Social listening of City Scale Events using the Streaming Linked Data Framework

Authors: Marco Balduini, Emanuele Della Valle, and Daniele Dell'Aglio, Mikalai Tsytsarau, Themis Palpanas, Cristian Confalonieri

Presented by: Ting Xiao 11/18/2013

What is City-Scale Events?

- a group of events (usually with a common topic) located in multiple venues around a city.

Who are interested in?

- City Managers
  - Want to assess the impact on the traffic, pollution, garbage collection
- Organizers
  - Want to monitor the appreciation and popularity of the events
- Sponsors
  - Want to know if their investments are given back in terms of perception and image
- Citizens
  - Want to know the current situation of the events
- Visitors
  - Want to find more popular events

Streaming Linked Data(SLD)

- The framework to collect data streams, analyze them and visualize the results in dashboards
- Exploit Several Semantic Technologies
  - RDF to model and integrate the data
  - SPARQL
  - Sentiment Mining Techniques
- Target
  - Collect All The Necessary Information From the Social Web To Predict More Accurate Results Of The Events
- Applications
  - London Olympic Games 2012
  - Milano Design Week 2013

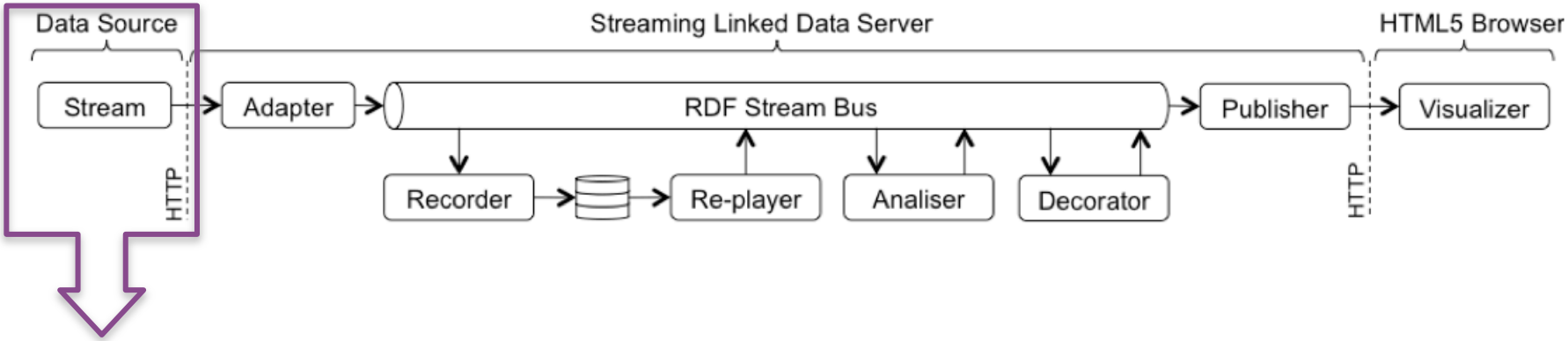
User Cases Analysis

London Olympic Games(2012)	Milano Design Week (2013)
Is it possible to detect the Olympic Games-related events analyzing the Twitter streams?	Is MDW visible in the social streams posted by people in Milano area? If yes in real-time, (a) What are the districts from which MDW visitors post the most? (b)The most frequently used hash-tags? (c) How do people feel before, during and after the event they join?
Is it possible to track the movement of the crowds through geo-tagged tweets?	Is the launch of ASUS products during MDW visible in the social streams posted by people around the world? If yes, not necessarily in real-time, (a) What are the products that attract more attention? (b) What is the global sentiment before, during and after the launch?

Technical Requirements

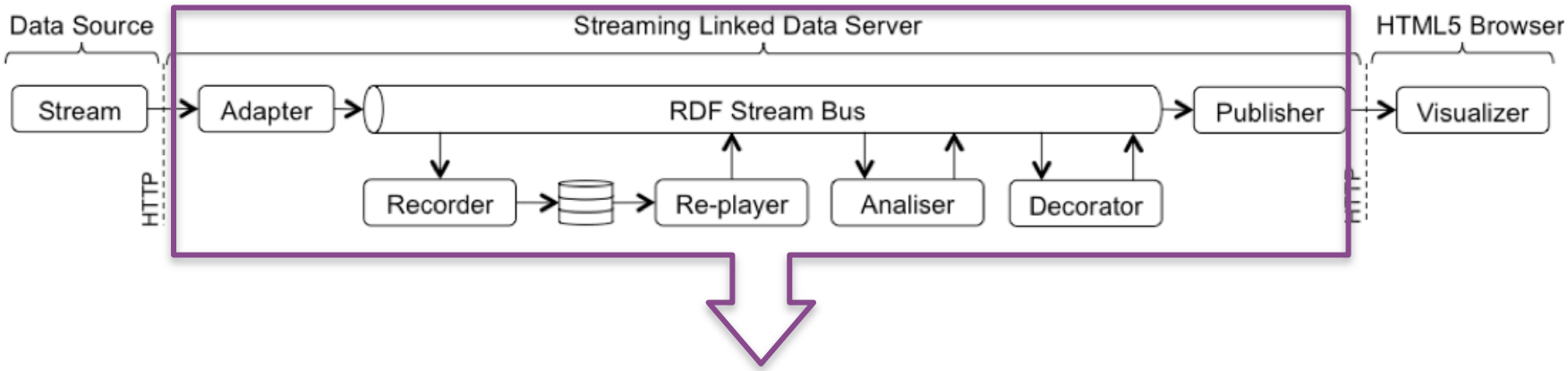
- Accessing the social stream
- Recoding and replaying portions of the social stream
- Decorating the social stream with sentiment information
- Continuously analyzing the social stream
- Internally streaming partial results of the analysis
- Publishing and visualizing continuous analysis results

Architecture of SLD Framework



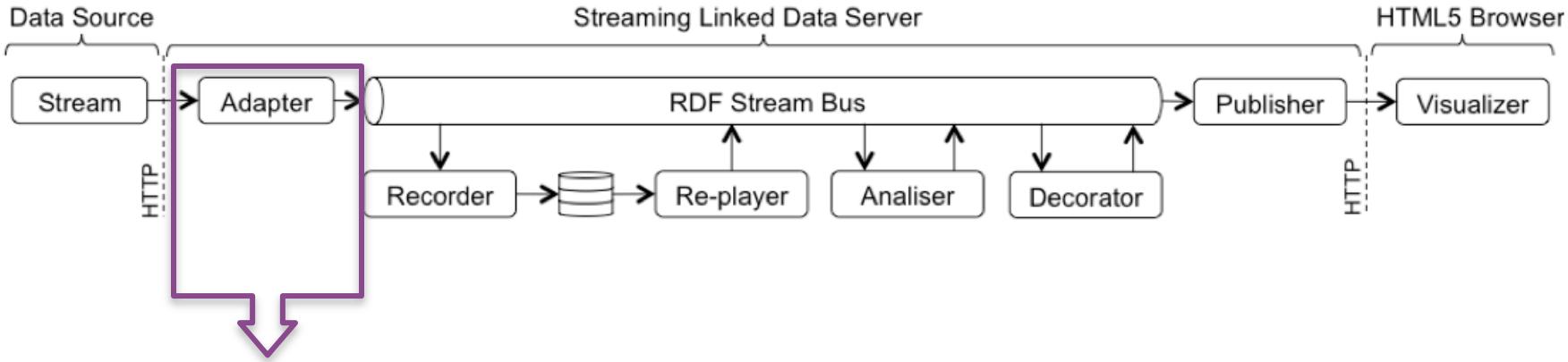
- Streaming Data Source
  - Only used the Twitter Stream Data

