European Commission

Big Data Integration and Management for the ATM Domain: The datAcron approach

G. Santipantakis, C.Doulkeridis,
G. A. Vouros
University of Piraeus (UPRC)

Data Enhanced TBO Workshop @ ICRAT 2018
www.datacron-project.eu

datAcron

HORIZON 2020

datAcron vision

… is to 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 significantly**

advance the capacities of systems to promote safety and effectiveness of critical operations for large numbers of moving entities in large geographical areas

datAcron

addresses core challenges of the European Big Data Vision

Data Management:
Data transformations, semantic integration, spatio-temporal query answering

Visual Analytics:
Multi-scale visualizations (time and space), visual data exploration, big data analytics

Predictive Analytics:
Forecasting trajectories and events

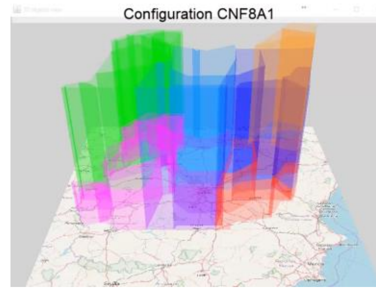
Data Processing:
In-situ data processing, synopses generation, integrated processing data-in-motion & data-at-rest

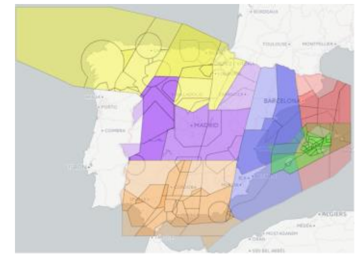
Data Management: big data challenges

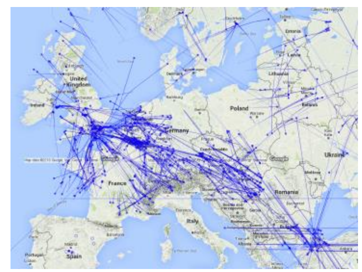
Variety

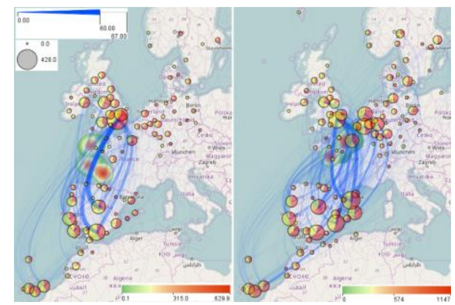
Volume and Velocity

180.000 distinct flights/day (OpenSky Network)

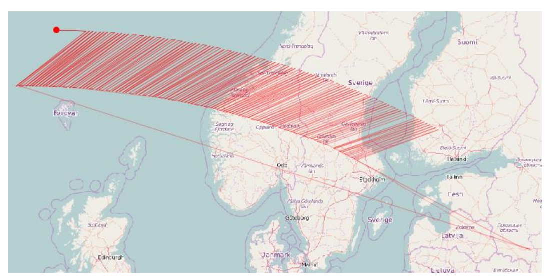








Historical & aggregated data, geographical & environmental data, contextual data, meta information



Noisy and error-prone data due to gaps in coverage, position errors, spatial distribution, repeated IDs

Veracity Issues

User-defined Challenges

- Aviation domain
  - Reduce costs by increasing the predictability of the overall system
  - Build accurate prediction models for aircraft trajectories
  - Discover patterns of predicted trajectories and events
  - Assess adherence to flight plan
  - Forecast demand-capacity imbalances and regulations.

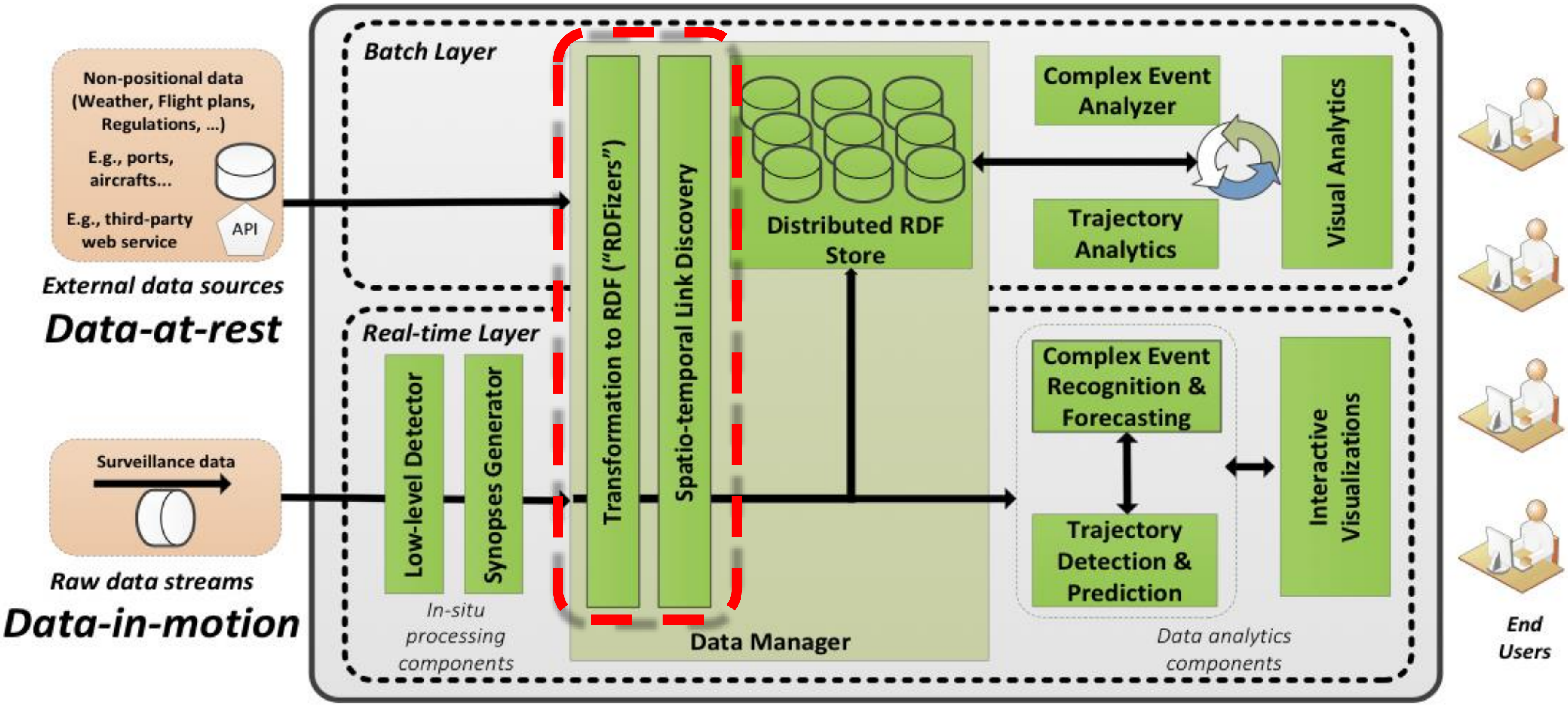
Research Challenges

- Scalable, automatic, real-time processing, semantic annotation and linking of data towards coherent views on integrated cross-streaming (data-in-motion) and archival (data-at-rest) data

  Focus of this talk

- Efficient distributed management and querying of integrated spatio-temporal data

