Stream Reasoning For Linked Data

M. Balduini, J-P Calbimonte, O. Corcho,
D. Dell'Aglio, E. Della Valle, and J.Z. Pan

http://streamreasoning.org/sr4ld2013

Stream Reasoning introduction

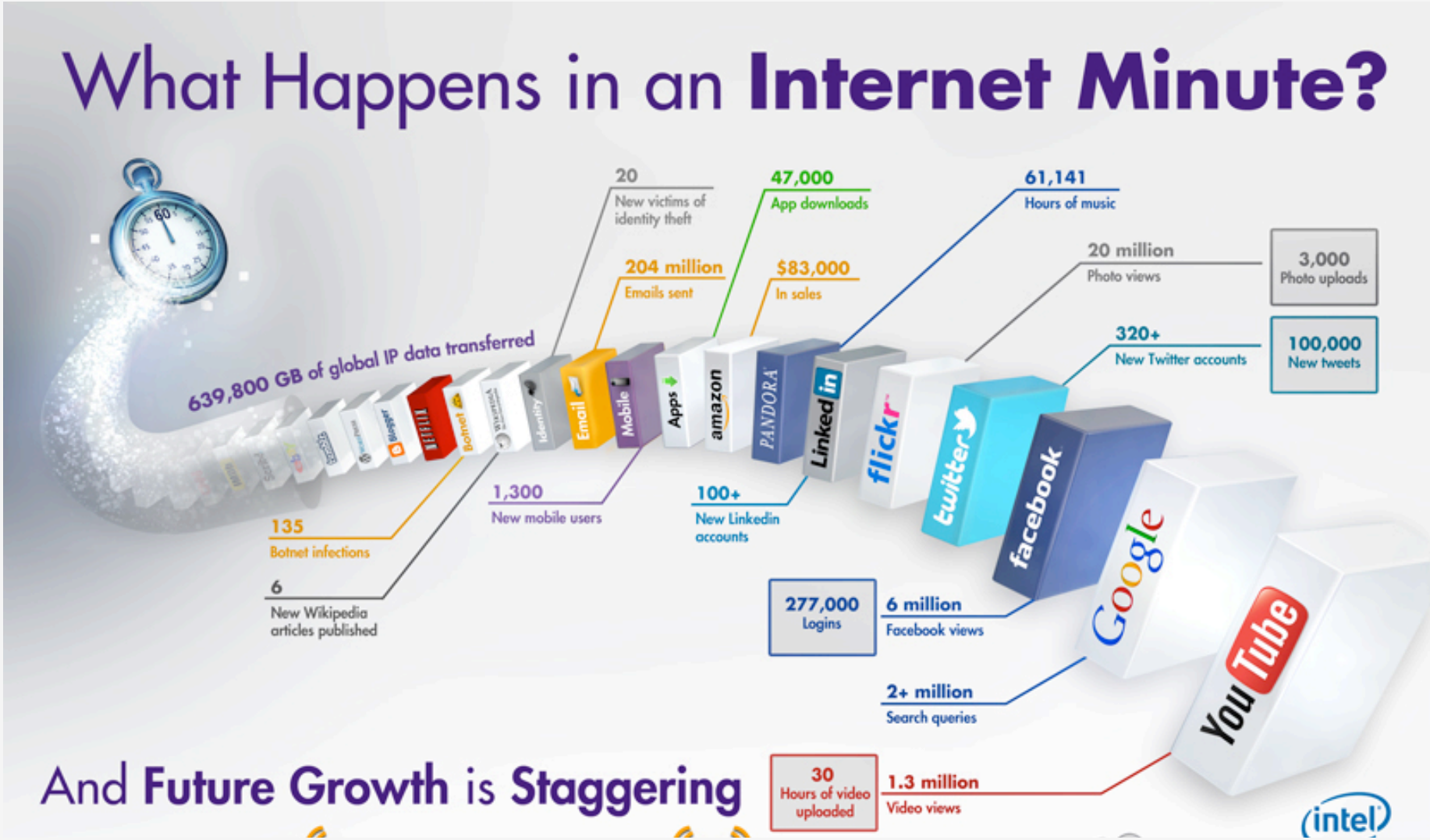
Emanuele Della Valle
emanuele.dellavalle@polimi.it
http://emanueledellavalle.org

Share, Remix, Reuse — Legally

Agenda

- It's a streaming world
- Continuous semantics
- Data Stream Management Systems and Complex Event Processors
- Stream Reasoning
- Research Challenges
- Approaches
- Structure of the tutorial
- More on Stream Reasoning at ISWC 2013

