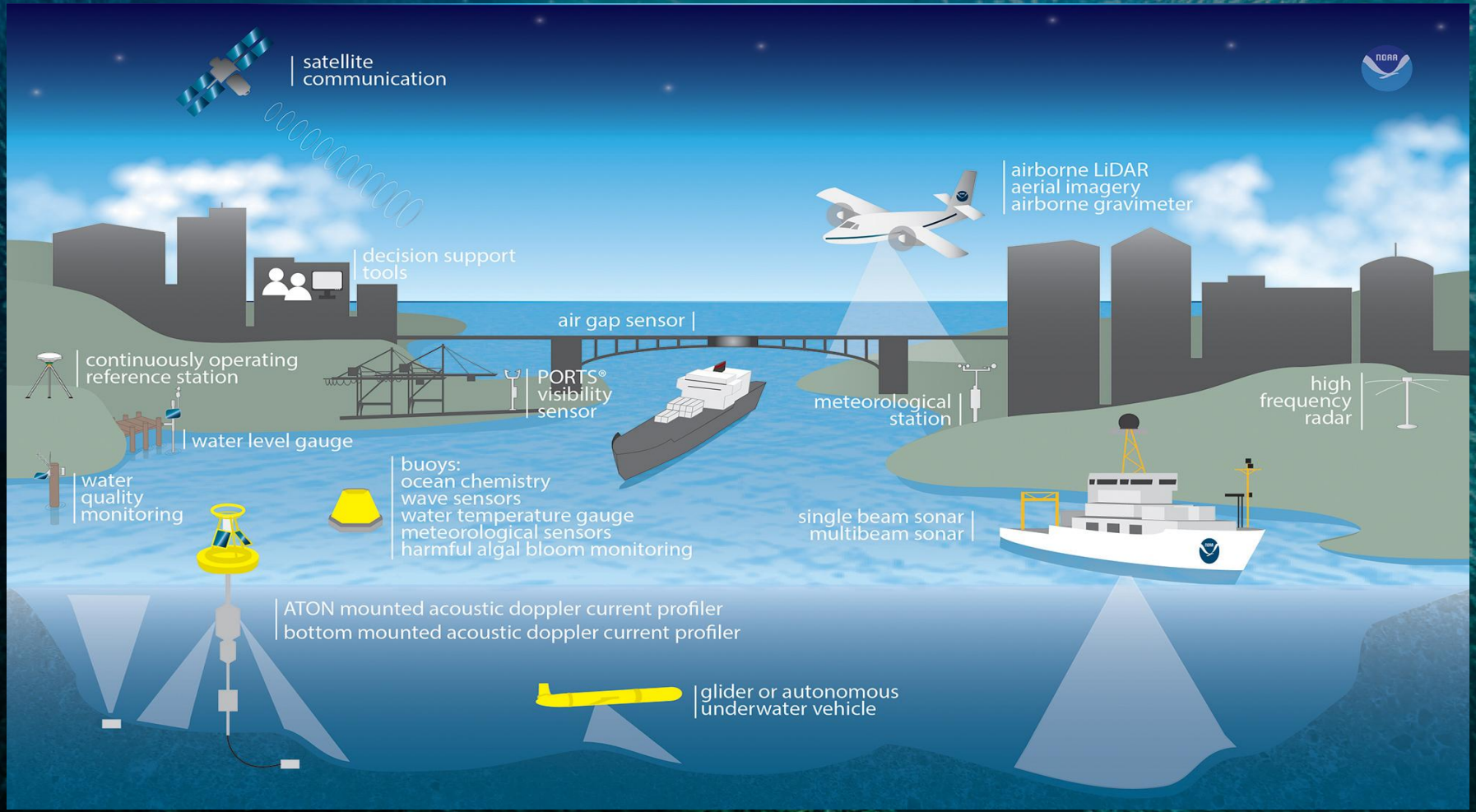
IST160 Specialists Meeting

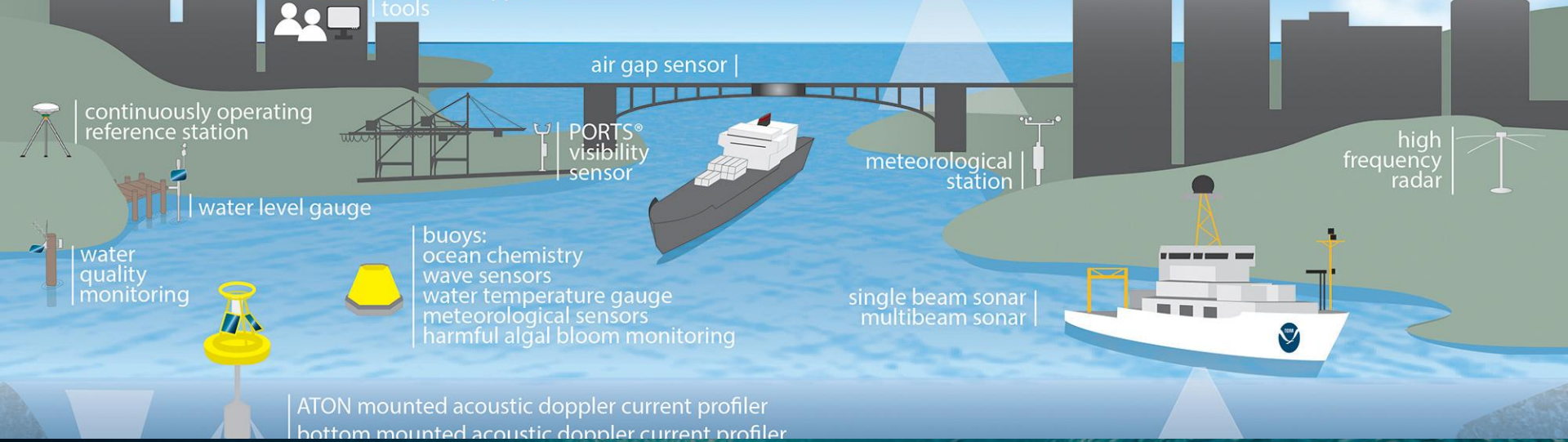
Semantic Integration of Real-Time Heterogeneous Data Streams for Ocean-related Decision Making

Renata Dividino, Amilcar Soares, Stan Matwin, Anthony W. Isenor, Sean Webb, and Matthew Brousseau