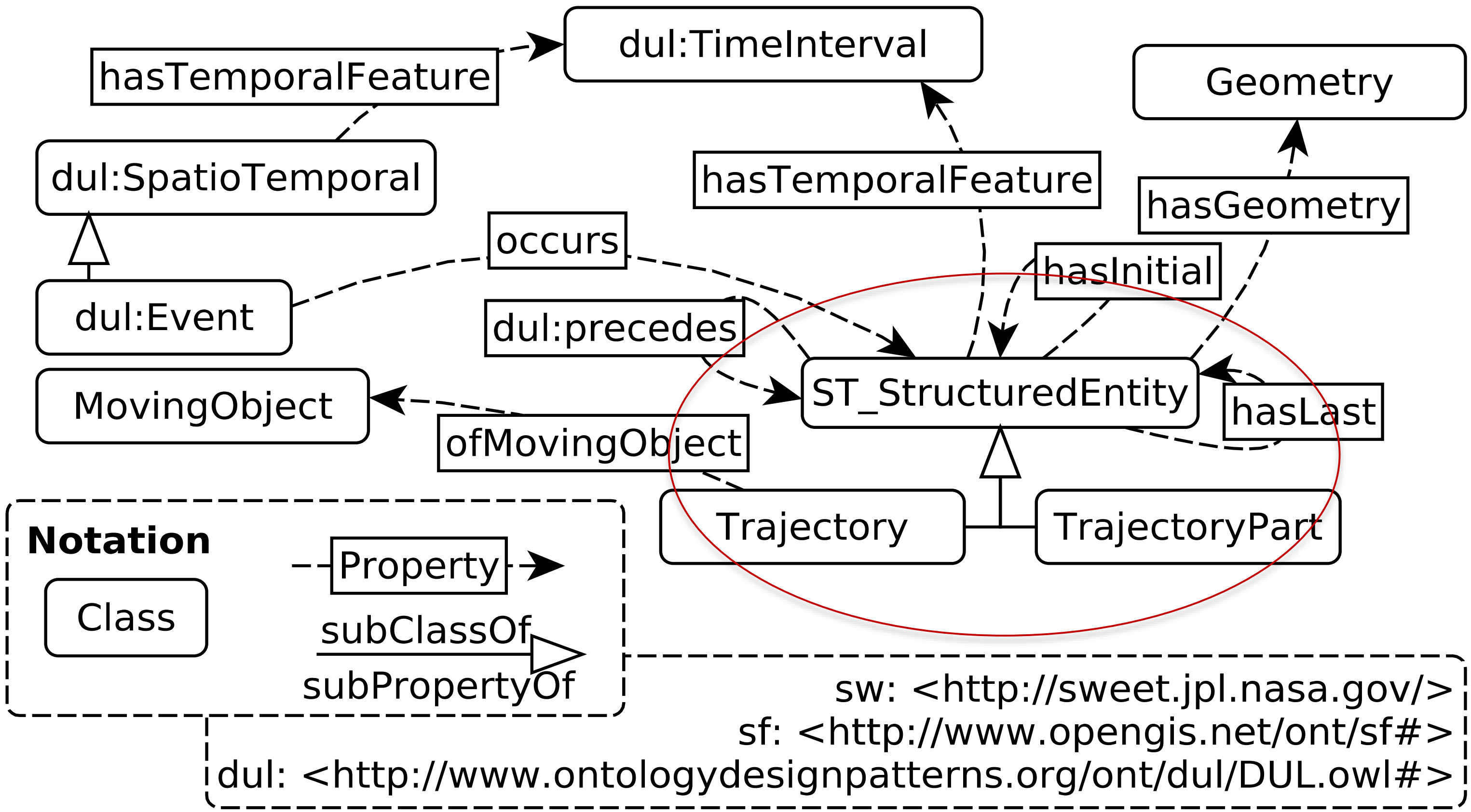
The datAcron System Architecture




2016-04-13 14:43:30
datAcron

The datAcron Ontology

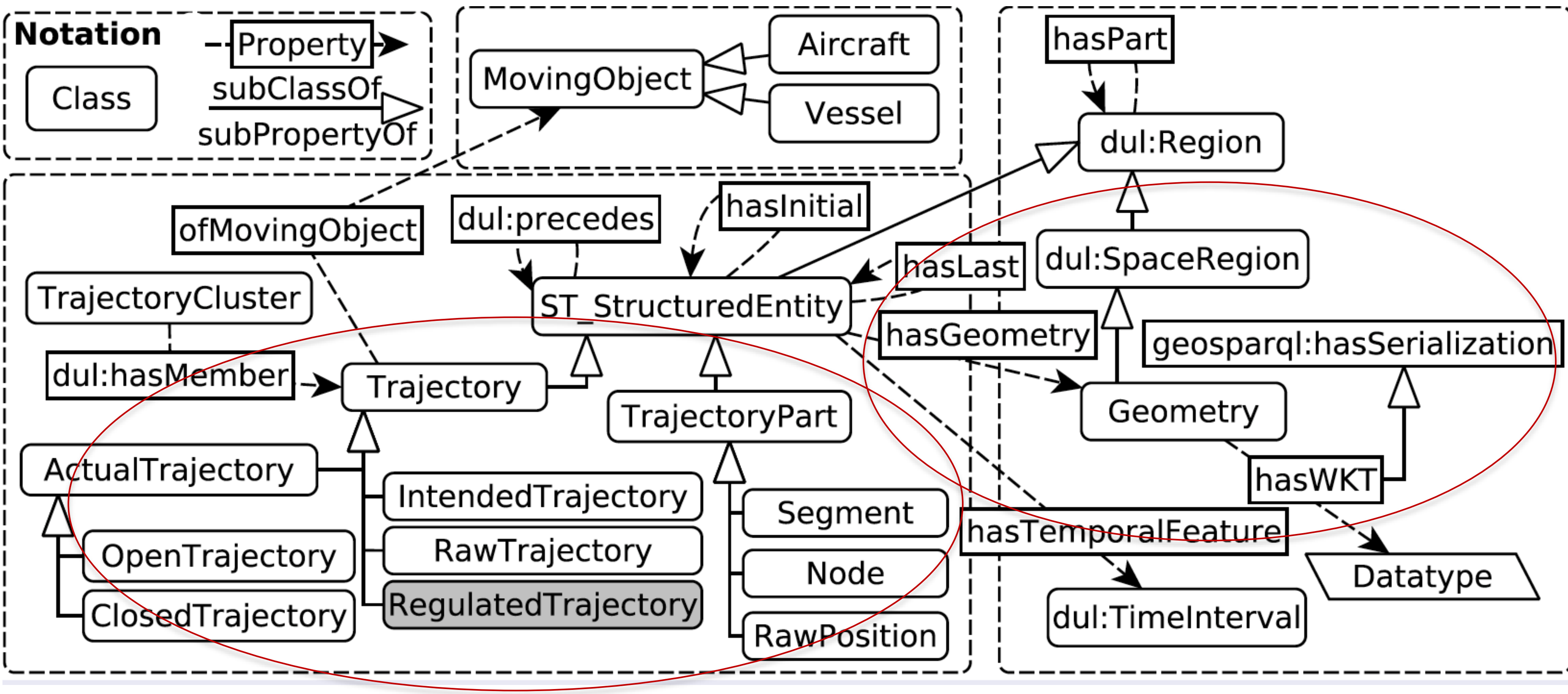
Main concepts and relations

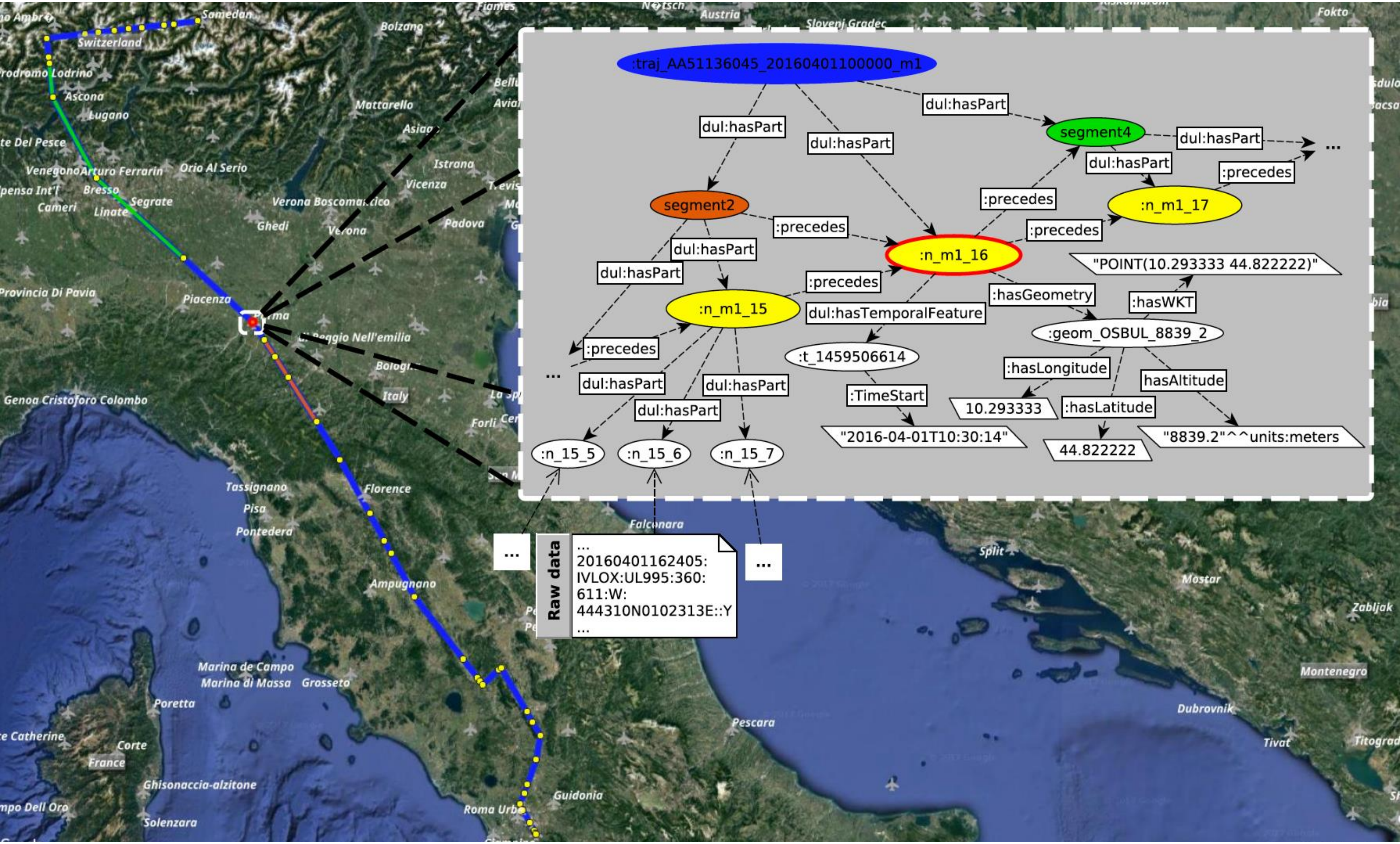


G.M. Santipantakis, G.A. Vouros, C.Doulkeridis, A.Vlachou, G.L.Andrienko, N.V. Andrienko, G.Fuchs, J.M.Cordero Garcia, M.G.Martinez: *Specification of Semantic Trajectories Supporting Data Transformations for Analytics: The datAcron Ontology*. **SEMANTICS 2017**: 17-24