Architecture of SLD Framework



- The Central Part Of the SLD
  - Accessing data stream sources
  - Internally streaming data
  - Register and replay the portion of data stream
  - Decorate and analyze time-boxed potion of the stream
  - Publish the result

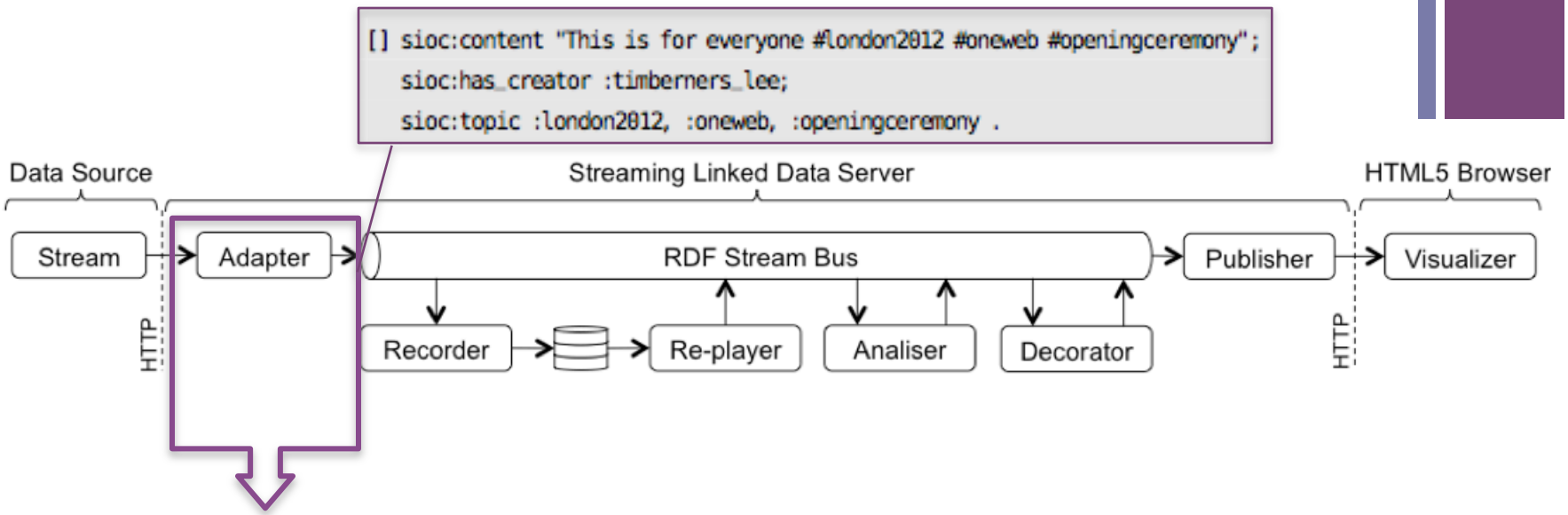
Architecture of SLD Framework



- Access data stream resources
- Do filtering operation to the data source
- Translate the stream data into a set of timestamped RDF triples

Ps, their framework includes several sensor networks adapters, such as Instagram, Foursqure

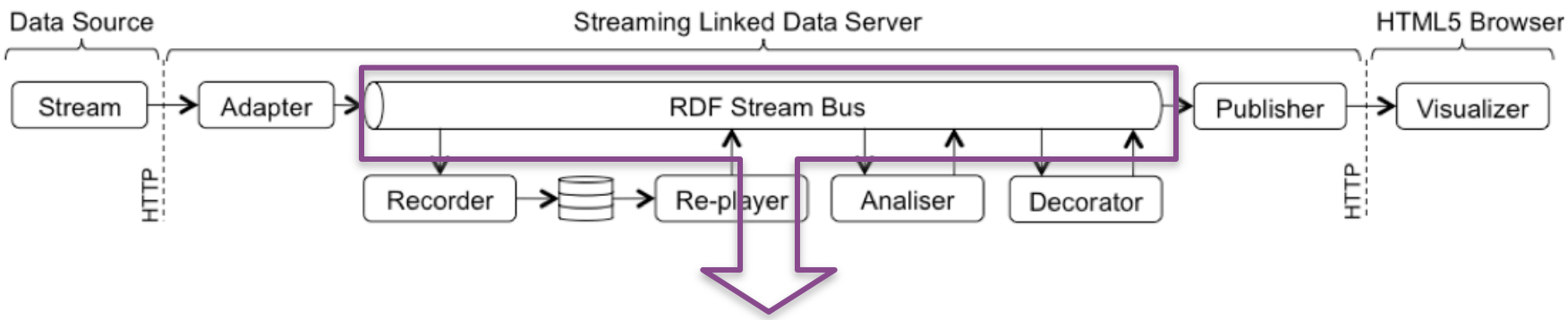
Architecture of SLD Framework



- Access data stream resources
- Do filtering operation to the data source
- Translate the stream data into a set of timestamped RDF triples

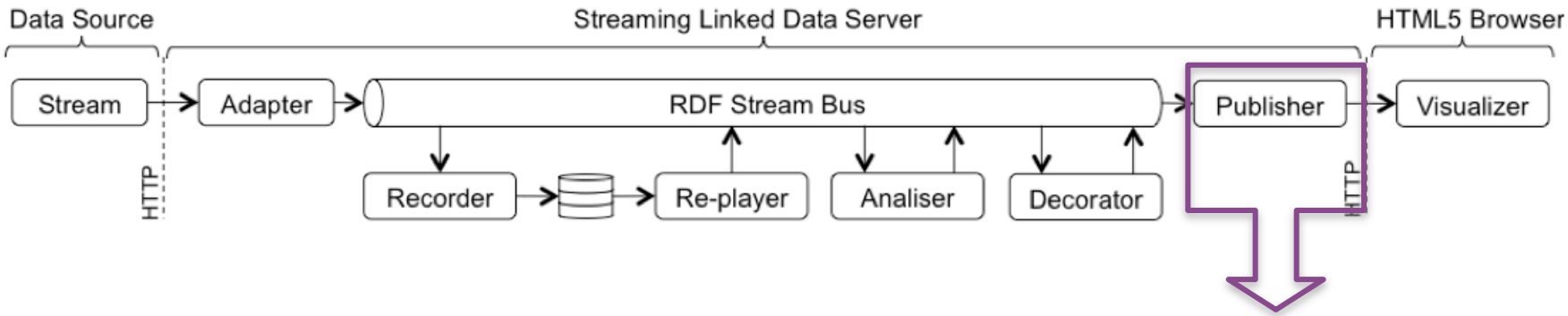
Ps, their framework includes several sensor networks adapters, such as Instagram, Foursqure

