

Tutorial on RDF Stream Processing 2016

M.I. Ali, J-P Calbimonte, D. Dell'Aglio,
E. Della Valle, and A. Mauri

<http://streamreasoning.org/events/rsp2016>



2016

Kobe, Japan

The 15th
International
Semantic Web
Conference

Insight 

Hes·so VALAIS WALLIS


 Universität
Zürich ^{UZH}

 POLITECNICO
MILANO 1863




Conclusions

Daniele Dell'Aglio

✉ dellaglio@ifi.uzh.ch

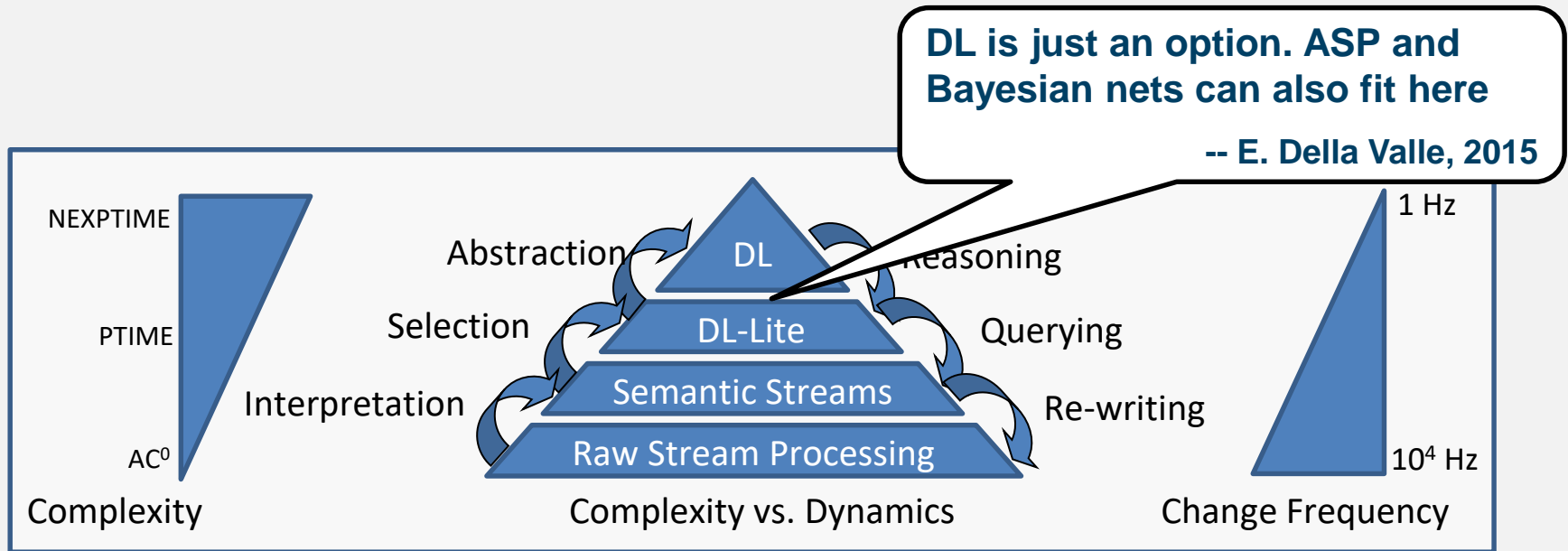
 <http://dellaglio.org>

 @dandellaglio

- This work is licensed under the Creative Commons Attribution 3.0 Unported License.
- **Your are free:**
 -  **to Share** — to copy, distribute and transmit the work
 -  **to Remix** — to adapt the work
- **Under the following conditions**
 -  **Attribution** — You must attribute the work by inserting
 - “[source <http://streamreasoning.org/events/rsp2016>]” at the end of each reused slide
 - a credits slide stating
 - These slides are partially based on “Tutorial on RDF Stream Processing 2016” by M.I. Ali, J-P Calbimonte, D. Dell'Aglio, E. Della Valle and Andrea Mauri <http://streamreasoning.org/events/rsp2016>
- To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>

- Research question and approach
- Research challenges
- Achievements & open issues
- What's next?
- More on RSP at ISWC 2016

- Proposed approach



H. Stuckenschmidt, S. Ceri, E. Della Valle, F. van Harmelen: **Towards Expressive Stream Reasoning**. Proceedings of the Dagstuhl Seminar on Semantic Aspects of Sensor Networks, 2010.