The datAcron Ontology

Main concepts and relations

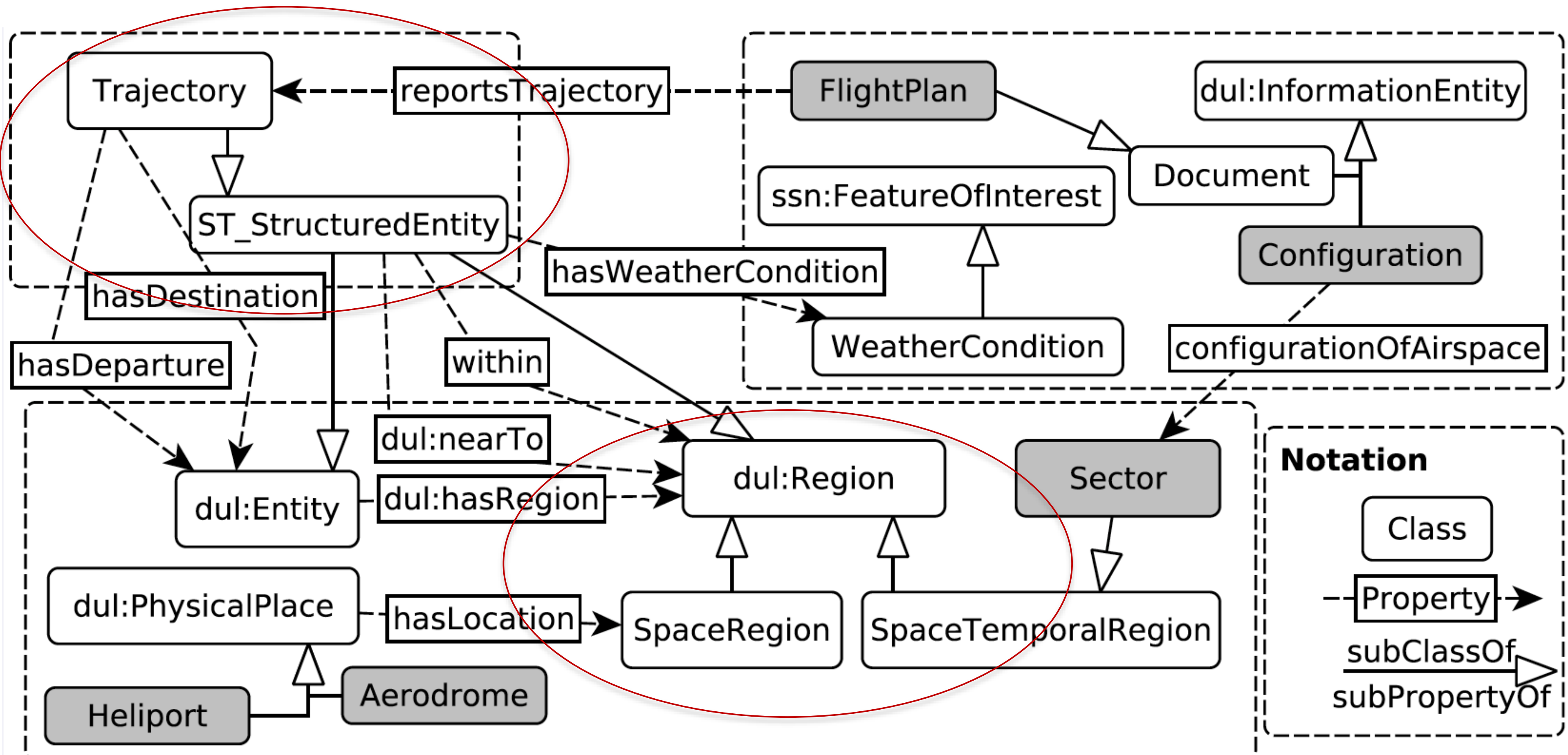




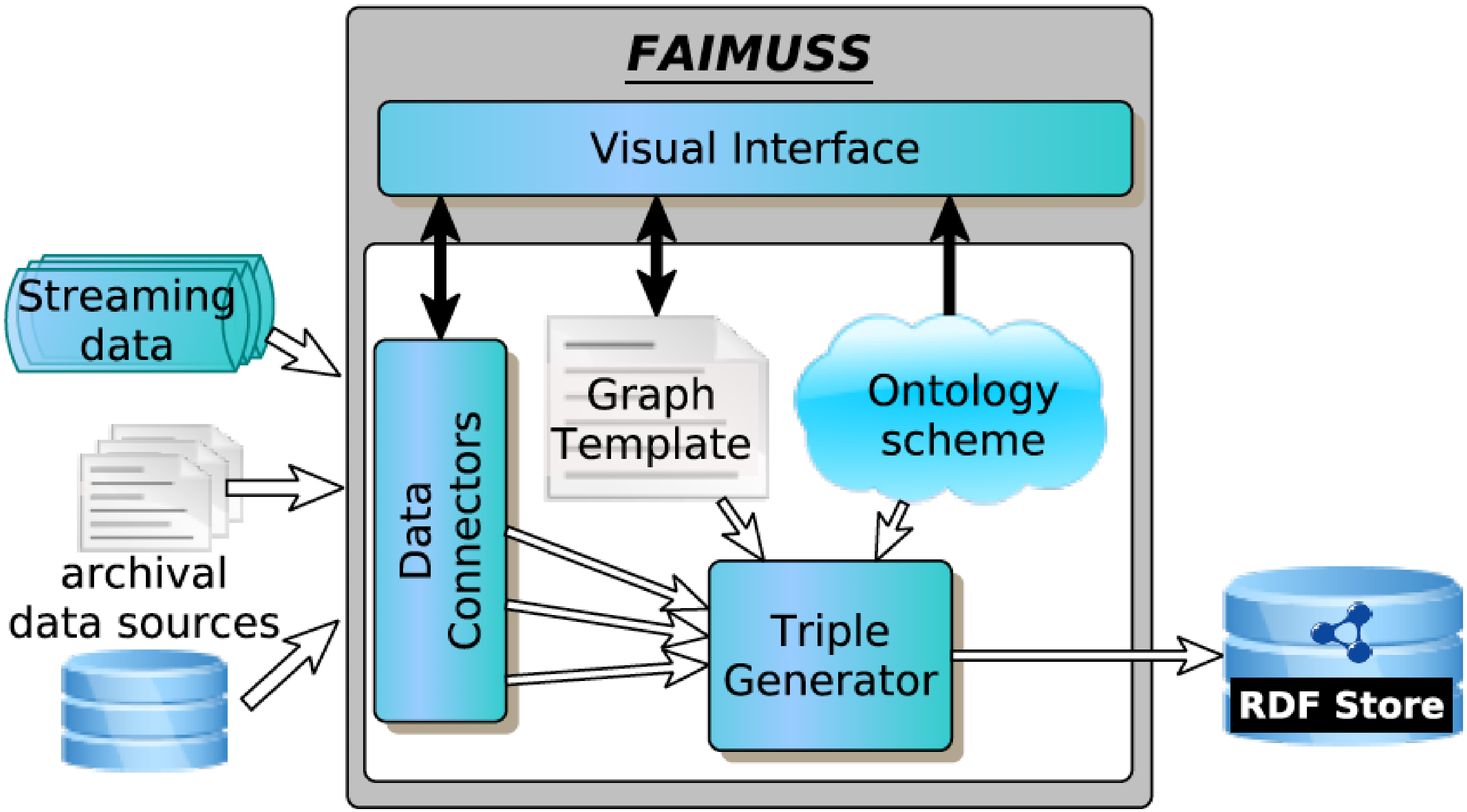
:traj_AA51136045_20160401100000_m1
dul:hasPart
dul:hasPart
dul:hasPart
segment4
dul:hasPart
...
dul:hasPart
:precedes
segment2
:precedes
:n_m1_17
:precedes
:precedes
dul:hasPart
:n_m1_16
dul:hasPart
"POINT(10.293333 44.822222)"
:precedes
:hasGeometry
:hasWKT
:n_m1_15
dul:hasTemporalFeature
:geom_OSBUL_8839_2
:precedes
:t_1459506614
:hasLongitude
...
dul:hasPart
dul:hasPart
hasAltitude
:TimeStart
10.293333
:hasLatitude
dul:hasPart
"2016-04-01T10:30:14"
"8839.2"^^units:meters
44.822222
:n_15_5
:n_15_6
:n_15_7
...
Raw data
...
20160401162405:
IVLOX:UL995:360:
611:W:
444310N0102313E::Y
...
...

