Conclusions

Daniele Dell’Aglio

dellaglio@ifi.uzh.ch  http://dellaglio.org  @dandellaglio
Share, Remix, Reuse — Legally

- 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/
Agenda

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

- Proposed approach

DL is just an option. ASP and Bayesian nets can also fit here

-- E. Della Valle, 2015

Research Challenges

- 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
Achievements and open issues
Relation with DSMSs and CEPs

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

<table>
<thead>
<tr>
<th>DB → Semantic Web</th>
<th>DSMS/CEP → Semantic Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational data → RDF</td>
<td>Data streams → RDF Streams</td>
</tr>
<tr>
<td>SQL → SPARQL</td>
<td>CQL/EPL/… → C-SPARQL/EP-SPARQL/…</td>
</tr>
<tr>
<td>Schema → OWL</td>
<td>Schema → OWL</td>
</tr>
</tbody>
</table>

- 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
Achievements and open issues

Data types for semantic streams - Achievements

- **RDF streams introduced** as new data type in the Semantic Web and Linked Data research
Achievements and open issues
Data types for semantic streams - Issues

- **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**

<table>
<thead>
<tr>
<th>System</th>
<th>Data item</th>
<th>Time model</th>
<th># of timestamps</th>
</tr>
</thead>
<tbody>
<tr>
<td>INSTANS</td>
<td>triple</td>
<td>Implicit</td>
<td>0</td>
</tr>
<tr>
<td>C-SPARQL</td>
<td>triple</td>
<td>Point in time</td>
<td>1</td>
</tr>
<tr>
<td>SPARQL_stream</td>
<td>triple</td>
<td>Point in time</td>
<td>1</td>
</tr>
<tr>
<td>CQELS</td>
<td>triple</td>
<td>Point in time</td>
<td>1</td>
</tr>
<tr>
<td>Sparkwave</td>
<td>triple</td>
<td>Point in time</td>
<td>1</td>
</tr>
<tr>
<td>Streaming Linked Data</td>
<td>RDF graph</td>
<td>Point in time</td>
<td>1</td>
</tr>
<tr>
<td>ETALIS</td>
<td>triple</td>
<td>Interval</td>
<td>2</td>
</tr>
</tbody>
</table>

- **Harmonization** of current proposals ongoing at the W3C RSP community group
Achievements and open issues

Query languages for semantic streams - Achievements

- **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
Achievements and open issues
Query languages for semantic streams - Issues

- Comparison between existing approaches

<table>
<thead>
<tr>
<th>System</th>
<th>S2R</th>
<th>R2R</th>
<th>Time-aware</th>
<th>R2S</th>
</tr>
</thead>
<tbody>
<tr>
<td>INSTANS</td>
<td>Based on time events</td>
<td>SPARQL update</td>
<td>Based on time events</td>
<td>Ins only</td>
</tr>
<tr>
<td>C-SPARQL Engine</td>
<td>Logical and triple-based</td>
<td>SPARQL 1.1 query</td>
<td>timestamp function</td>
<td>Batch only</td>
</tr>
<tr>
<td>SPARQL\textsubscript{stream}</td>
<td>Logical and triple-based</td>
<td>SPARQL 1.1 query</td>
<td>no</td>
<td>Ins, batch, del</td>
</tr>
<tr>
<td>CQELS</td>
<td>Logical and triple-based</td>
<td>SPARQL 1.1 query</td>
<td>no</td>
<td>Ins only</td>
</tr>
<tr>
<td>Sparkwave</td>
<td>Logical</td>
<td>SPARQL 1.0</td>
<td>no</td>
<td>Ins only</td>
</tr>
<tr>
<td>Streaming Linked Data</td>
<td>Logical and graph-based</td>
<td>SPARQL 1.1</td>
<td>no</td>
<td>Batch only</td>
</tr>
<tr>
<td>ETALIS</td>
<td>no</td>
<td>SPARQL 1.0</td>
<td>SEQ, PAR, AND, OR, DURING, STARTS, MEETS, FINISHES</td>
<td>Ins only</td>
</tr>
</tbody>
</table>

- Harmonization of current proposals ongoing at the W3C RSP community group
Achievements and open issues

Query languages for semantic streams - issues

- The existing engines
  - adopts different architectural choices and it is still unclear when each choice is best
    - C-SPARQL, ETALIS, SPARQL\textsubscript{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
Achievements and open issues

Reasoning on Streams - Achievements

- **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
Achievements and open issues
Reasoning on Streams - Issues

- **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
Achievements and open issues
Dealing with incomplete & noisy data

- Data streams are incomplete and noisy!

- Achievements
  - **probabilistic ASP** (see morning slides)
  - Combining **deductive and inductive Stream Reasoning**

- Issues
  - More research required!

Achievements and open issues
Distributed and parallel processing

- Data streams are **parallel and distributed** in nature!

- **Achievements**
  - Active Field of research
    - 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 and open issues
Engineering Stream Reasoning Applications

- **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
Achievements and open issues
Wrap-up

- 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

http://streamreasoning.org/events/rsp2016
What's next? order matters!

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

<table>
<thead>
<tr>
<th>Combinations</th>
<th>Order-aware reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous top-k Q/A</td>
<td>Top-k Reasoning</td>
</tr>
<tr>
<td>Top-k Q/A</td>
<td>Stream reasoning</td>
</tr>
<tr>
<td>DSMS/CEP</td>
<td>Scalable reasoning</td>
</tr>
</tbody>
</table>

Types of orders:
- Recency
- Relevance, Trustability, etc.
- Indexes

No | Yes

Semantic Technologies


http://streamreasoning.org/events/rsp2016
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**

Wednesday evening – **Poster and Demos**
- 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*
More on RSP at ISWC 2016

- **Wednesday afternoon – Main Conference Papers**
  - L. M. Daniele et al: Interoperatibility 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*
  - E. Kharlamov et al: *Towards Analytics Aware Ontology Based Access to Static Streaming Data*
Please fill the questionnaire at:

http://tinyurl.com/rspquestionnaire