- Relation with DSMSs and CEPs
 - Just as RDF relates to data-base systems?
- Data types and query languages for semantic streams
 - Just RDF and SPARQL but with continuous semantics?
- Reasoning on Streams
 - Theory: formal semantics
 - Efficiency
 - Scalability and approximation
- Dealing with incomplete & noisy data
 - Even more than on the current Web of Data
- Distributed and parallel processing
 - Streams are parallel in nature, data stream sources are distributed, ...
- Engineering Stream Reasoning Applications
 - Development Environment
 - Integration with other technologies
 - Benchmarks as rigorous means for comparison

■ Achievement

- Somehow just as RDF, SPARQL, and OWL relate to data-base systems

| DB → Semantic Web | DSMS/CEP → Semantic Web |
|--------------------------|--------------------------------------|
| Relational data → RDF | Data streams → RDF Streams |
| SQL → SPARQL | CQL/EPL/... → C-SPARQL/EP-SPARQL/... |
| Schema → OWL | Schema → OWL |

- But with some differences
 - Queries are registered → opportunity for query optimizations
 - Many application requires a network of queries → opportunity for inter-query optimizations

■ Issues

- Relational and RDF streams should live together on the Web and in the RDF Stream Processing engines

- **RDF streams introduced** as new data type in the Semantic Web and Linked Data research

- **Multiple notions of RDF stream** proposed
 - Ordered sequence (implicit timestamp)
 - One timestamp per triple (point in time semantics)
 - Two timestamps per triple (interval base semantics)

- Comparison between existing approaches

| System | Data item | Time model | # of timestamps |
|--------------------------|-----------|---------------|-----------------|
| INSTANS | triple | Implicit | 0 |
| C-SPARQL | triple | Point in time | 1 |
| SPARQL _{stream} | triple | Point in time | 1 |
| CQELS | triple | Point in time | 1 |
| Sparkwave | triple | Point in time | 1 |
| Streaming Linked Data | RDF graph | Point in time | 1 |
| ETALIS | triple | Interval | 2 |

- **Harmonization** of current proposals **ongoing** at the **W3C RSP community group**

- **Languages for continuous querying** of and **event processing** on RDF streams **proposed**
- Window base selection outperforms filter base selection
- Dynamic optimization of query plans and incremental evaluation is possible
- Multiple RDF stream processor **prototypes** implemented and deployed

- Comparison between existing approaches

| System | S2R | R2R | Time-aware | R2S |
|--------------------------|--------------------------|------------------|---|-----------------|
| INSTANS | Based on time events | SPARQL update | Based on time events | Ins only |
| C-SPARQL Engine | Logical and triple-based | SPARQL 1.1 query | timestamp function | Batch only |
| SPARQL _{stream} | Logical and triple-based | SPARQL 1.1 query | no | Ins, batch, del |
| CQELS | Logical and triple-based | SPARQL 1.1 query | no | Ins only |
| Sparkwave | Logical | SPARQL 1.0 | no | Ins only |
| Streaming Linked Data | Logical and graph-based | SPARQL 1.1 | no | Batch only |
| ETALIS | no | SPARQL 1.0 | SEQ, PAR, AND, OR, DURING, STARTS, EQUALS, NOT, MEETS, FINISHES | Ins only |

- Harmonization** of current proposals **ongoing** at the **W3C RSP community group**

- The **existing engines**
 - adopts **different architectural** choices and it is still unclear when each choice is best
 - C-SPARQL, ETALIS, SPARQL_{stream} are wrappers for existing systems thus they are more reliable and maintainable
 - CQELS, Streaming Linked Data, INSTANS, Sparkwave are native implementations, thus they are more efficient and offer optimizations not possible in the other system
 - They have **different operational semantics**
- Proposal
 - Daniele Dell'Aglio, Emanuele Della Valle, Jean-Paul Calbimonte, Oscar Corcho: **RSP-QL Semantics: A Unifying Query Model to Explain Heterogeneity of RDF Stream Processing Systems**. Int. J. Semantic Web Inf. Syst. 10(4): 17-44 (2014)