Architecture of SLD Framework



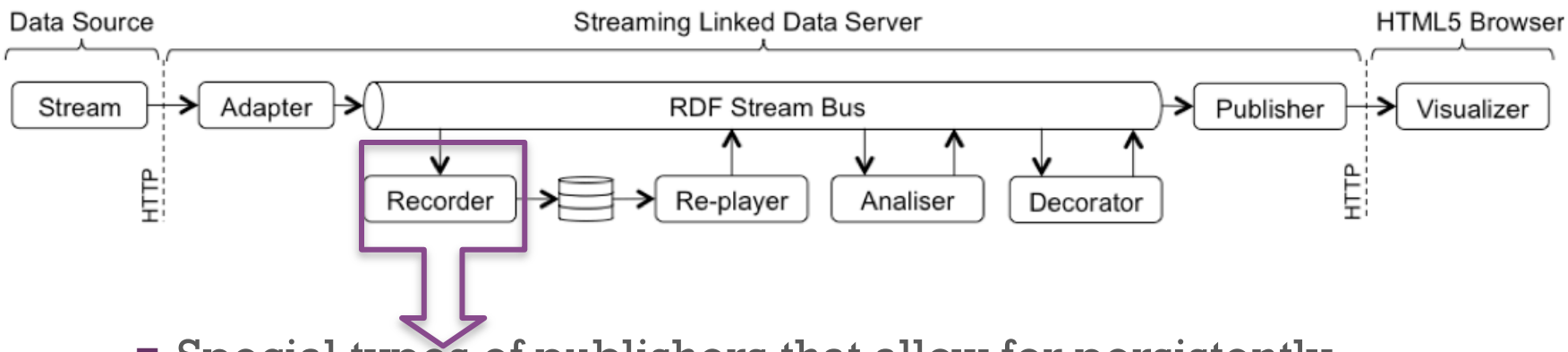
- Support the publish/subscribe communication among the internal components of streaming linked data.

Architecture of SLD Framework



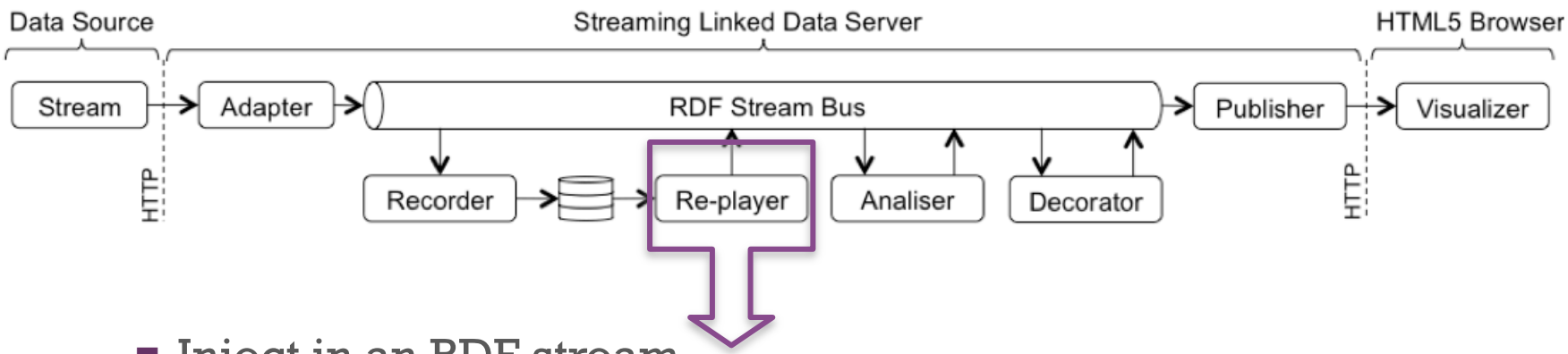
- Make available on Web the content of chosen RDF stream(R.6)
  - Two SLD formats:
    - Instantaneous Graph(iGraphs) : a set triples with same timestamp
    - Stream Graph(sGraphs): a set triples point to one or more timestamped iGraphs

Architecture of SLD Framework



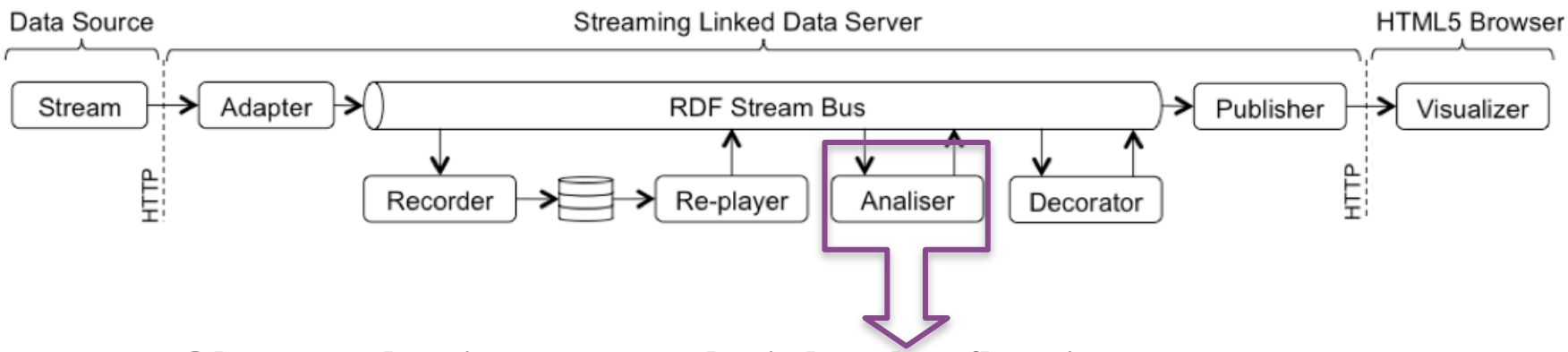
- Special types of publishers that allow for persistently storing a part of an RDF stream
  - Data format: iGraphs + rGraphs
  - rGraphs: similar with sGraphs but they include pointers to all the iGraph recorded and such pointers do not have a time interval of validity.

Architecture of SLD Framework



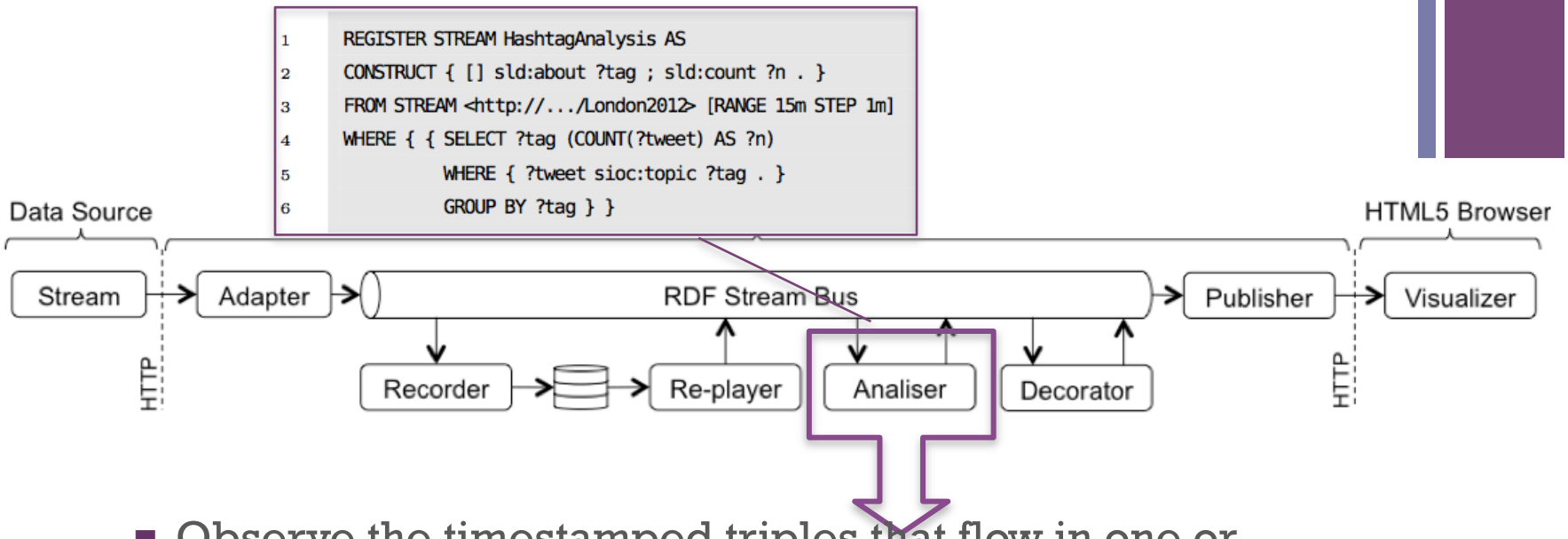
- Inject in an RDF stream
  - Data format: iGraphs