It ‘s a streaming World! 1/3



[source http://y2socialcomputing.files.wordpress.com/2012/06/social-media-visual-last-blog-post-what-happens-in-an-internet-minute-infographic.jpg]

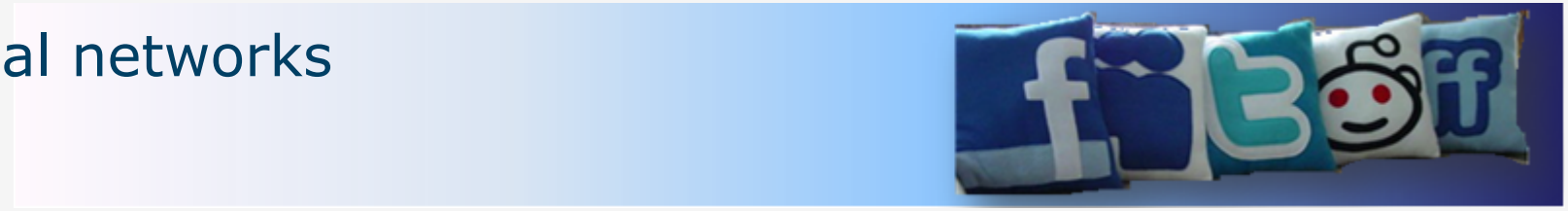
It ‘s a streaming World! 2/3

- Oil operations

- Traffic

- Financial markets

- Social networks



- Generate data streams!

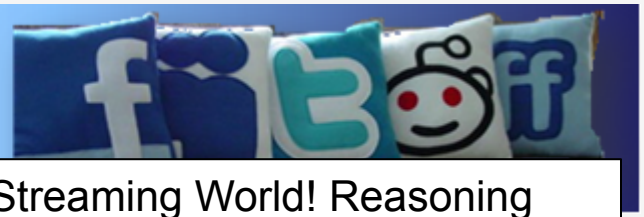
It ‘s a streaming World! 3/3

- … want to analyse data streams in real time and to receive answers in push mode
- In a well in progress to drown, how long time do I have given its historical behavior?

- Is public transportation where the people are?

- Can we detect any intra-day correlation clusters among stock exchanges?

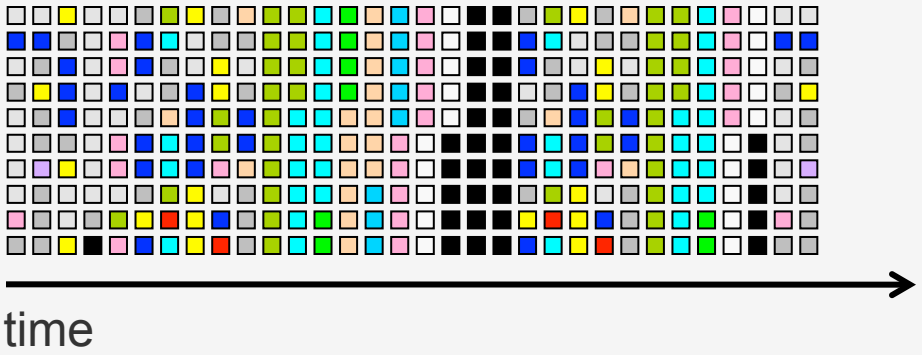
- Who is driving the discussion about the top 10 emerging topics ?



E. Della Valle, S. Ceri, F. van Harmelen, D. Fensel It's a Streaming World! Reasoning upon Rapidly Changing Information. IEEE Intelligent Systems 24(6): 83-89 (2009)

What are data streams anyway?

- Formally:
  - Data streams are unbounded sequences of time-varying data elements



- Less formally:
  - an (almost) “continuous” flow of information
- Assumption
  - recent information is more relevant as it describes the current state of a dynamic system