The datAcron Ontology

Main concepts and relations

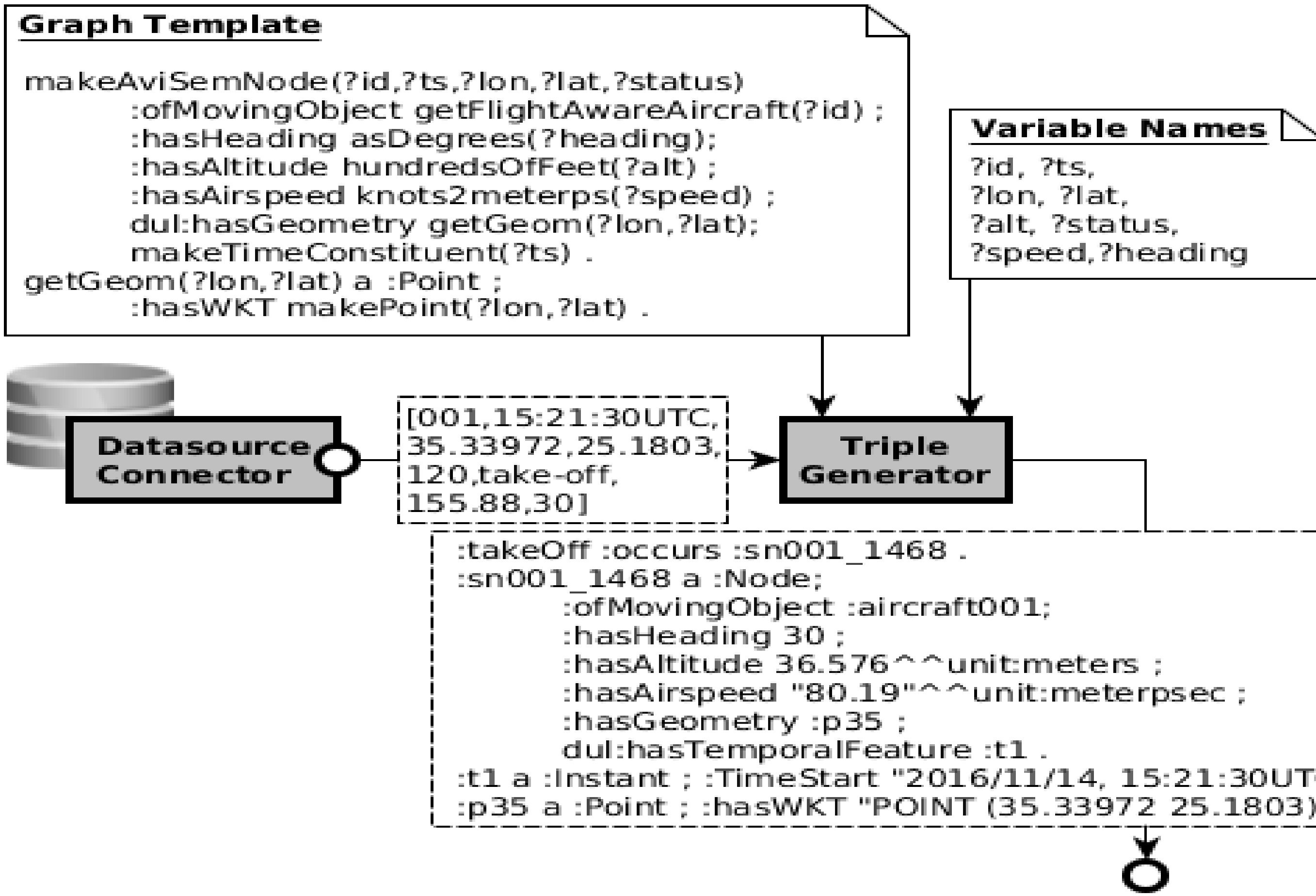


Data Transformation to RDF: RDFGen



G.M. Santipantakis, A.Glenis, N.Kalaitzian, A.Vlachou, C.Doulkeridis, G.A.Vouros: *FAIMUSS: Flexible Data Transformation to RDF from Multiple Streaming Sources*. **EDBT 2018 (demo)**

RDFGen



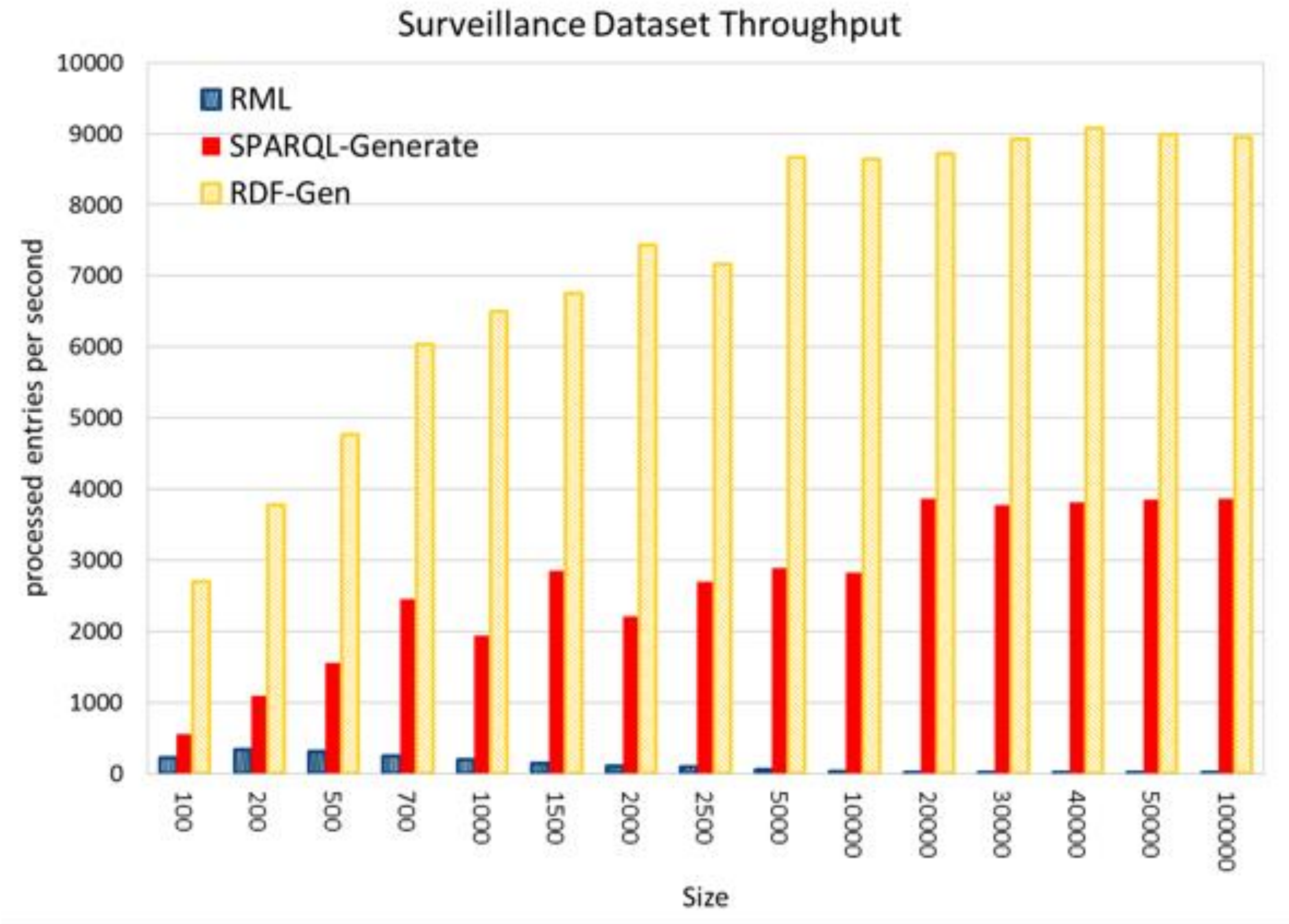
G.M. Santipantakis, A.Glenis, N.Kalaitzian, A.Vlachou, C.Doulkeridis, G.A.Vouros: *FAIMUSS: Flexible Data Transformation to RDF from Multiple Streaming Sources*. **EDBT 2018 (demo)**

RDFGen

	O1	O2	O3	O4	O5	O6	O7
RML [1]		✓	✓		✓	✓	✓
SPARQL-Generate [2]						✓	✓
KR2RML [3]	✓	✓		✓	✓	✓	
RMLProcessor [4]		✓				✓	✓
DataLift [5]						✓	✓
RDF-Gen	✓	✓	✓	✓	✓	✓	✓

O1 Inherently supports the RDF generation of both streaming and archival datasets.
O2 Provides facilities for close-to-source data processing tasks, e.g. for data cleansing, data manipulation and conversion, and generation of URIs.
O3 Supports close-to-source link discovery functionality.
O4 Demonstrates computational efficiency in terms of high throughput and low data-generation latency.
O5 Demonstrates the scalability which is necessary for the transformation of big data.
O6 Demonstrates extensibility, in the sense that (i) it can integrate custom data processing and manipulation functions, and (ii) it can be instantiated to new data formats.
O7 Supports reusability of solutions across data sources of the same domain.