Institute for Big Data Analytics, Dalhousie University
&
Defence Research and Development Canada (DRDC)



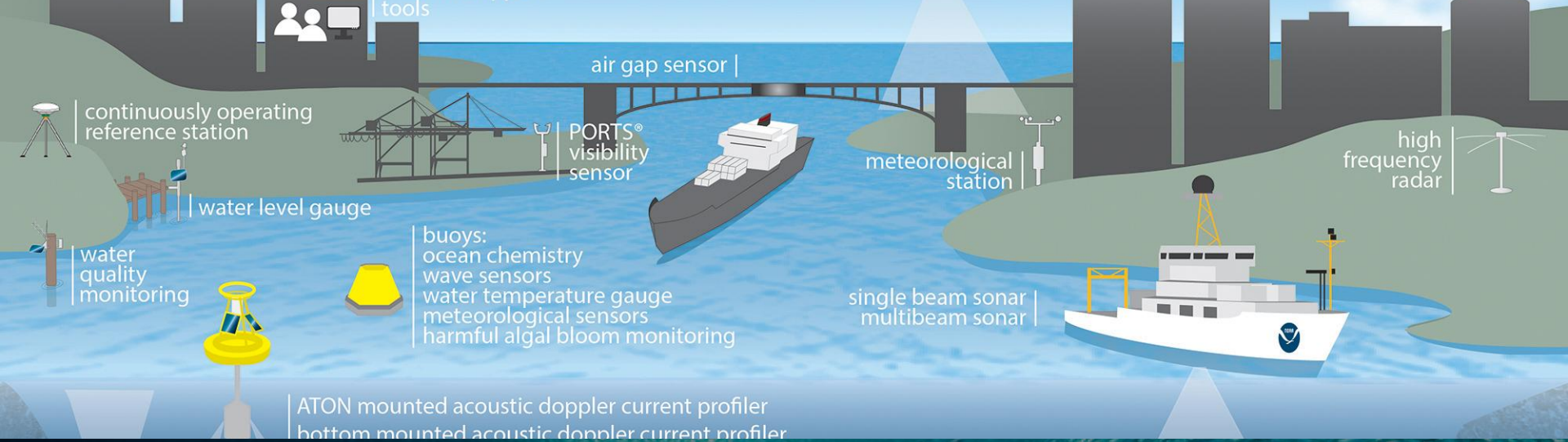
Picture taken from: http://www.noaanews.noaa.gov/stories2014/20140505_ioos_ports.html