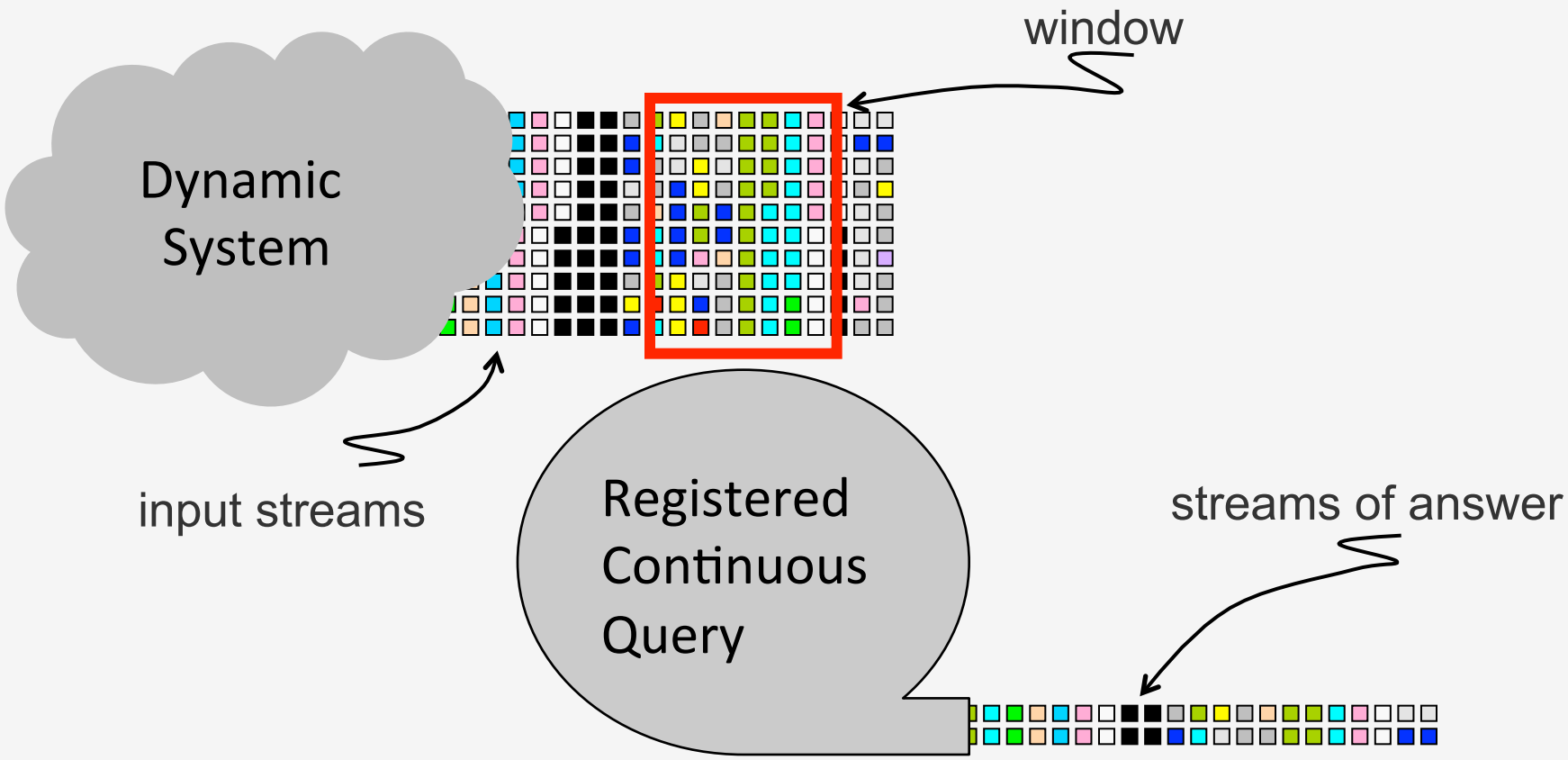
The continuous nature of streams

- The nature of streams requires a paradigmatic change*
  - **from persistent data**
    - to be stored and queried on demand
    - a.k.a. one time semantics
  - **to transient data**
    - to be consumed on the fly by continuous queries
    - a.k.a. continuous semantics

- * This paradigmatic change first arose in DB community [Henzinger98]

Continuous Semantics

- Continuous queries registered over streams that, in most of the cases, are observed trough windows



Example

- Input
  - Smoke and Temperature sensors in many areas
- Query
  - Alert me when there is a fire, i.e. smoke and temp>50
- DSMS formulation
  - Stream the areas where smoke is detected over two windows open on smoke and temperature streams

```
Select IStream(Smoke.area)
From Smoke[Rows 30 Slide 10], Temp[Rows 50 Slide 5]
Where Smoke.area = Temp.area AND Temp.value > 50
```