RDFGen: Comparative Evaluation



G.M. Santipantakis, K.I.Kotis, G.A. Vouros, C.Doulkeridis: *RDF-Gen: Generating RDF from Streaming and Archival Data*. **WIMS 2018**

Link Discovery

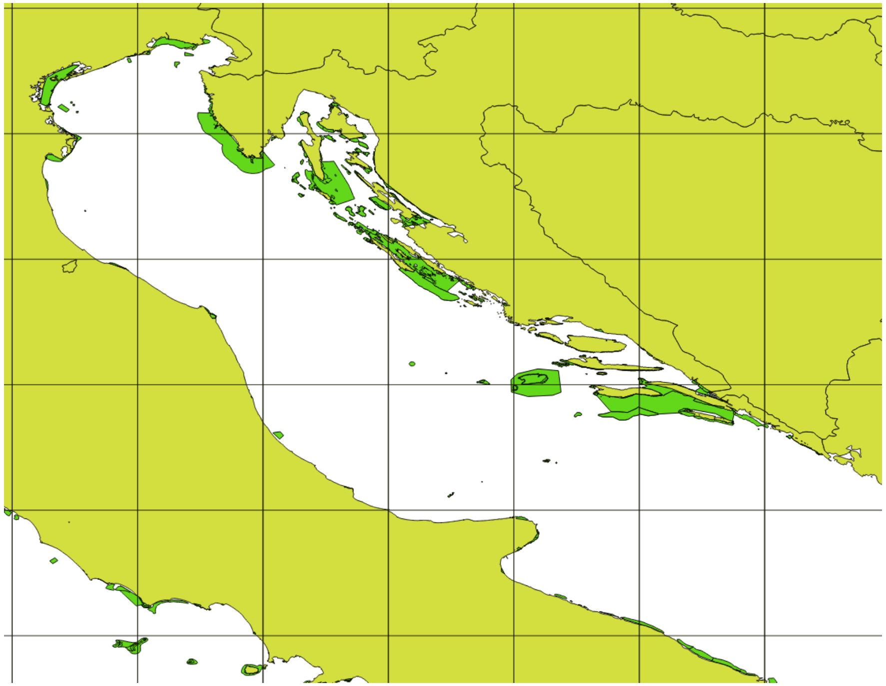
- The Link Discovery problem:

Given two data sets, namely a **target** T and **source** S data set, and a set of **relations** R, we want to detect all the associations <t r s>

- such that $t \in T$, $s \in S$ and $r \in R$,
- and (t, s) satisfy the relation r

State-of-the-art Approach

- The brute force approach is costly: O(|T| |S|)
- The state-of-the-art is to employ the blocking technique (filter-refine approach)
- Blocking technique (grid) to organize the target data set T
- For each record in source data set S
  - Find enclosing grid cell (Filter)
  - Check against Target records in the grid cell (Refine)

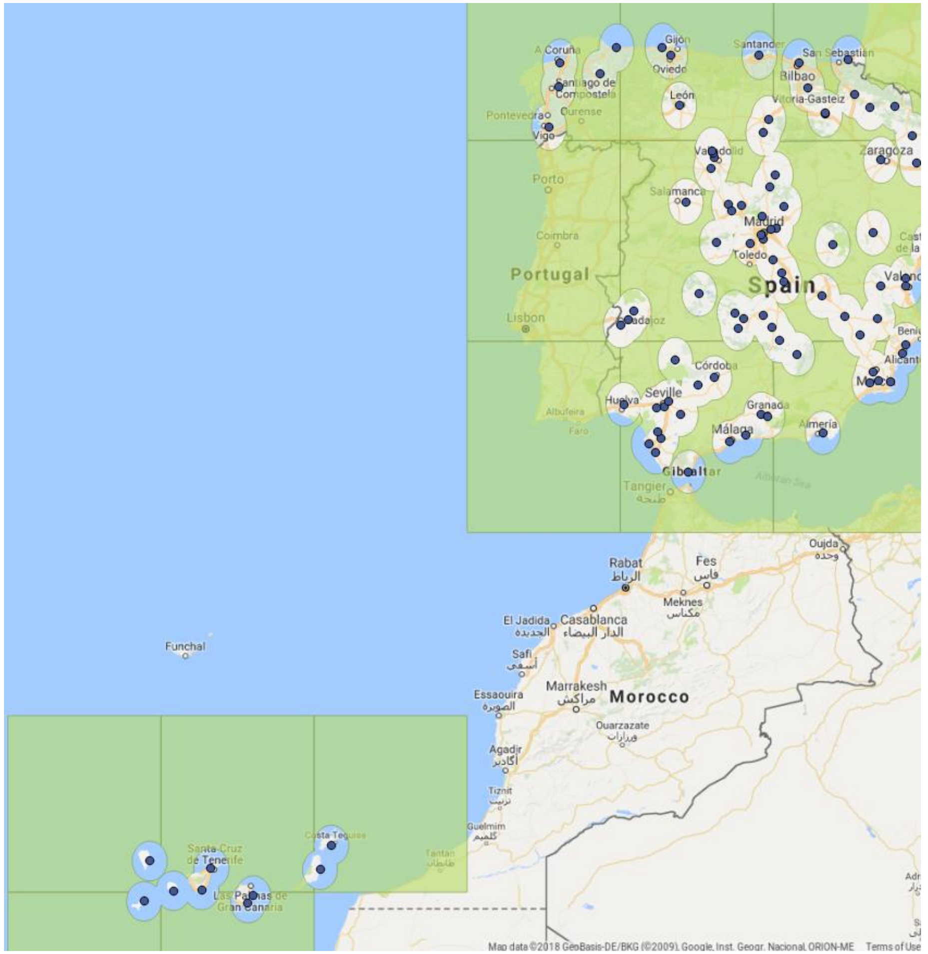


Improving the Blocking Method

- Although the blocking method can reduce the number of comparisons (compared to brute force), it still entails many unnecessary comparisons
  - Those that will not yield to a (t, s) pair satisfying a given relation
- Based on this observation, we introduced the MaskLink technique which filters candidates of T within a cell