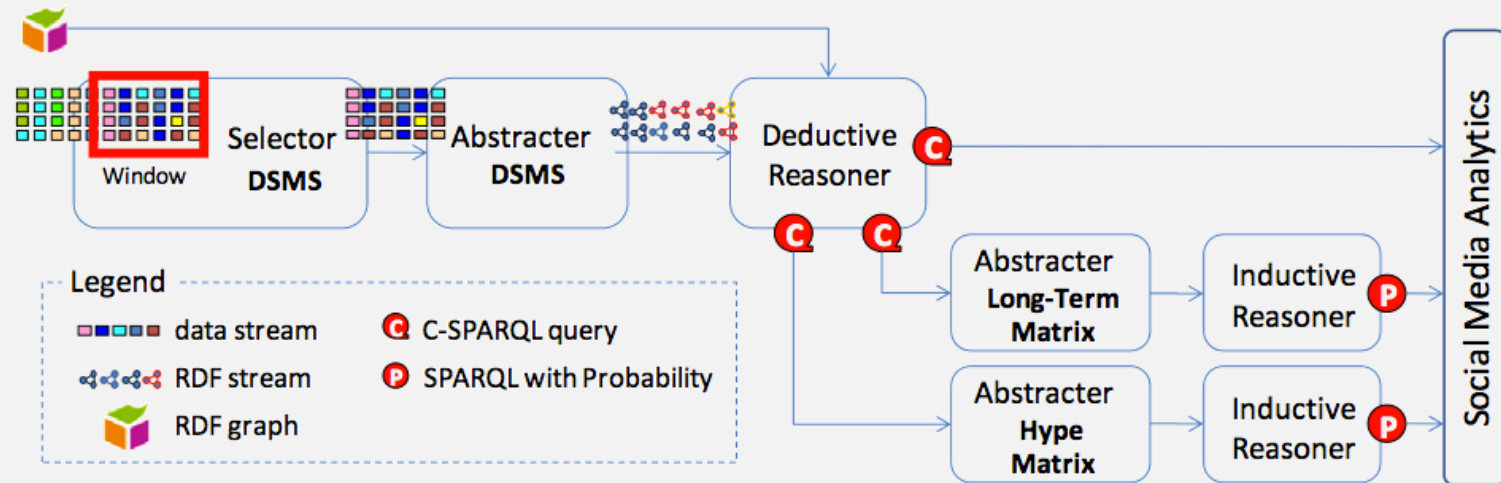
- **Stream Reasoning** research field is getting momentum
- Efficient **continuous reasoning algorithm** on RDF streams for RDFS, RDFS++, EL++, Answer Set Programming were proposed
- Formal semantics of Stream Reasoning is under investigation
 - Stream Reasoning with ASP
 - STARQL
 - LARS
- Multiple Stream Reasoning **proofs of concept** were implemented

- **Continuous reasoning** requires more investigations
 - **Rewriting** of continuous conjunctive queries under OWL2QL entailment regime
 - **inconsistency** and **negation** in continuous reasoning tasks
 - Remove the assumption that **ontologies cannot change**
 - Extend ontological languages to allow for modelling aggregates and temporal operators

- **Logic based time-management**
 - From point in time to interval based semantics
 - More expressive specification, e.g., calendar algebra
 - Windows that logically resize at runtime

- Explore **more reasoning form beyond Q/A**, e.g., planning

- Data streams are incomplete and noisy!
- Achievements
 - **probabilistic ASP** (see morning slides)
 - Combining **deductive and inductive Stream Reasoning**



D.F. Barbieri, et a. Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics. IEEE Intelligent Systems 25(6): 32-41 (2010)

- Issues
 - More research required!