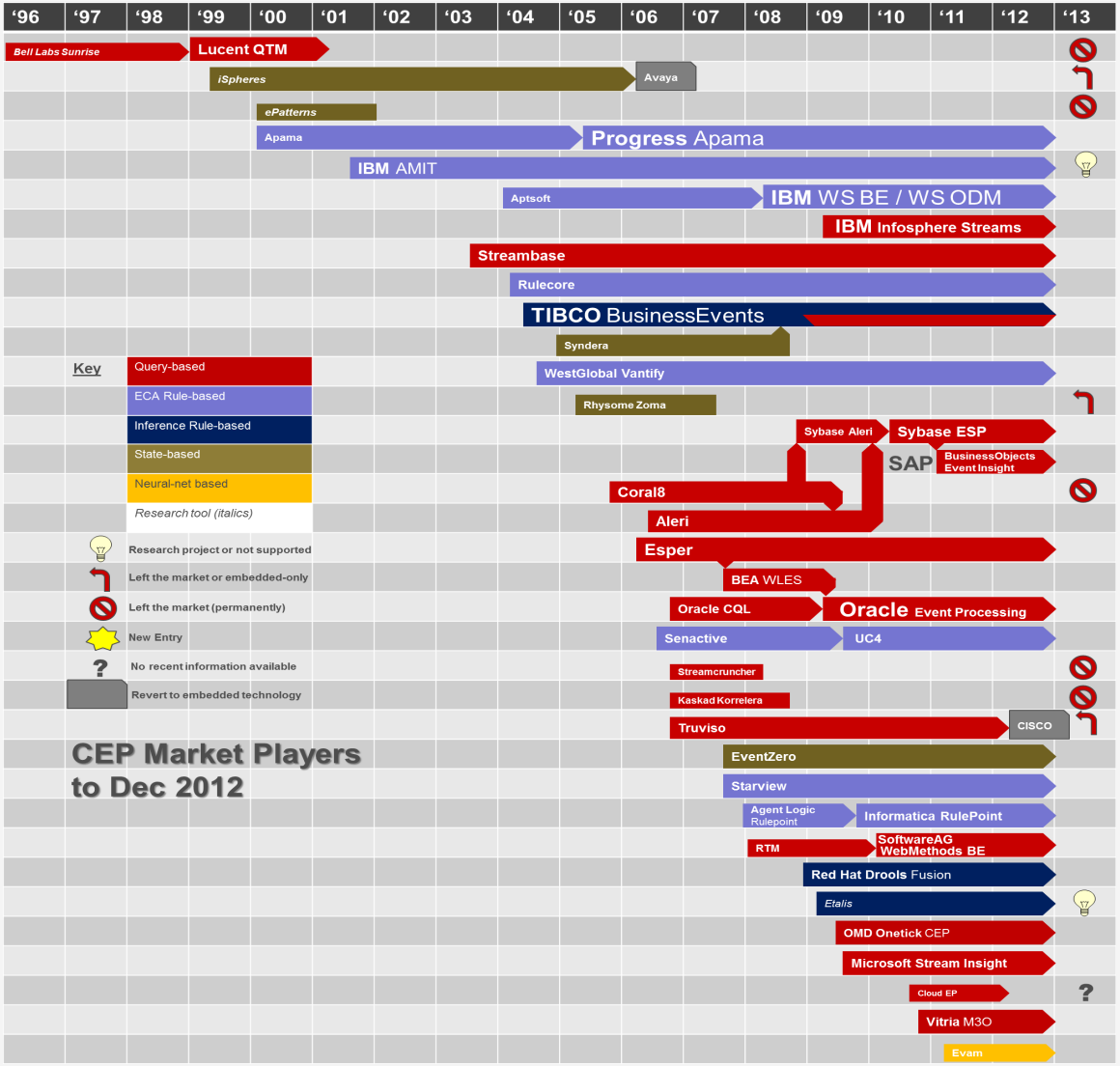
- CEP formulation
  - Rise a fire event in an area when smoke and high temperature events are received within 1 minute

```
define  Fire(area: string, measuredTemp: double)
from    Smoke(area=$a) and
        each Temp(area=$a and val>50) within 1min.
where   area=Smoke.area and measuredTemp=Temp.value
```

DSMS/CEP State of the Art

- Gianpaolo Cugola, Alessandro Margara: Processing flows of information: From data stream to complex event processing. ACM Comput. Surv. 44(3): 15 (2012)
- Content
  - Type of models compared
    - Functional and processing
    - Deployment and interactions
    - Data, Time, and Rule
    - Language
  - # of systems surveyed:
    - Academic: 24
    - Industrial: 9
    - Total: 33
  - To learn more:
    - http://home.dei.polimi.it/margara/papers/survey.pdf

DSMS/CEP Market Players



[source https://ctrlaltcep.files.wordpress.com/2013/01/cepmarket1212.png]

New Requirements → New Challenges

Typical Requirements	Challenge
Processing Streams	Continuous semantics
Large datasets	Scalable processing
Heterogeneous data	Data Integration
Incomplete and noisy data	Uncertainty mng.
Reactivity	Real-time systems
Fine-grained information access	Powerful query languages
Modeling complex application domains	Rich ontology languages

Are DSMS/CEP ready to address them?