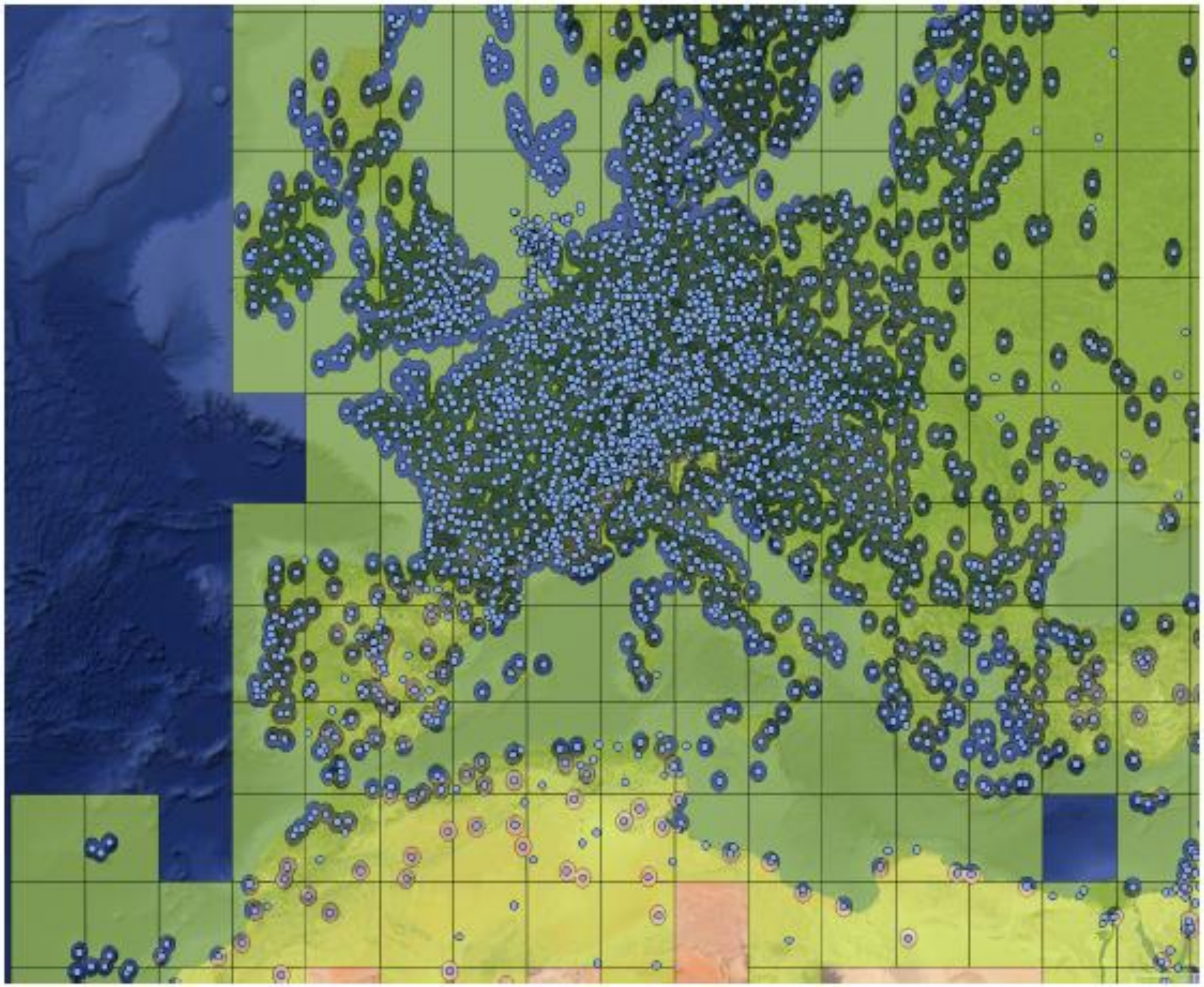
The MaskLink Technique

- The MaskLink technique:
  - We compute the empty region of each cell
  - If an element s is within the empty region, it does not need to be compared to any candidate element of T

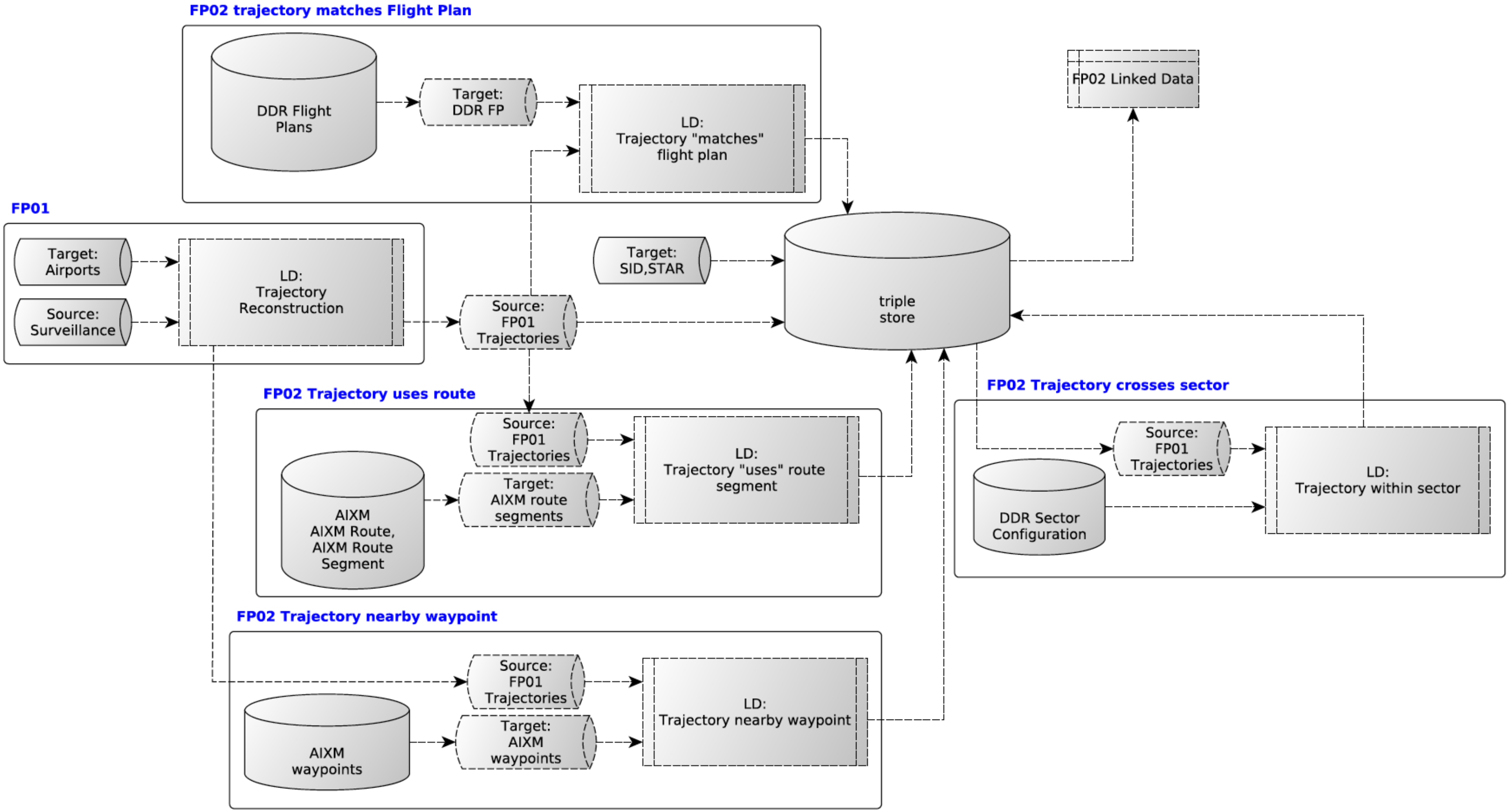


G.M. Santipantakis, C.Doulkeridis, G.A.Vouros, A.Vlachou: *MaskLink: Efficient Link Discovery for Spatial Relations via Masking Areas*. **arXiv:1803.01135 (3 Mar 2018)**