Architecture of SLD Framework



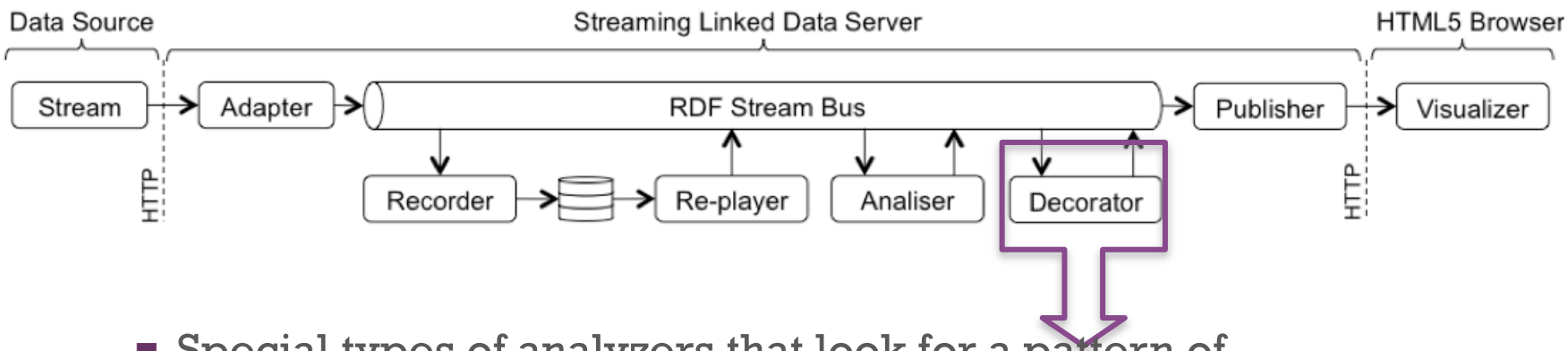
- Observe the timestamped triples that flow in one or more RDF stream
- Perform analysis on them
- Generate a continuous stream of answers
- SPARQL can be plugged in here and used for analyzing

Architecture of SLD Framework



- Observe the timestamped triples that flow in one or more RDF stream
- Perform analysis on them
- Generate a continuous stream of answers
- SPARQL can be plugged in here and used for analyzing

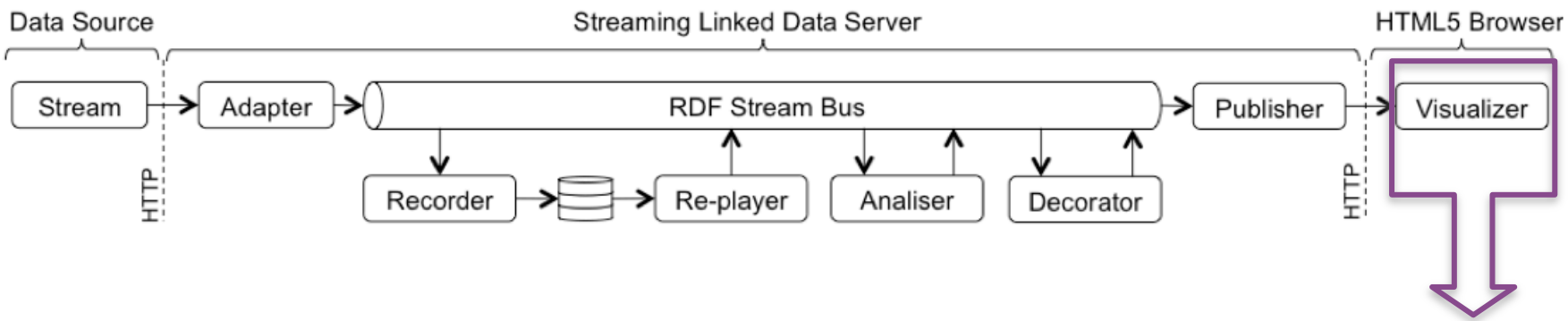
Architecture of SLD Framework



- Special types of analyzers that look for a pattern of triples in RDF stream. When the pattern matches, it run a computation of the matching and add new triples to the stream.