Maritime Data

Sensor observations are used to monitor and track vessels, and to characterize the maritime environment.

Local data is used to report shipping status, transactions, routing, collisions, regulations, ...



Turning Maritime Data into Actionable Insights

Sensor observations combined with other available data can be used in applications to provide safe and efficient movement of vessel traffic, mitigating risk of pollution by preventing collisions and grounding, as well as to understand cause and effect between environment and vessels.

We need...

Real-time tracking ships and the marine environment (weather, fish migrations, etc) to improve entity screening, and to identify and predict unusual/suspicious/unexpected activities at sea.

Challenges

- **Data Deluge**
  Sensors generate thousands of data points every second!
- **Data silos**
  Data is inaccessible, unfindable, unsearchable.
- **Data is Distributed**
  Data is heterogeneous, spatially distributed, and from different IoT infrastructures
- **Need of Real-time Processing**
  Data should be analyzed and processed in real-time

Requirements: turning data into actionable insights

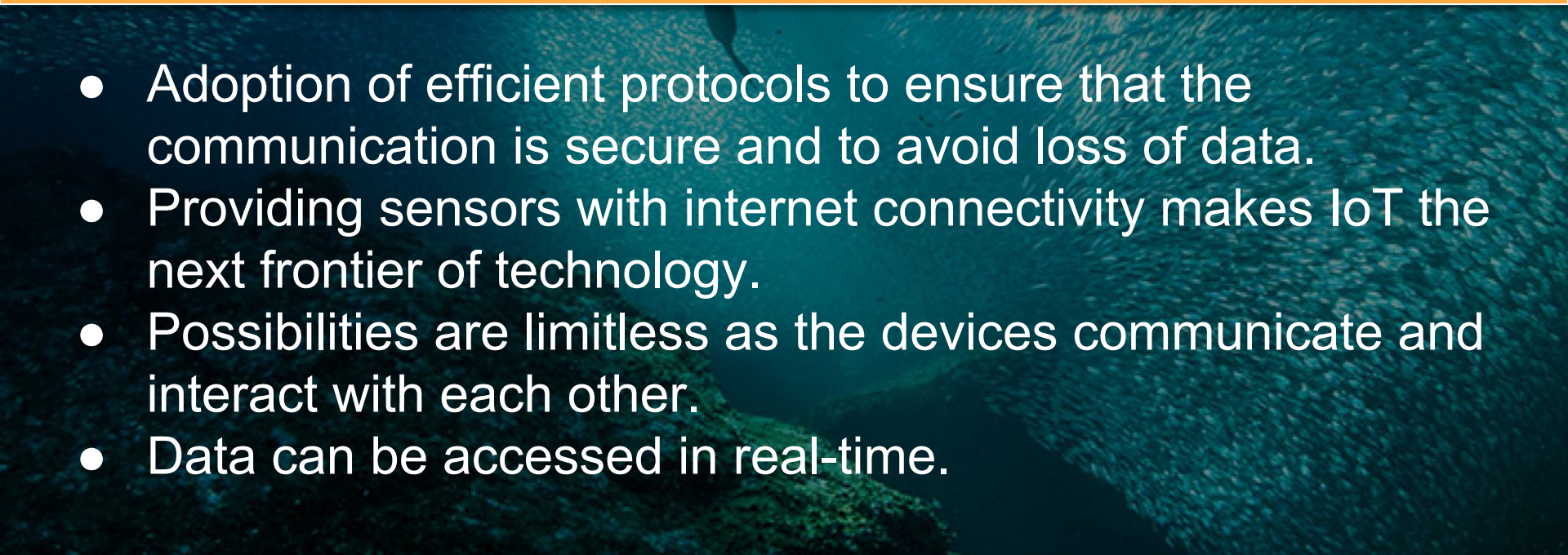
Data is accessible

Cross-dataset Queries

Data is interpretable

Queries run over data streams

Data is accessible



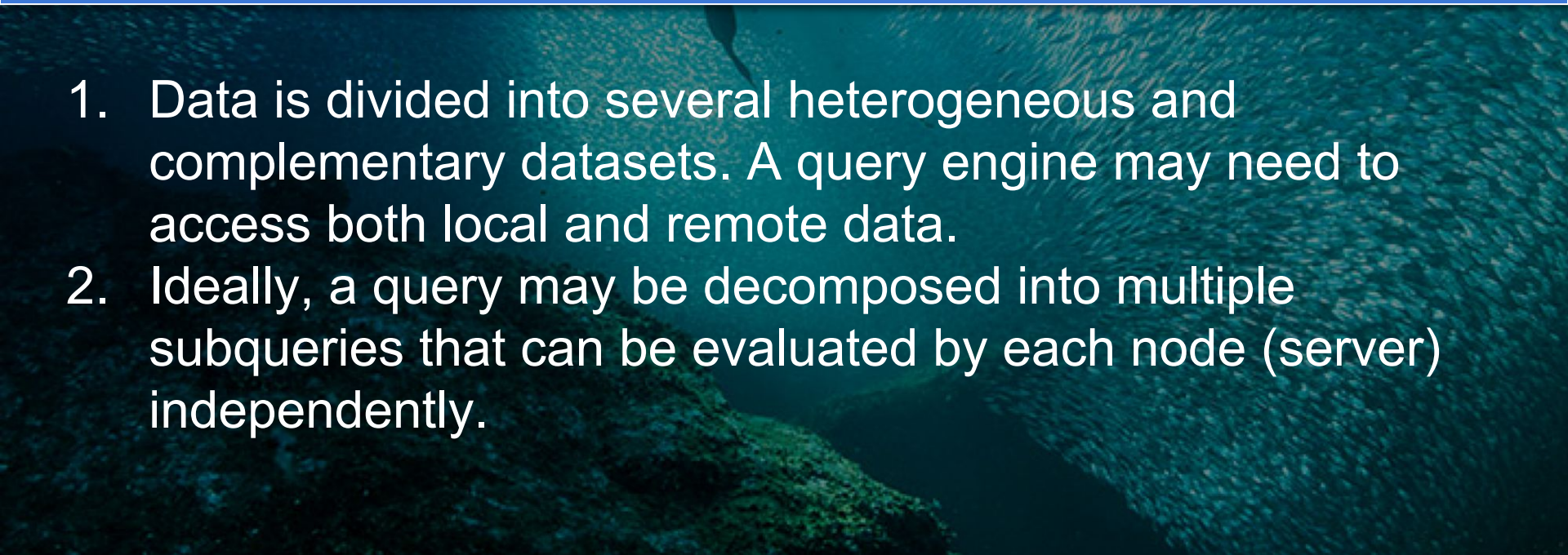
- Adoption of efficient protocols to ensure that the communication is secure and to avoid loss of data.
- Providing sensors with internet connectivity makes IoT the next frontier of technology.
- Possibilities are limitless as the devices communicate and interact with each other.
- Data can be accessed in real-time.