Typical Requirements	DSMS/CEP
Processing Streams	✓ Continuous semantics
Large datasets	✓ Scalable processing
Heterogeneous data	✗ Data Integration
Incomplete and noisy data	✓ Uncertainty mng.
Reactivity	✓ Real-time systems
Fine-grained information access	✓ Powerful query languages
Modeling complex application domains	✗ Rich ontology languages

Is Semantic Web/Linked Data ready?

- Data streams can be just another form of Linked Data
- The Semantic Web/Linked Data fields are doing fine
  - RDF, RDF Schema, SPARQL, OWL
  - well understood theory
  - rapid increase in scalability
  - rapid adoption of Linked Data to publish data on the Web
- BUT they (largely) pretends that the world is static or at best a low change rate both in change-volume and change-frequency
  - SPARQL UPDATE
  - time stamps on named graphs
  - ontology versioning
  - belief revision
- They sticks to the traditional one-time semantics

New Requirements → New Challenges

Typical Requirements	Semantic Web
Processing Streams	 Continuous semantics
Large datasets	 Scalable processing
Heterogeneous data	 Data Integration
Incomplete and noisy data	 Uncertainty mng.
Reactivity	 Real-time systems
Fine-grained information access	 Powerful query languages
Modeling complex application domains	 Rich ontology languages

New Requirements call for Stream Reasoning

Typical Requirements	Stream Reasoning
Processing Streams	Continuous semantics
Large datasets	Scalable processing
Heterogeneous data	Data Integration
Incomplete and noisy data	Uncertainty mng.
Reactivity	Real-time systems
Fine-grained information access	Powerful query languages
Modeling complex application domains	Rich ontology languages

Stream Reasoning Definition

- Making sense
  - in real time
  - of multiple, heterogeneous, gigantic and inevitably noisy data streams
  - in order to support the decision process of extremely large numbers of concurrent user

- Note: making sense of streams necessarily requires processing them against rich background knowledge, an unsolved problem in database

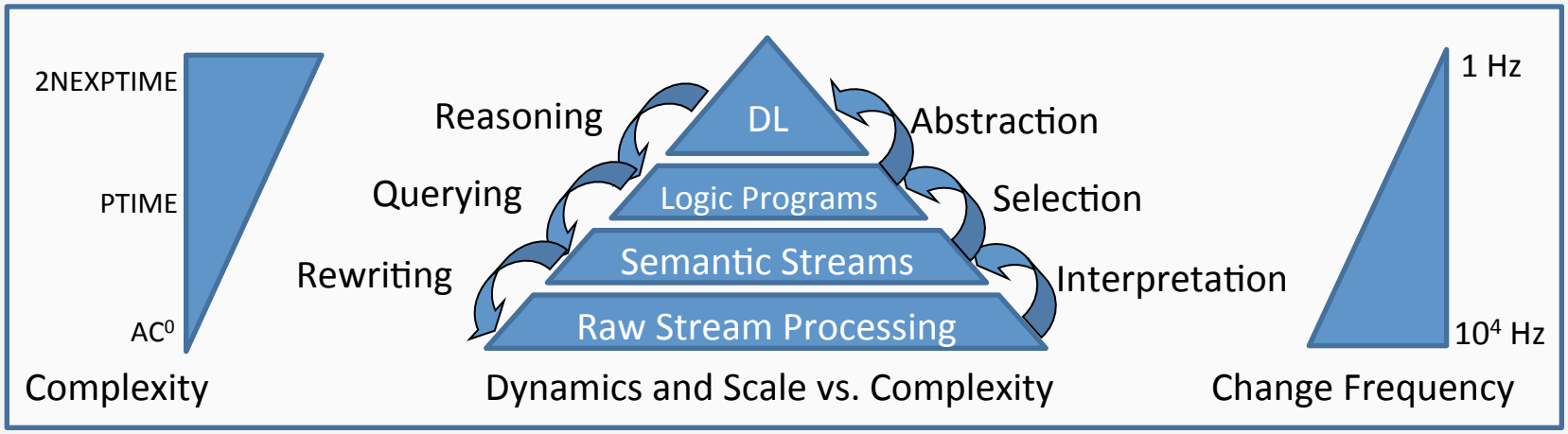
D. Barbieri, D. Braga, S. Ceri, E. Della Valle, Y. Huang, V. Tresp, A.Rettinger, H. Wermser: Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics IEEE Intelligent Systems, 30 Aug. 2010.