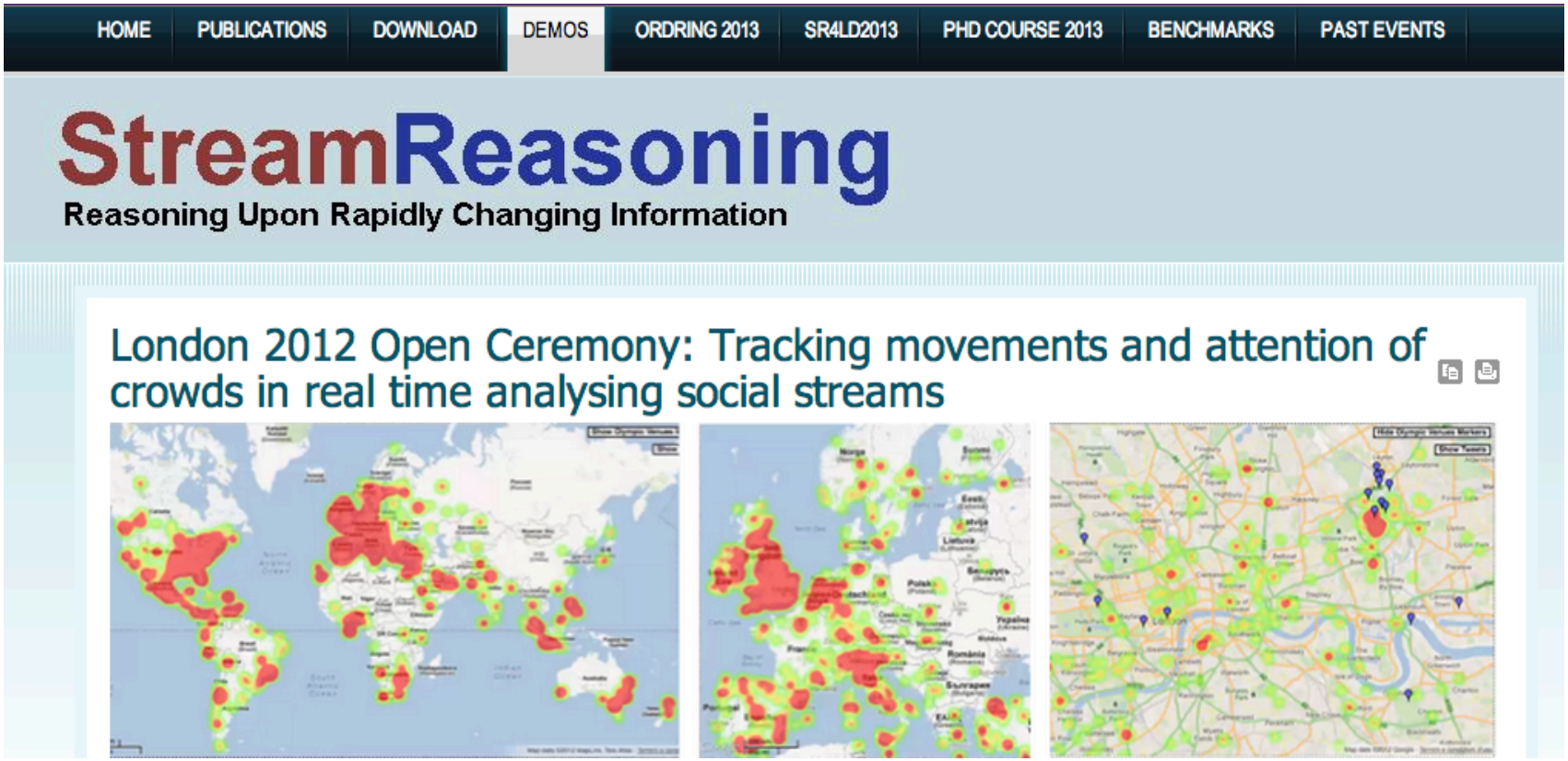
- Include a sentiment mining component, using a dictionary-based sentiment classifier(University of Trento)

Architecture of SLD Framework



- A library of visual widgets using HTML5 visualize what is published periodically.

Application(1)
London Olympic Games 2012



http://streamreasoning.org/demos/london2012

Application(1)
London Olympic Games 2012

- **Detecting Events**: detect the events given the position of a set of venues and socially listen their surroundings
  - 3 million tweets used
  - July 25th to August 13th (2012)
  - Three venues: big, medium and small
  - Ground truth: calendar of Olympic Games
- **Visualizing Crowd Movements:**
  - The method looks for a sequence of bursts detected first at public transport stations, then in the walkable areas outside the venues and finally in one of the venues.

Application(1)
London Olympic Games 2012

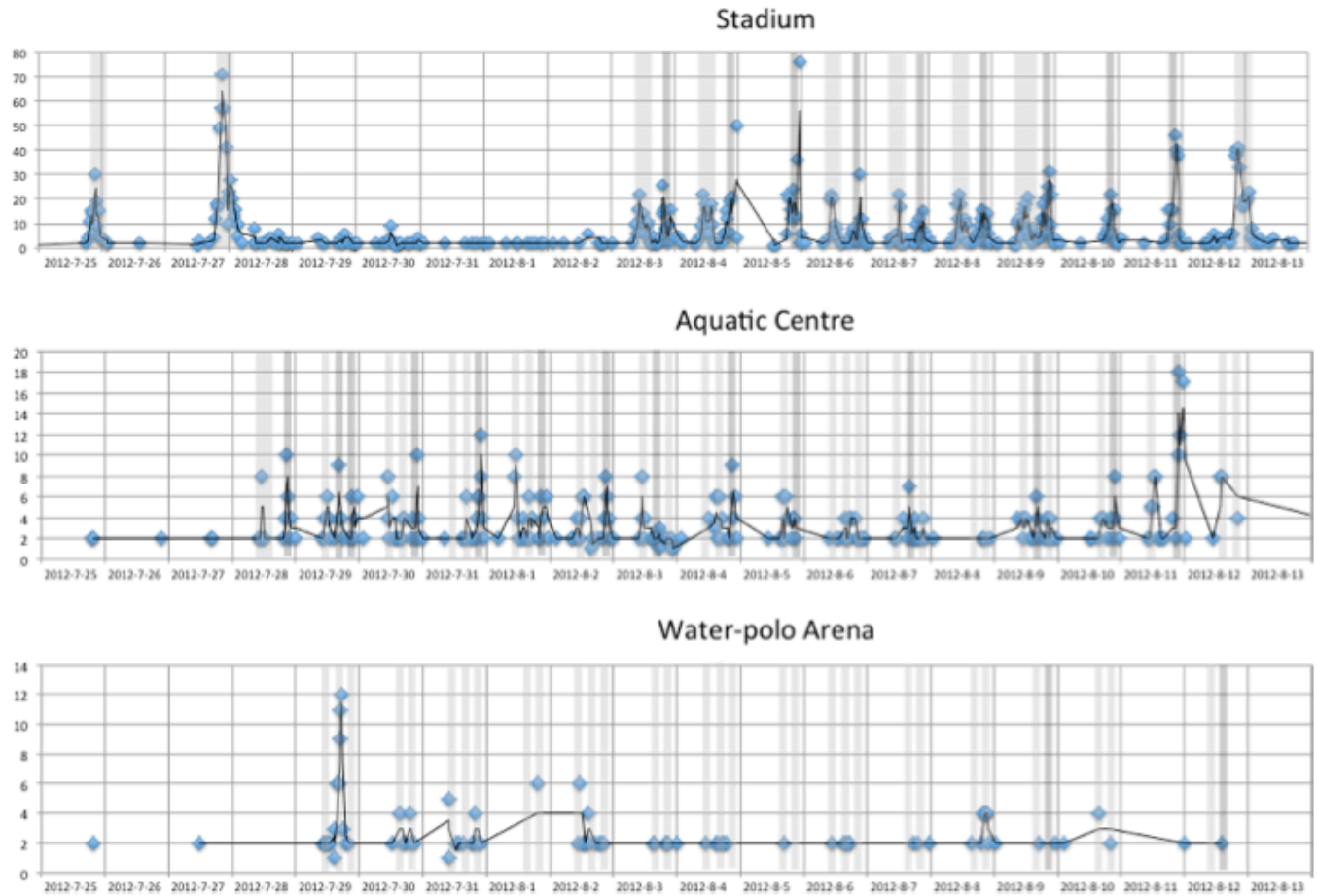
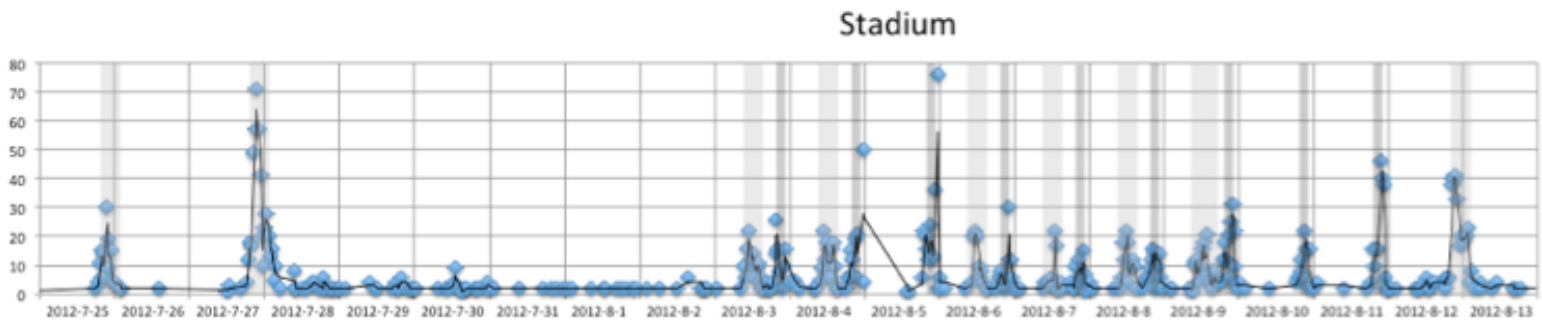