Data is interpretable

Use of Semantic Web languages and technologies:

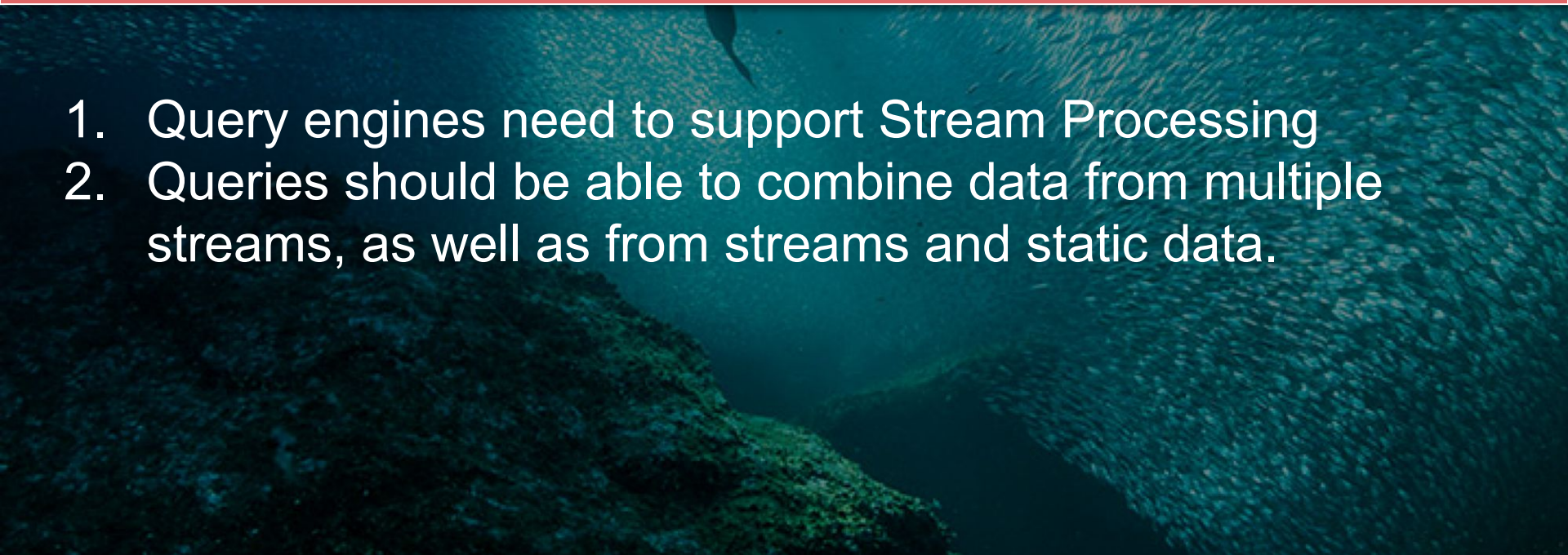
1. Using a common syntax for machine (and human) understandable statements.
2. Using established common vocabularies.
3. Using common logical language.
4. Using the language for exchanging proofs.

Cross-Dataset Queries



1. Data is divided into several heterogeneous and complementary datasets. A query engine may need to access both local and remote data.
2. Ideally, a query may be decomposed into multiple subqueries that can be evaluated by each node (server) independently.

Queries run over data streams



1. Query engines need to support Stream Processing
2. Queries should be able to combine data from multiple streams, as well as from streams and static data.

Our Play Scenario

We track vessels, weather and ice conditions in real-time.

We want to receive alerts any time the system identifies:

a. Vessels which are speeding (in a given area, e.g. port)
b. Vessels heading to regions with severe weather conditions
c. Vessels heading to regions with deteriorating ice conditions

When the user selects a specific vessel for inspection, we execute queries over remote and local database as:

a. an internal IMO database to get information about vessel owner, insurance, and specifications, etc.
b. the Web, more specific the DBpedia (open and structured Wikipedia)

Our Play Scenario

We track vessels, weather and ice conditions in real-time.