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

Stream Reasoning feasibility (intuition)

- Many relevant reasoning methods are not able to deal with high frequency data streams
- However, trade-off exists between the complexity of the reasoning method and the frequency of the data stream the reasoner



Heiner Stuckenschmidt, Stefano Ceri, Emanuele Della Valle, Frank van Harmelen: Towards Expressive Stream Reasoning. Proceedings of the Dagstuhl Seminar on Semantic Aspects of Sensor Networks, 2010.

Approaches (a selection) 1/4

- RDF Stream Processors (ordered by year)
  - **C-SPARQL**
    - Davide Francesco Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle, Michael Grossniklaus: Querying RDF streams with C-SPARQL. SIGMOD Record 39(1): 20-26 (2010)
  - **SPARQL$_{stream}$**
    - Jean-Paul Calbimonte, Óscar Corcho, Alasdair J. G. Gray: Enabling Ontology-Based Access to Streaming Data Sources. International Semantic Web Conference (1) 2010: 96-111
  - **CQELS**
    - Danh Le Phuoc, Minh Dao-Tran, Josiane Xavier Parreira, Manfred Hauswirth: A Native and Adaptive Approach for Unified Processing of Linked Streams and Linked Data. International Semantic Web Conference (1) 2011: 370-388
  - It continues in next slide …

Approaches (a selection) 2/4

- … it continues from previous slide
- **INSTANS**
  - Rinne, M., Nuutila, E., Törma, S.: INSTANS: High-Performance Event Processing with Standard RDF and SPARQL. Poster in ISWC2012.
- **Streaming Linked Data**
  - Marco Balduini, Emanuele Della Valle, Daniele Dell'Aglio, Mikalai Tsytsarau, Themis Palpanas, Cristian Confalonieri: Social listening of City Scale Events using the Streaming Linked Data Framework. ISWC 2013