- Data streams are **parallel and distributed** in nature!
- Achievements
 - Active Field of research
 - Jacopo Urbani, Alessandro Margara, Cerial J. H. Jacobs, Frank van Harmelen, Henri E. Bal: DynamITE: Parallel Materialization of Dynamic RDF Data. International Semantic Web Conference (1) 2013: 657-672
 - Danh Le Phuoc, Hoan Nguyen Mau Quoc, Chan Le Van, Manfred Hauswirth: Elastic and Scalable Processing of Linked Stream Data in the Cloud. International Semantic Web Conference (1) 2013: 280-297
 - Chang Liu, Jacopo Urbani, Guilin Qi: Efficient RDF stream reasoning with graphics processing units (GPUs). WWW (Companion Volume) 2014: 343-344
- Issues
 - More research required!

- Achievements
 - Deployments for
 - **semantic sensor networks**
 - **social media analytics**
 - **fusions** of those above
 - Multiple **benchmarks** proposed
 - SRbench, CSRbech
 - LSbench
 - CityBench

- Issues
 - It is still **unclear when and where** it is convenient to **adopt Stream Reasoning** technologies
 - **Benchmarks KPI** too focused on throughput; correctness and memory allocation cost, too
 - Lack of an community **shared infrastructures** to run repeatable, reproducible, comparable experiments on



- Data types and query languages for semantic streams
 - Notion of RDF stream :-)
 - Languages for continuous querying :-)
 - Prototypes :-)
 - Standardization :-|

- Reasoning on RDF streams
 - Theory :-|
 - Algorithms :-)
 - Prototypes :-|

- Dealing with incomplete & noisy data
 - Theory :-)
 - Algorithms :-|
 - Prototypes :-|

- Engineering Stream Reasoning Applications
 - Deployments :-)
 - Benchmarks :-|

- Observation: order reflects recency, relevance, trustability ...

| | | | |
|--|-------------------------------|--|-----------------------|
|  <p>Types of orders</p> | Combinations | Continuous top-k Q/A | Order-aware reasoning |
| | Relevance, Trustability, etc. | Top-k Q/A | Top-k Reasoning |
| | Recency | DSMS/CEP | Stream reasoning |
| | Indexes | Traditional solutions | Scalable reasoning |
| | | No | Yes |
| | |  Semantic Technologies | |

Emanuele Della Valle, Stefan Schlobach, Markus Krötzsch, Alessandro Bozzon, Stefano Ceri, Ian Horrocks: **Order matters! Harnessing a world of orderings for reasoning over massive data.** Semantic Web 4(2): 219-231 (2013)

- Tuesday morning – Workshop: **Stream Reasoning 2016**
 - Keynotes by
 - F. Lecue: Ontology Stream Reasoning for Diagnosis and Predictive Inference
 - J. Z. Pan: The Maze of Deletion in Stream Reasoning
- Tuesday Morning – Tutorial: **Semantic Web meets Internet of Things and Web of Things**
 - A. Gyrard, P. Patel, S. K. Datta, M. I. Ali
- Wednesday evening – **Poster and Demos**
 - M. Watanabe et al.: *Working Process Quantification in Factory Using Wearable Sensor Device and Ontology-based Stream Data Processing*
 - T. Morita et al.: *Implementing Customer Reception Service in Robot Cafe using Stream Reasoning and ROS based on PRINTEPS*
 - X. Ren et al.: *Apache Spark and Apache Kafka at the rescue of distributed RDF Stream Processing engines*
 - R. Taelman et al.: *Querying Dynamic Datasources with Continuously Mapped Sensor Data*
 - R. Keskisarkka: *Representing RDF Stream Processing Queries in RSP-SPIN*

- Wednesday afternoon – **Main Conference Papers**
 - L. M. Daniele et al: Interoperability for Smart Appliances in the IOT World

- Thursday afternoon – **Main Conference Papers**
 - S. Gao et al: Planning Ahead: *Stream-Driven Linked-Data Access under Update-Budget Constraint*
 - E. Siow et al: *SPARQL-to-SQL on Internet of Things Databases and Streams*
 - D. Le-Phuoc: *Operator-aware approach for boosting performance in RDF stream processing*
 - A. Mauri, et al: *TripleWave: Spreading RDF Streams on the Web*
 - E. Kharlamov et al: *Towards Analytics Aware Ontology Based Access to Static Streaming Data*

Please fill the questionnaire at:
<http://tinyurl.com/rspquestionnaire>