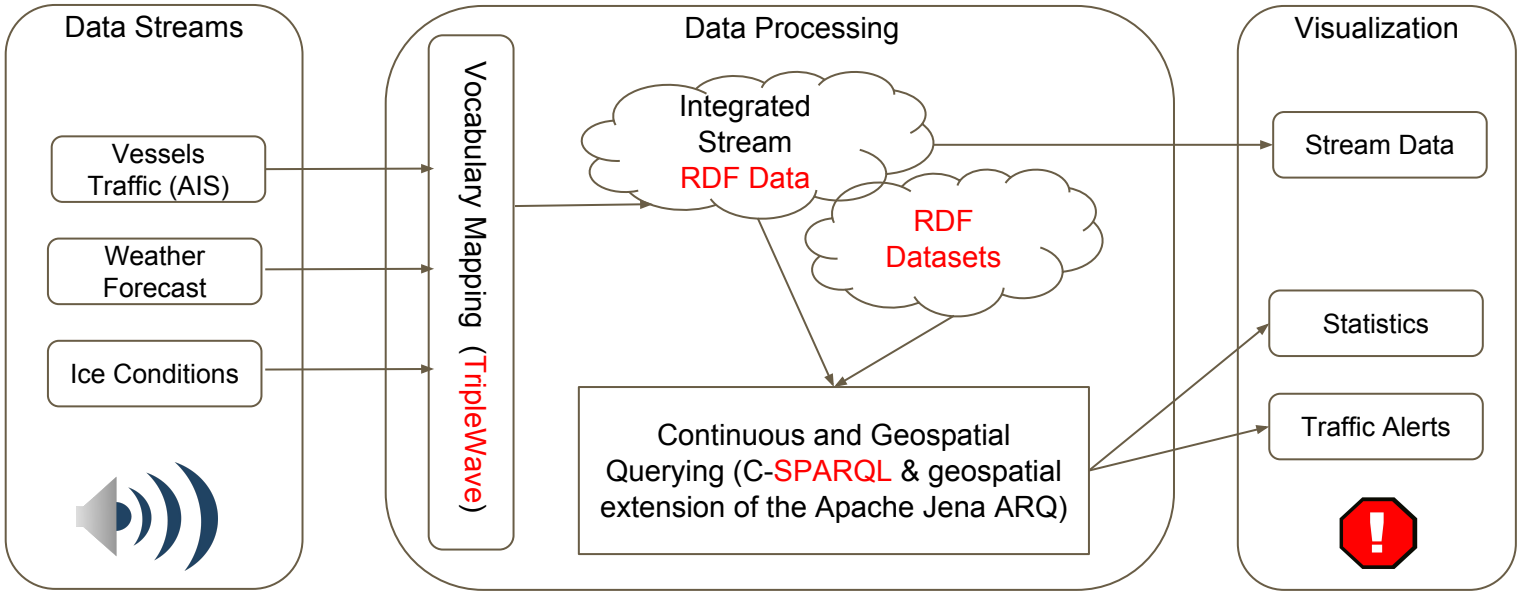
We want to receive alerts any time the system identifies:

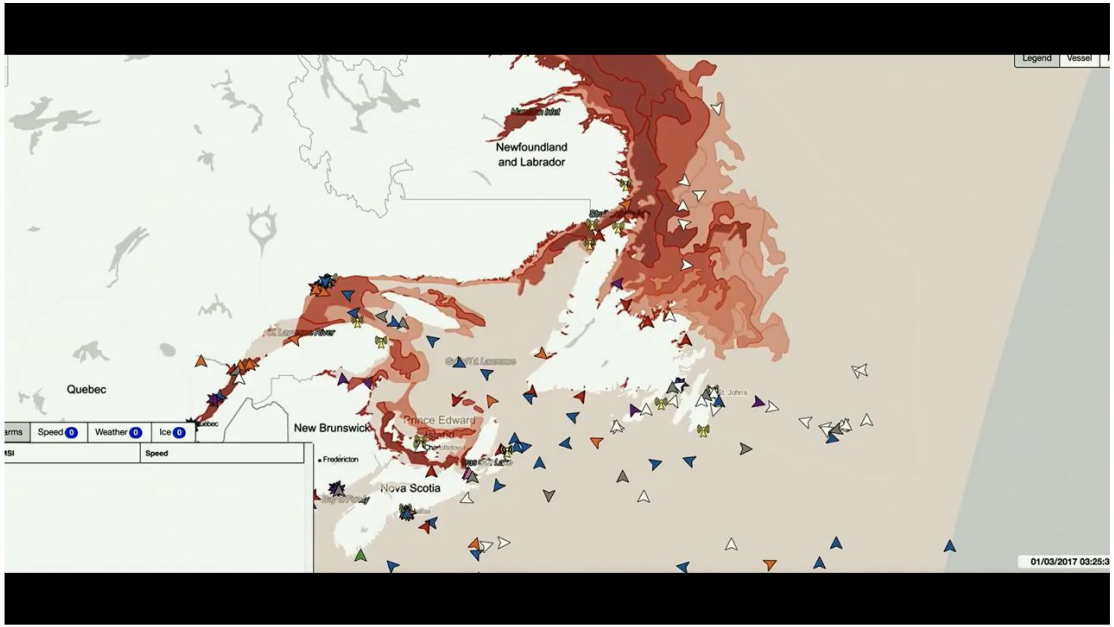
a. Vessels which are speeding (in a given area, e.g. port)
b. Vessels heading to regions with severe weather conditions
c. Vessels heading to regions with deteriorating ice conditions

When the user selects a specific vessel for inspection, we execute queries over remote and local database as:

a. an internal IMO database to get information about vessel owner, insurance, and specifications, etc.
b. the Web, more specific the DBpedia (open and structured Wikipedia)