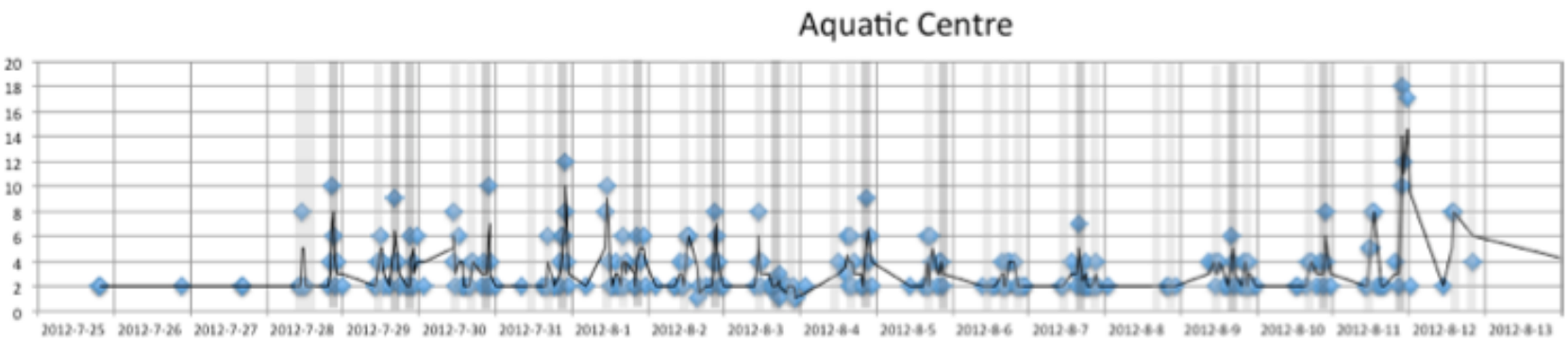
Fig. 2: The results of the event detection experiment.

Application(1)
London Olympic Games 2012



- In the stadium, SLD was able to detect all events in the ground truth:
  - the rehearsal for the opening ceremony on July 25th; the opening ceremony on July 27th;
  - the pair of events scheduled on August 3rd, 4th, and 6th to 9th;
  - the single event on August 5th, 10th, and 11th;
  - The closing ceremony on August 12th.
- The magnitude of the burst is related to the importance of the event
  - On August 4th took place the women's 100 metres final, and on August 5th the men's 100 metres final.

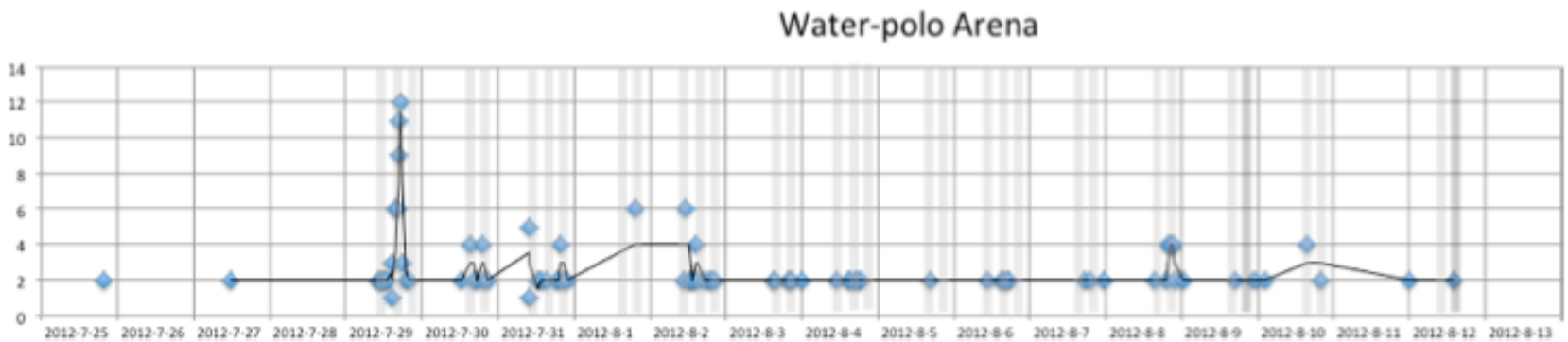
Application(1)
London Olympic Games 2012



- In the aquatic arena, which attracts less attention in terms of tweets, our method performed with a high precision(i.e., only three unscheduled events were detected before the opening ceremony), but with a recall of 76% (32 events out of the 42 planned).

- In this case the magnitude of the burst speaks for the importance of the event: most of the finals have high peaks.

Application(1)
London Olympic Games 2012



- In the water polo arena, which is a small venue hosting a single sport, our method was still precise, but the recall was very low (32%, i.e., 11 events out of the 34 planned).
- The only event that generated a large burst was on July 29th.