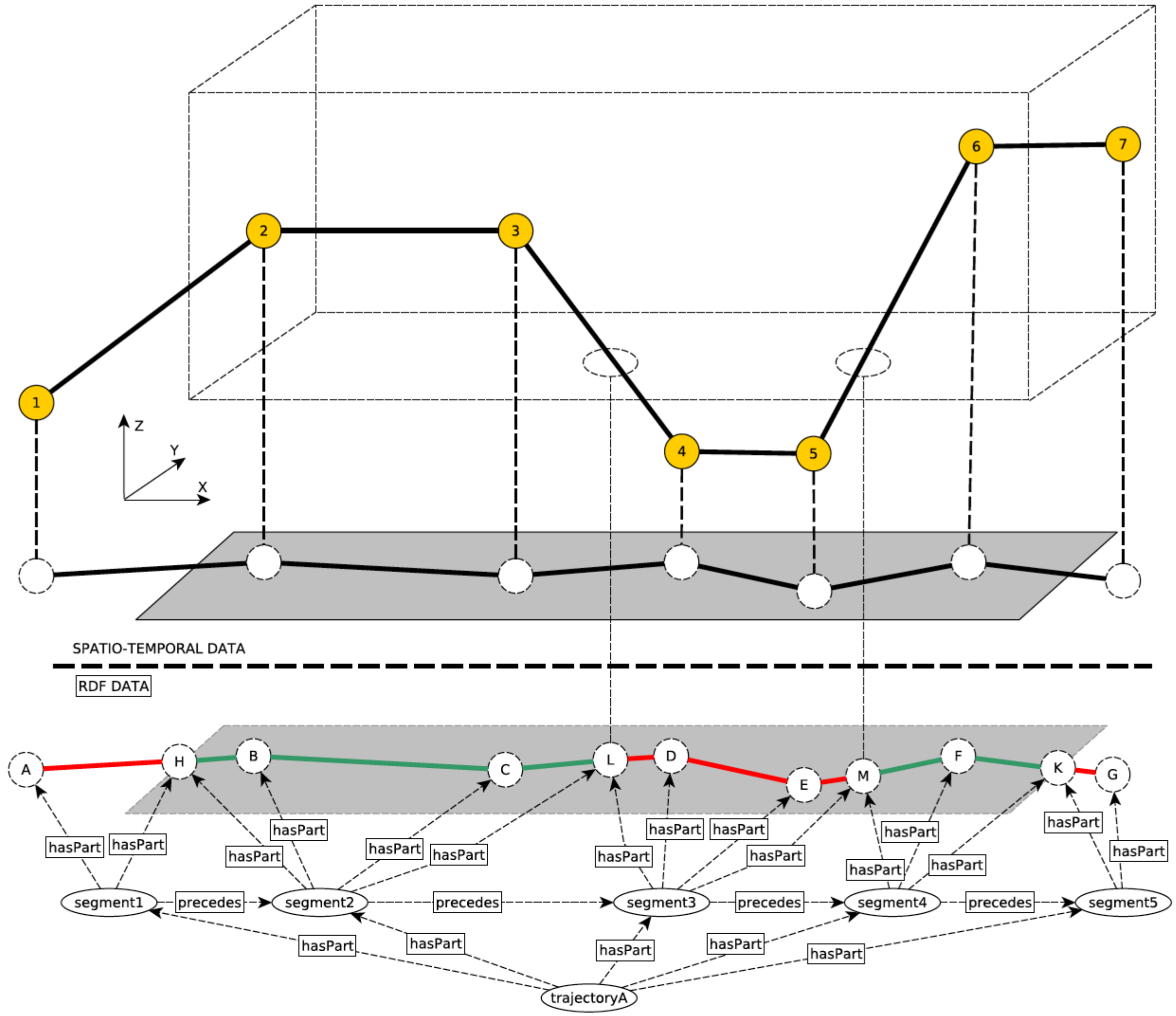
MaskLink on Airports Dataset



Trajectory Enrichment



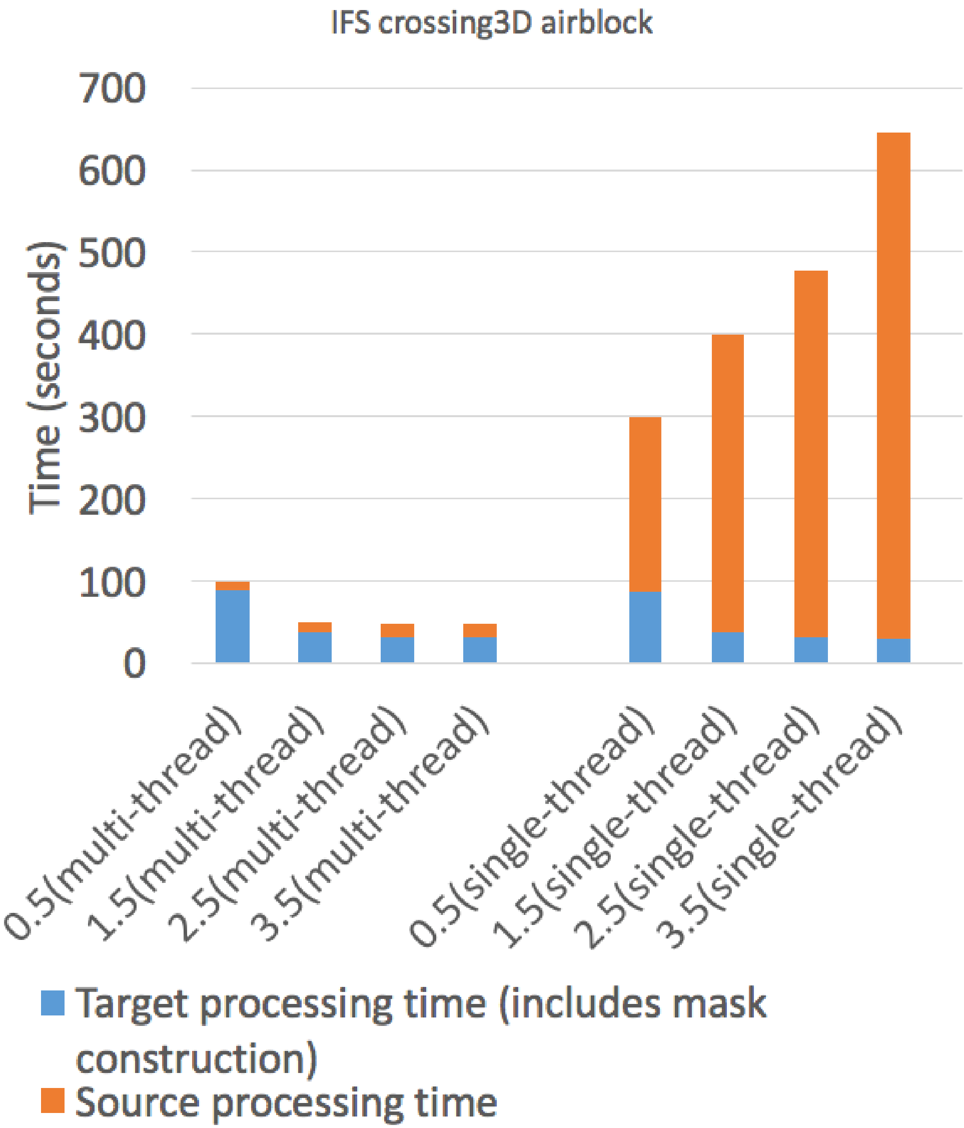
Linking Trajectories with Airblocks



HORIZON 2020

Linking Trajectories with Airblocks

- Source:
  - 1,689,541 positions
  - 8652 trajectories
- Target:
  - 20025 Airblocks
- Throughput (entities per ms):
  - Single threaded: max ~2
  - Multi-threaded: max ~70



datAcron

Online Trajectory Reconstruction

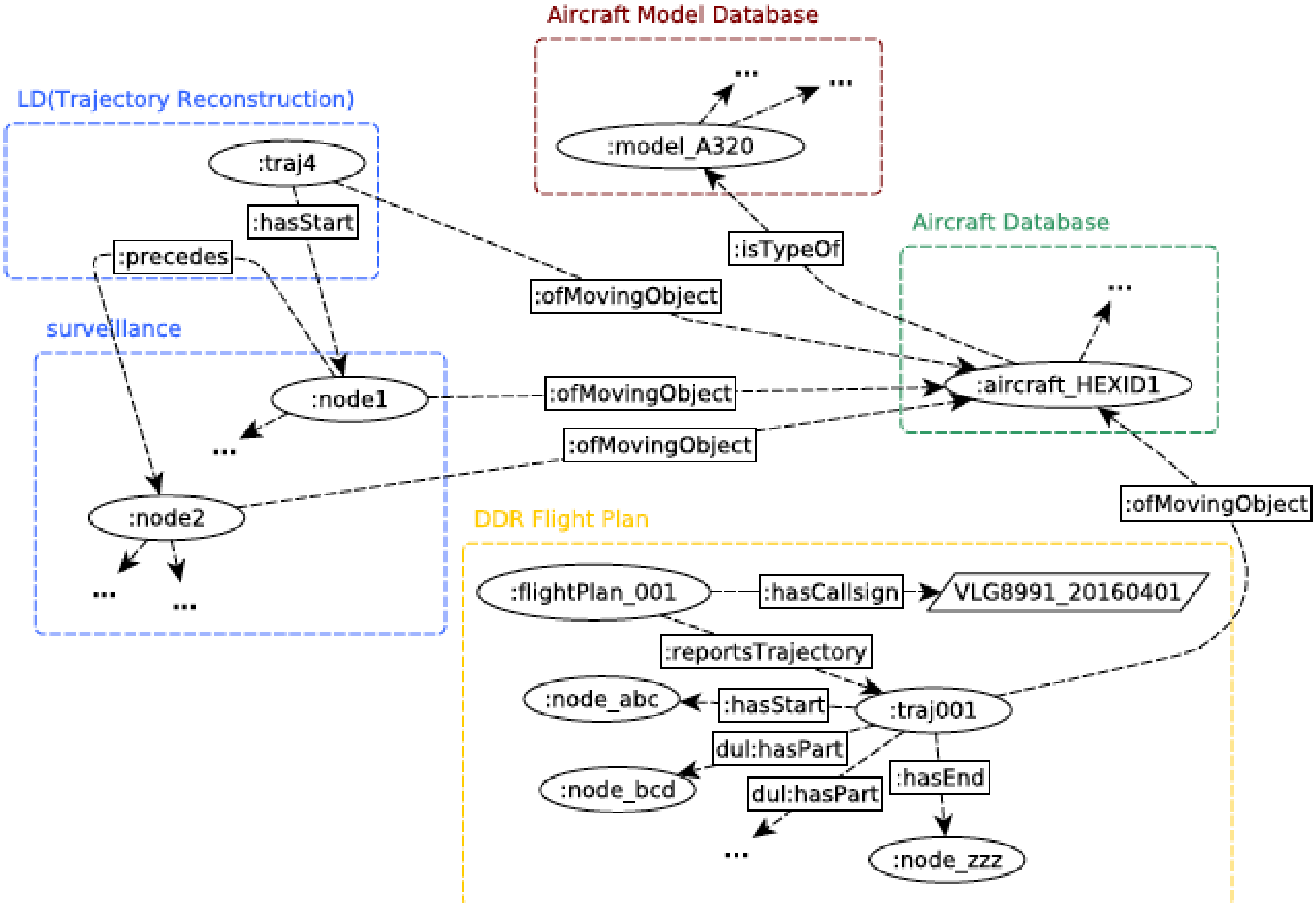
Example of input data:

IFS Radar

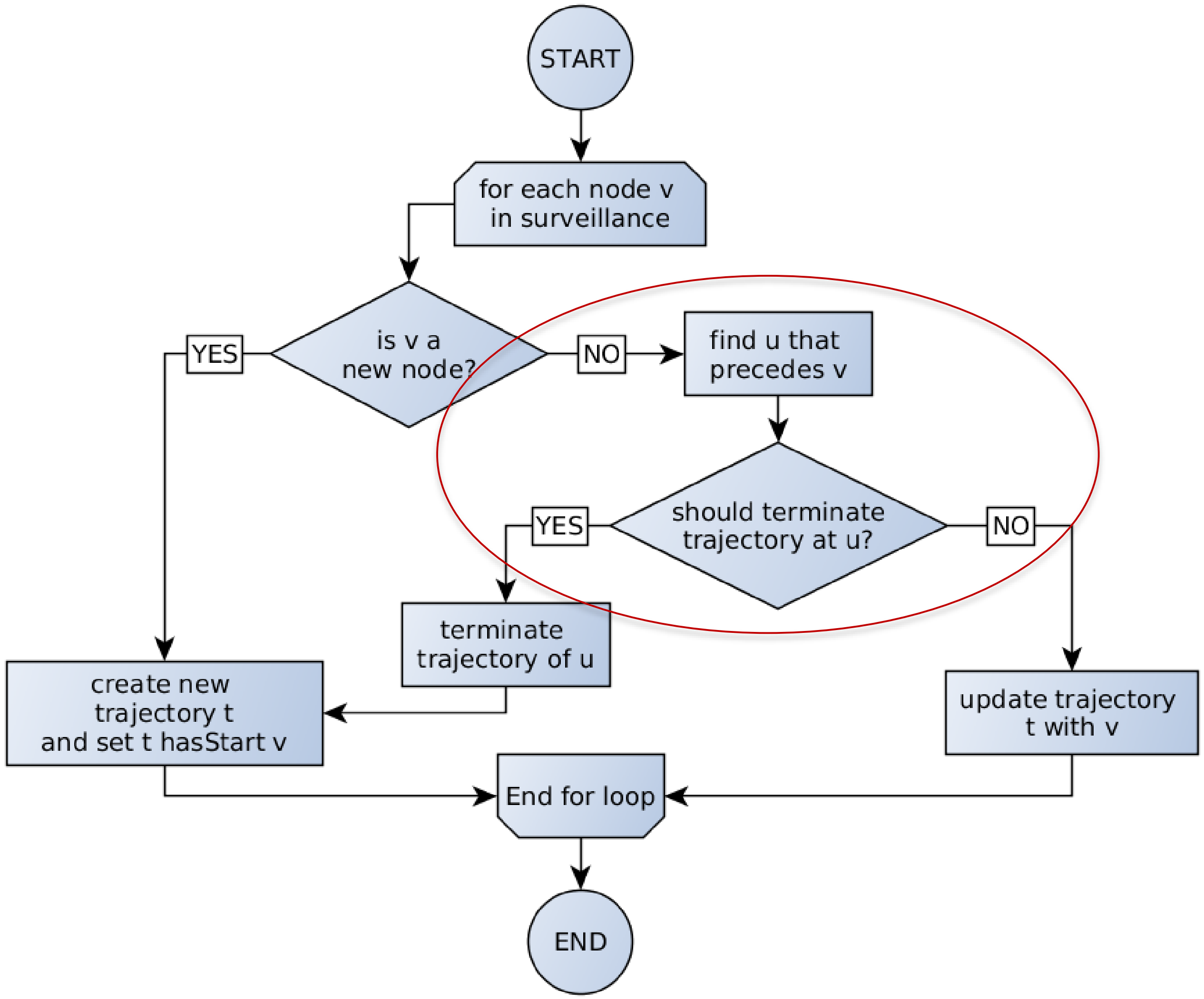
factId;flightKey;callsign;adep;ades;flightRule;wake;aircraft;processDateReference;date_value;time_value;latitude;longitude;modo_c;vel_mod;hdg;vel_x;vel_y;vel_z

4209542619;6737113;IBE6856;SAEZ;LEMD;I;H;A343;2016-04-01;2016-04-01; 01:56:00.0000000;
26.585888;-15.593530;360;464.086;23.198;182.812;426.562;0

Linking ATM Data In Action

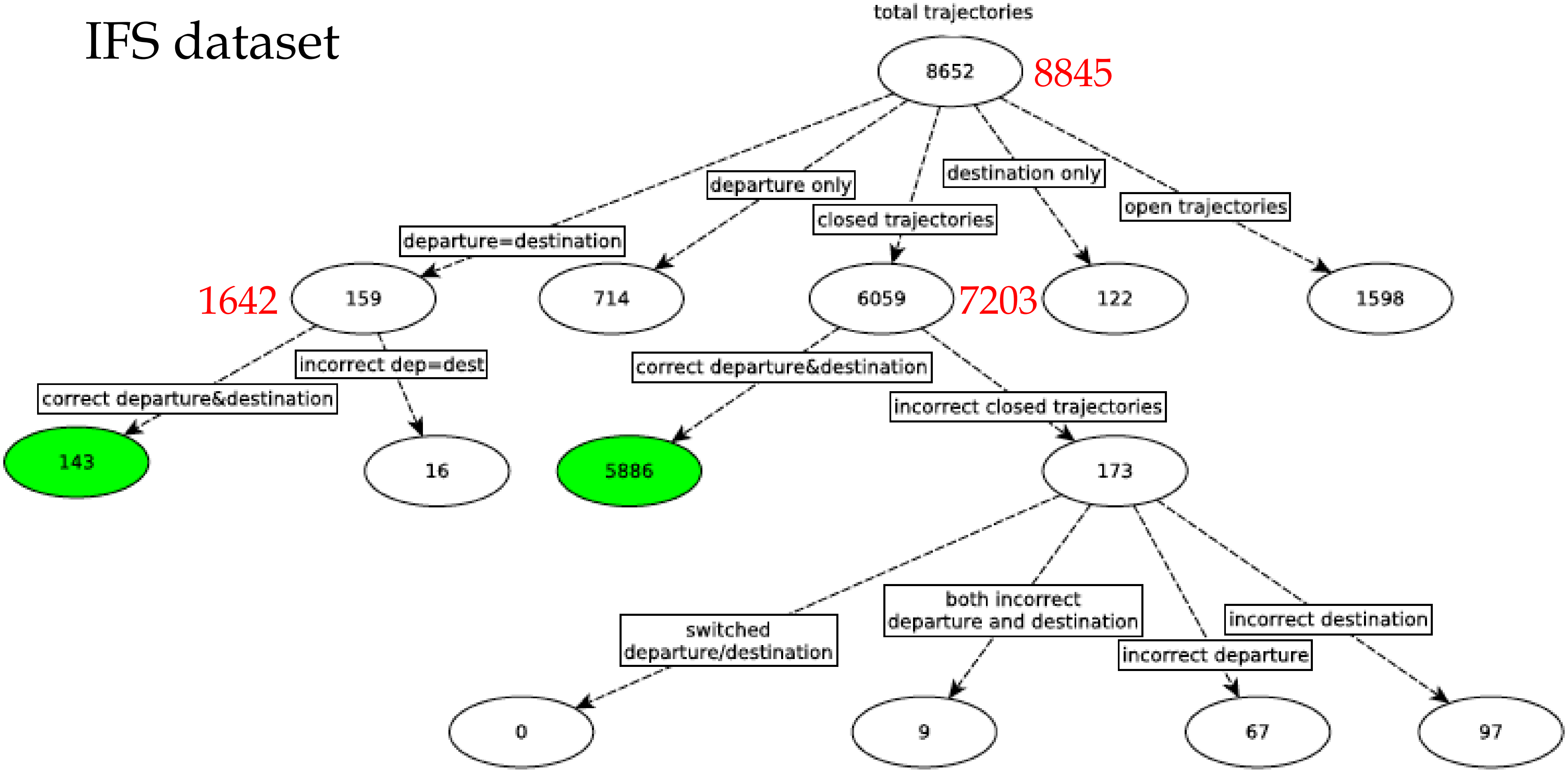


Online Trajectory Reconstruction



Online Trajectory Reconstruction Results

IFS dataset



Conclusions

- datAcron advances data management
  - For streaming and archival data sources
  - Providing integrated and enriched views of data towards trajectory based analytics
  - At different levels of abstraction
  - According to the needs of predictive analytics techniques

Challenges ahead

- Link trajectories in real time
  - E.g. Flight plans/predicted trajectories with actual trajectories (per waypoint)
  - Calculate "distances" between enriched trajectories in real time (and update)
  - Refine trajectory re-construction using more advanced techniques (e.g. based on complex events recognition techniques).

Thank you for your attention!

Follow datAcron developments in:

www.datacron-project.eu

Twitter

linkedin

ResearchGate