Framework Architecture





https://www.youtube.com/watch?v=RuvKvp1UwVg

CRISIS - Cross Queries Over Streams

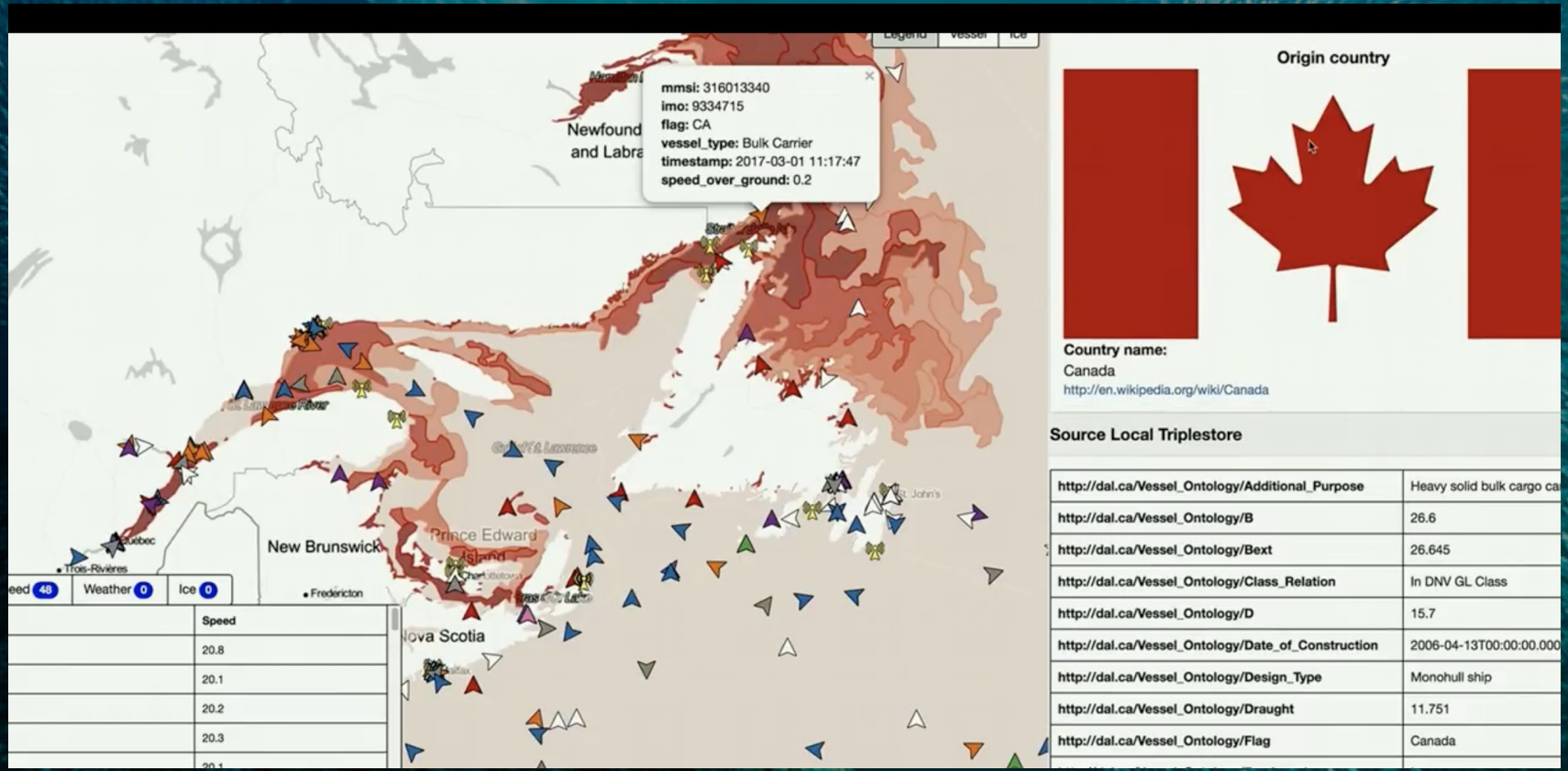
30 Aufrufe

0 0 TEILEN

Institute for Big Data Analytics
Am 30.04.2018 veröffentlicht

ABONNIEREN 1

The goal of CRISIS is to support maritime related decision making, which is accomplished by processing real-time cross-dataset queries over various streaming, heterogeneous and spatially distributed maritime sensors from multiple IoT infrastructures.


mmsi: 316013340
imo: 9334715
flag: CA
vessel_type: Bulk Carrier
timestamp: 2017-03-01 11:17:47
speed_over_ground: 0.2
Legend
Vessel
Ice
Origin country
Country name:
Canada
http://en.wikipedia.org/wiki/Canada
Source Local Triplestore
http://dal.ca/Vessel_Ontology/Additional_Purpose
Heavy solid bulk cargo ca
http://dal.ca/Vessel_Ontology/B
26.6
http://dal.ca/Vessel_Ontology/Bext
26.645
http://dal.ca/Vessel_Ontology/Class_Relation
In DNV GL Class
http://dal.ca/Vessel_Ontology/D
15.7
http://dal.ca/Vessel_Ontology/Date_of_Construction
2006-04-13T00:00:00.000
http://dal.ca/Vessel_Ontology/Design_Type
Monohull ship
http://dal.ca/Vessel_Ontology/Draught
11.751
http://dal.ca/Vessel_Ontology/Flag
Canada
Newfound
and Labra
New Brunswick
Prince Edward
Island
Nova Scotia
Fredericton
Trois-Rivières
St. John's
Weather 0
Ice 0
Speed
20.8
20.1
20.2
20.3

Our Goal

To build a platform that must identify, integrate, and interpret heterogeneous and distributed IoT data streams and open data, with information flowing from these data sources automatically expressed on the basis of rich background knowledge and analyzed in real-time.

Conclusion

Our framework seeks the seamless integration of data streams of the many IoT sensors infrastructure with other information providers and for practical applications for Situation Awareness in Maritime Domain.

Future Work - Use of Machine learning for:

1. Automatically discover unusual behaviours
2. Learn to rank risks


Thanks for your attention
Questions:
dividino@dal.ca