Application(1)
London Olympic Games 2012

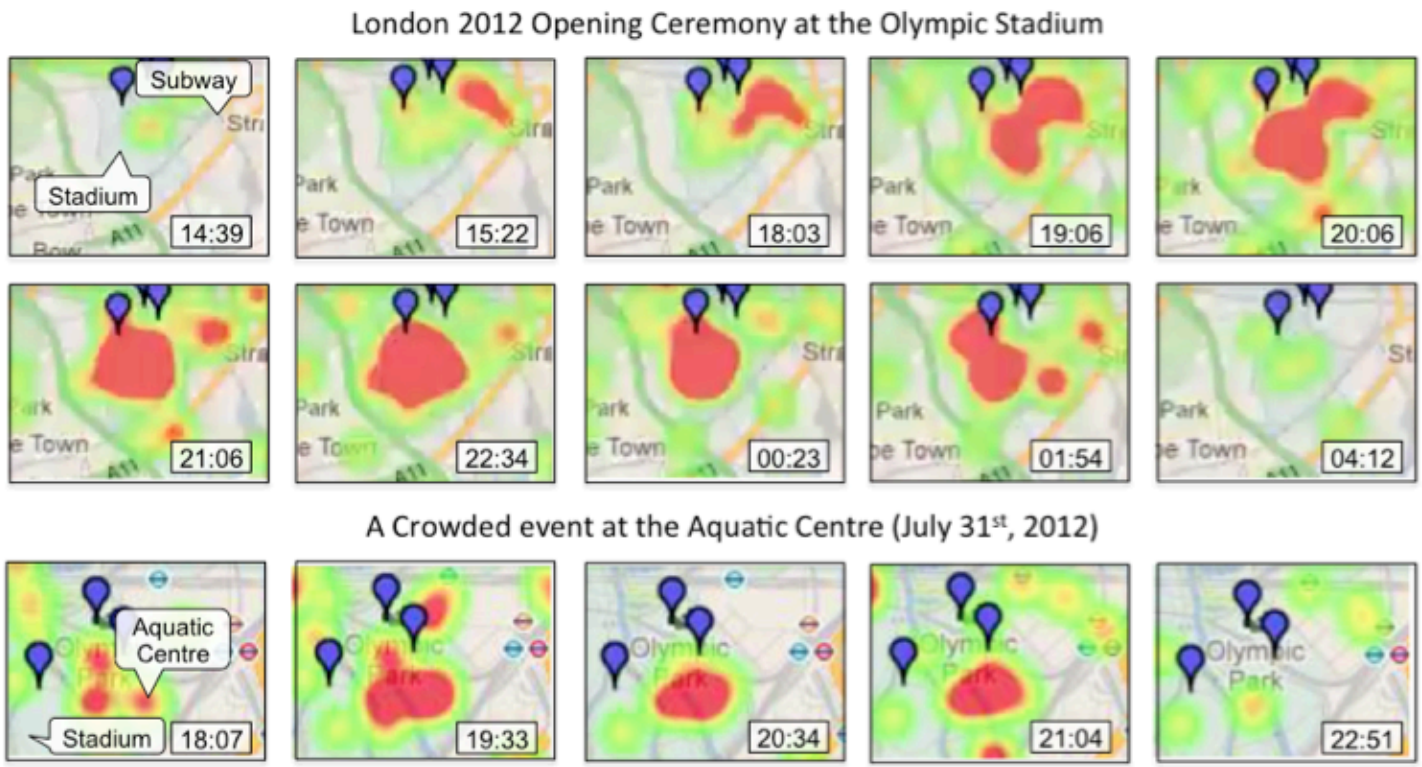
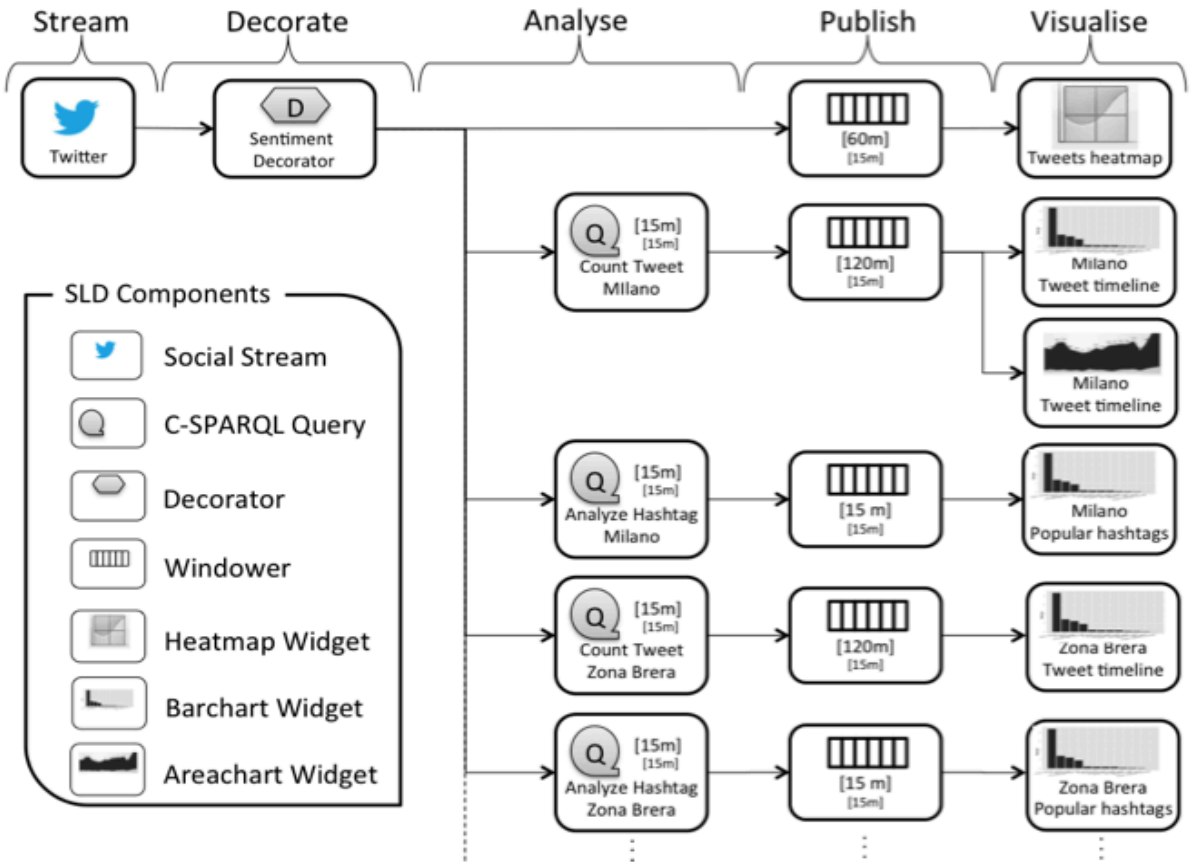


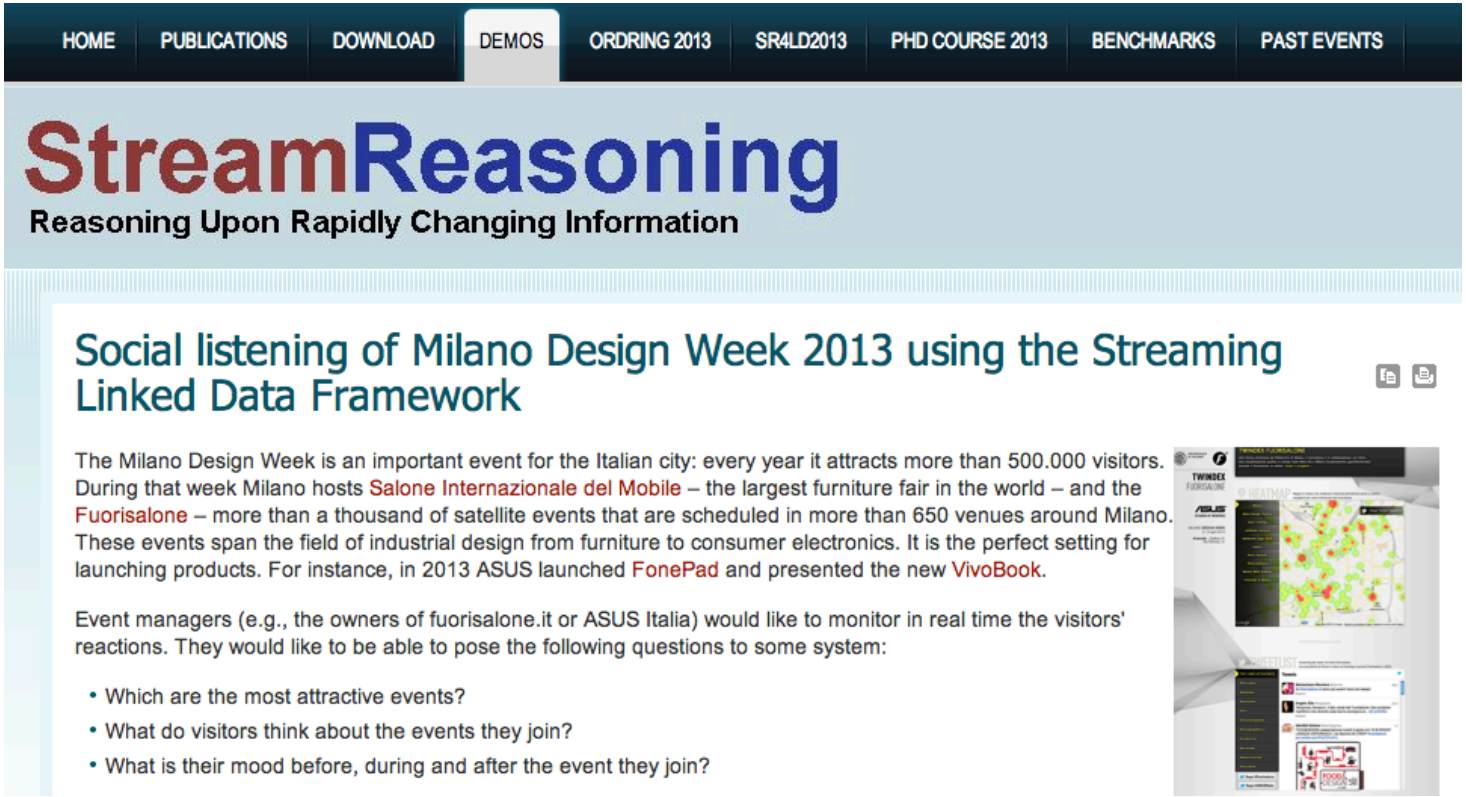
Fig. 3: The sequence of heatmaps visualises the flows of crowd from the public transports to the Olympic venues in two different scenarios.