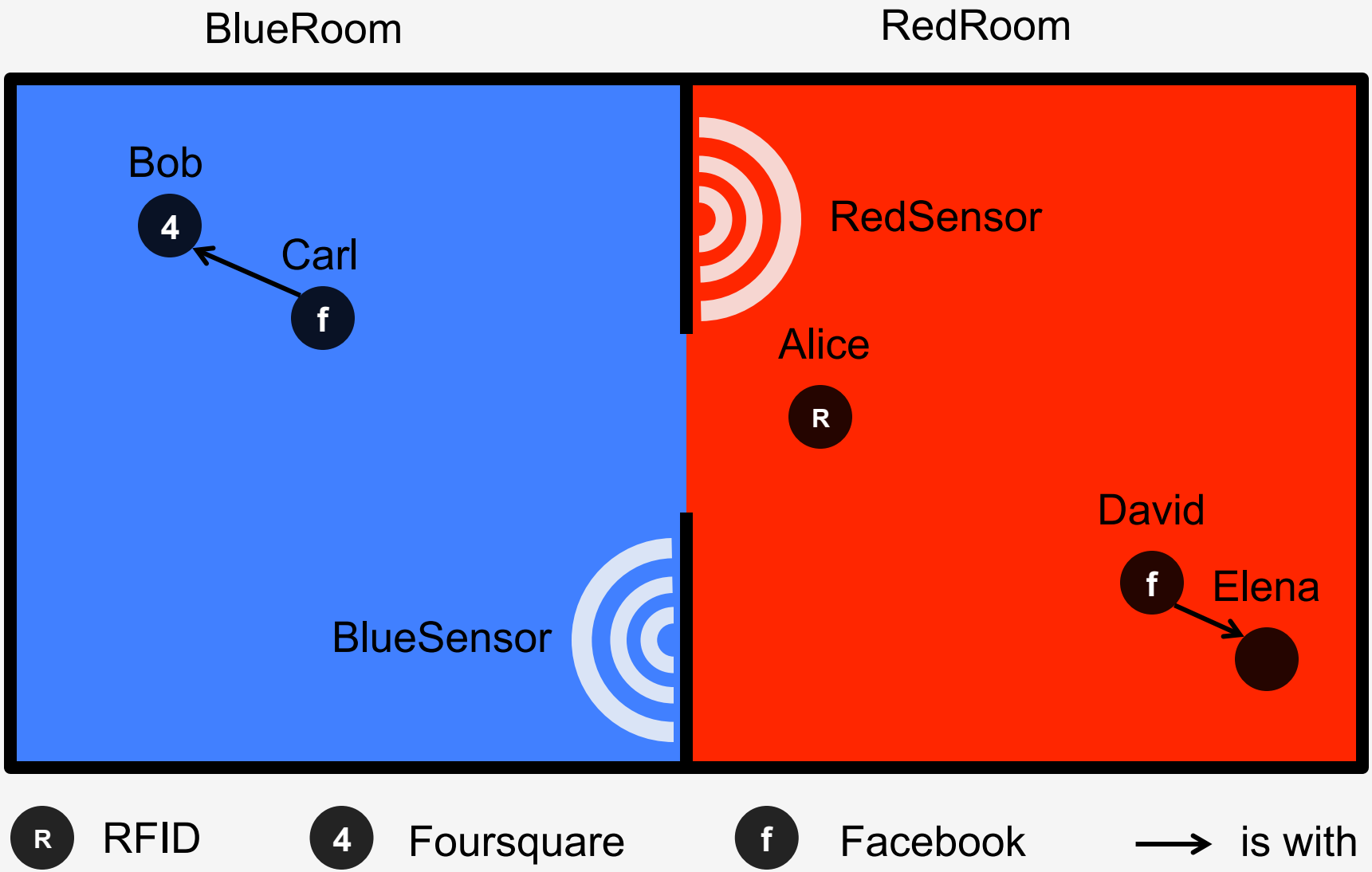
Approaches (a selection) 3/4

- Stream Reasoners (ordered by year)
  - **Streaming Knowledge Bases**
    - Walavalkar, O., Joshi, A., Finin, T., Yesha, Y., 2008. Streaming knowl- edge bases. In: In International Workshop on Scalable Semantic Web Knowledge Base Systems
  - **IMaRS**
    - Davide Francesco Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle, Michael Grossniklaus: Incremental Reasoning on Streams and Rich Background Knowledge. ESWC (1) 2010: 1-15
  - **TrOWL**
    - Yuan Ren, Jeff Z. Pan: Optimising ontology stream reasoning with truth maintenance system. CIKM 2011: 831-836
  - **ETALIS** (EP-SPARQL)
    - Darko Anicic, Paul Fodor, Sebastian Rudolph, Nenad Stojanovic: EP-SPARQL: a unified language for event processing and stream reasoning. WWW 2011: 635-644
  - It continues in next slide …

Approaches (a selection) 4/4

- … continues from previous slide
- **Sparkwave**
  - Srdjan Komazec, Davide Cerri, Dieter Fensel: Sparkwave: continuous schema-enhanced pattern matching over RDF data streams. DEBS 2012: 58-68
- **SR-Based on Answer Set Programming**
  - Martin Gebser, Torsten Grote, Roland Kaminski, Philipp Obermeier, Orkunt Sabuncu, Torsten Schaub: Stream Reasoning with Answer Set Programming: Preliminary Report. KR 2012

