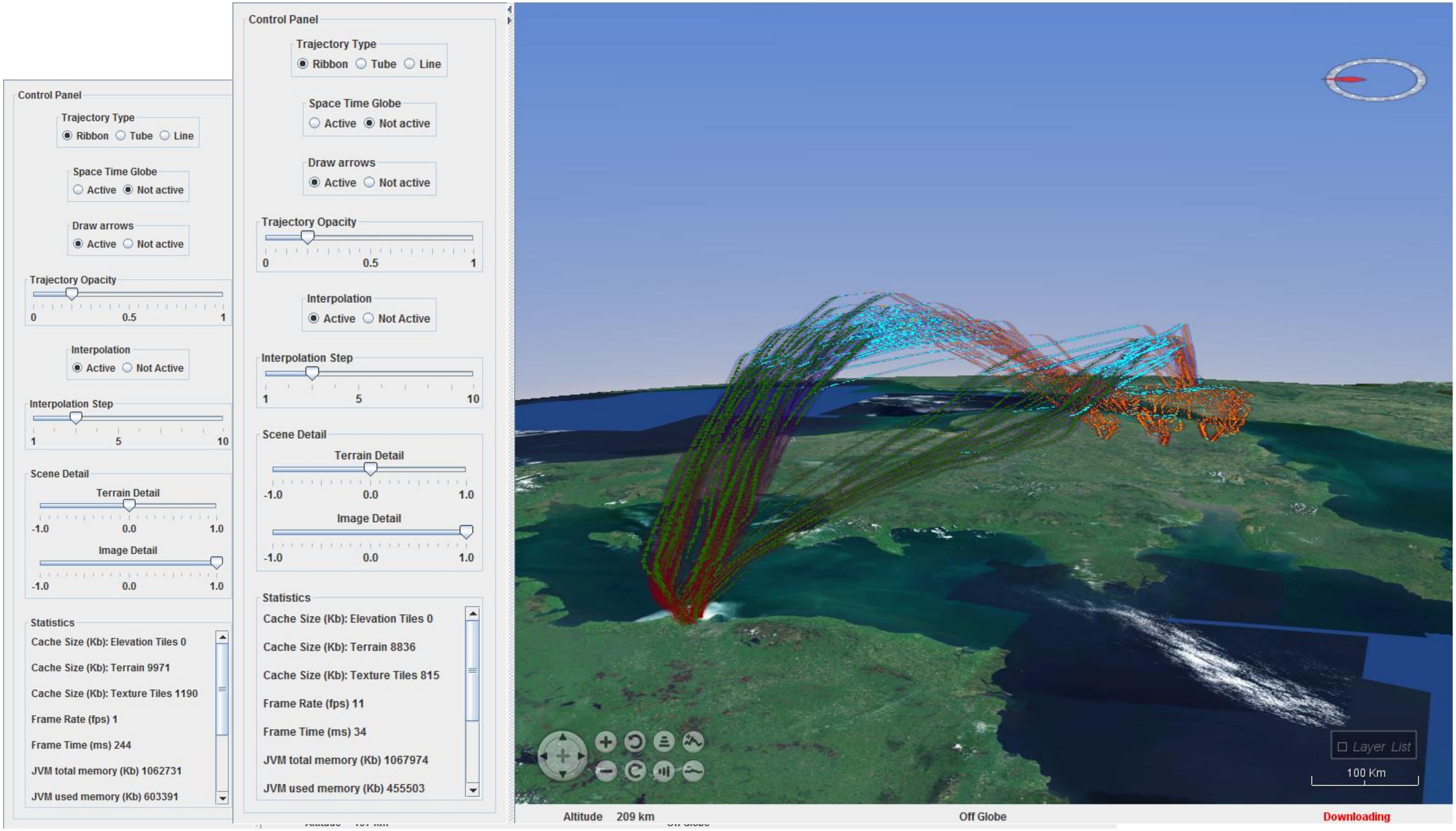








Back to the air flight domain…



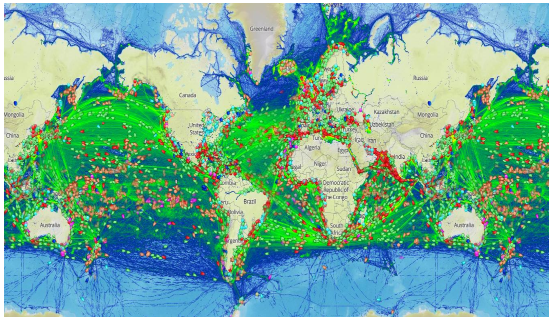
Conclusion

The maritime environment provides many research application opportunities and research challenges !

We introduced a series of current computational issues still opened and suggested several research directions for a successful integration, manipulation and analysis of maritime trajectories.

The scientific domains covered are very large thus opening several opportunities for pluri-disciplinary research, let us mention:

- **ontology and conceptual data models** at the data integration level,
- **data mining and visual analytics** for the ability to discover patterns within large volume of data,
- **machine learning** for streaming data, **information fusion** for the ability to combine information from different sources and deal with **uncertainty**,
- **human factor** and **decision-aided systems**
- **transportation** & **maritime sciences**



Big Data Management and Analytics for Mobility Forecasting in **dataCron**

- **More details at:**
- **http://www.datacron-project.eu**
- **@datacron_eu**

dataCron

Selected publications (extract)

- Andrienko G et al. (2016) **Understanding movement data quality**. Journal of Location Based Services. 10:31-46.
- Claramunt C et al. (2017) **Maritime data integration and analysis: recent progress and research challenges**. Proc. EDBT Conference.
- Nikitopoulos P et al. (2016) **BigCAB: Distributed Hot Spot Analysis over Big Spatio-temporal Data using Apache Spark (GIS Cup)**. Proc. ACM SIGSPATIAL GIS.
- Patroumpas K et al. (2017) **Online event recognition from moving vessel trajectories**. Geoinformatica, 21(2): 389-427.
- Pelekis N et al. (2017) **In-DBMS sampling-based sub-trajectory clustering**. Proc. EDBT Conference.
- Pelekis N et al. (2016) **Simulating our LifeSteps by example**. ACM Trans. Spatial Algorithms Syst.. 2:11:1–11:39.
- Salmon L et al. (2016) **Continuous detection of black holes for moving objects at sea**. Proc. IWGS Workshop.