Application(2)
Milano Design Week 2013



Twindex Fuorisalone Layout

Application(2)
Milano Design Week 2013



http://www.streamreasoning.org/demos/mdw2013

Application(2)
Milano Design Week 2013

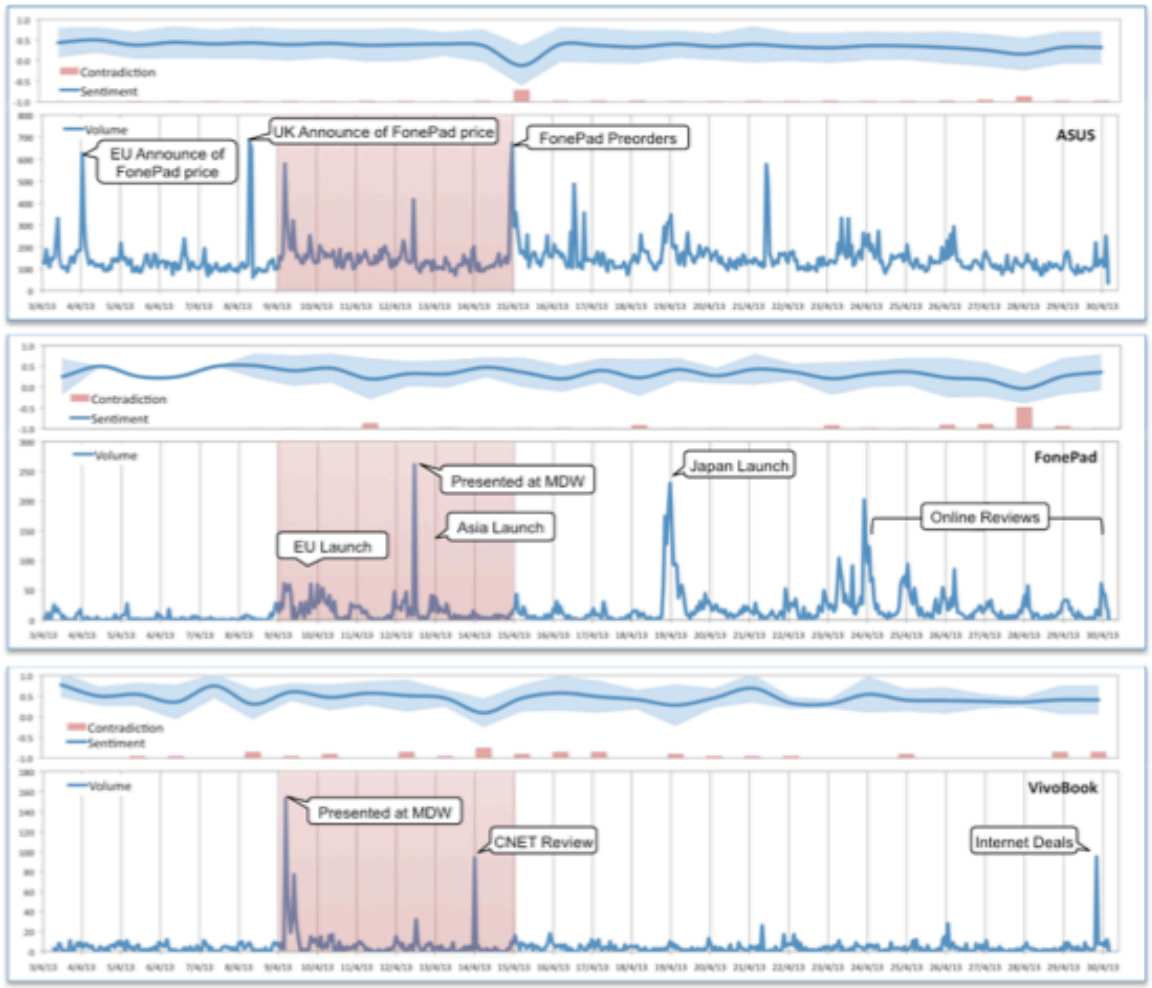
- Monitor in real time during the MDW 2013 on the tweets posted from Milano
  - 106,770 tweets
  - April 8th and April 17th, 2013 (Still Running)
- HTML5 dashboard was deployed and it was accessible to organizers and visitors of the event
  - http://www.streamreasoning.org/live/slm/
- The post-event analysis
  - Results collection of 107,044,487 tweets(MDW, ASUS and its products)
  - April 8th and April 30th, 2013

Application(2)
Milano Design Week 2013

April 9th, 2013 at 18.00		April 15th, 2013 at 18.00	
fuorisalone	30	inter	20
designweek	28	diretta	11
nabasalone	20	cagliari	6
milano	9	milan	4
design	6	seriea	3

- A comparison between the top-5 hashtags used in geo-tagged tweets in Milano during a late afternoon of MDW 2013 and one after MDW

Application(2)
Milano Design Week 2013



- A comparison Results of the sentiment analysis carried out on the tweets about ASUS and two of its products: FonePad and VivoBook

Conclusion

- Describe and analyze the concrete problems and user requirements for social listening of city-scale events.
- Provide the Streaming Linked Data (SLD) framework and sentiment mining techniques adapted for streaming.
- Report on the pragmatics of deploying and using of SLD to monitor two city-scale events: the London Olympic Games 2012 , and the Milano Design Week 2013. These use cases prove the feasibility of the framework.

Thanks

- Any Questions?