Running Example



Running Example

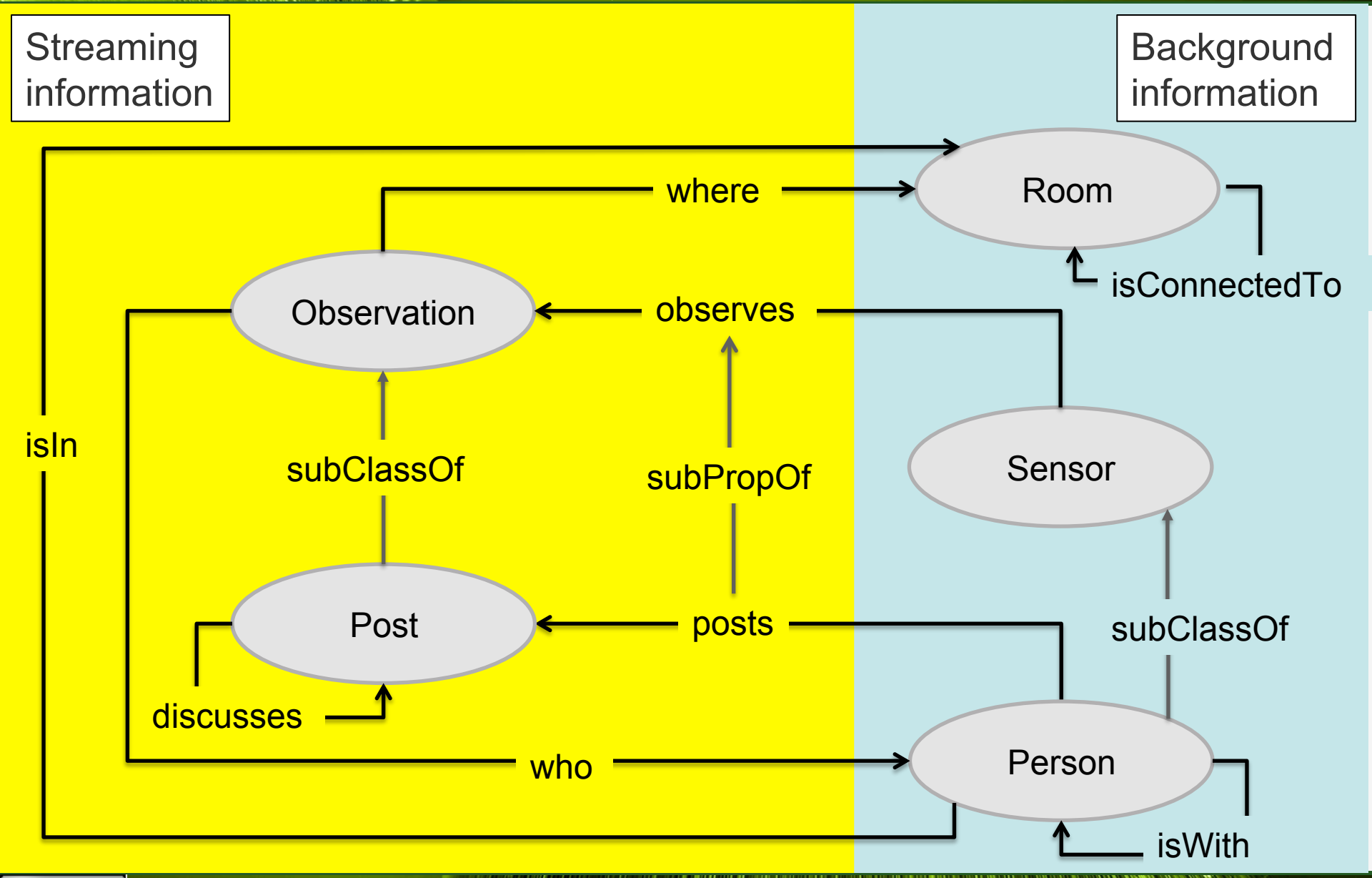
- Four ways to learn who is where

Sensor	Room	Person	Time-stamp
RedSensor	RedRoom	Alice	T_1
…	…	…	…

Person	ChecksIn	Time-stamp
Bob	BlueRoom	T_2
…	…	…

Person	IsIn	With	Time-stamp
Carl	null	Bob	T_2
David	RedRoom	Elena	T_3
…	…	…	…

Running Example – Data Model



Running Example – Data Model (formally)

- Details about hands-on ontology
  - *isConnectedTo* is a symmetric property
  - *discusses* is a transitive property
  - *isWith* is a composition of *posts* and *who*
  - *isIn* is either a composition of *posts* and *where*
    or a composition of *isWith* and *isIn*

- Available online
  - http://www.streamreasoning.org/ontologies/sr4ld2013-onto.rdf

Structure of the tutorial

- 9.00 - 10.30
  - Stream Reasoning introduction (30 min)
  - RDF stream processing models (45 min)
  - Naive reasoning on RDF streams (25 min)
- 11.00 - 12.45
  - C-SPARQL: A Continuous Extension of SPARQL (20 min)
  - $SPARQL_{stream}$: Ontology-based streaming data access (40 min)
  - Hands on session (45 min)
- 13:45 - 15.30
  - Approximate Reasoning and Approximate Stream Reasoning for OWL2-DL (70m)
  - Hands on session (20 min)
- 16:00 - 17.30
  - IMaRS: Incremental Materialization for RDF Streams (30m)
  - Other Stream Reasoning approaches (30 min)
  - Wrap-up and conclusions (30 min)

Have fun! Any question?

Water, water, every where,

Nor any drop to drink.

-- The Rime of the Ancient Mariner
Samuel Taylor Coleridge, 1798

Streams, streams everywhere

nor any actionable fact to use

-- Emanuele and Daniele :-P

Stream Reasoning For Linked Data

M. Balduini, J-P Calbimonte, O. Corcho,
D. Dell'Aglio, E. Della Valle, and J.Z. Pan

http://streamreasoning.org/sr4ld2013

Stream Reasoning introduction

Emanuele Della Valle
emanuele.dellavalle@polimi.it
http://emanueledellavalle.org