Research Plan for 2014

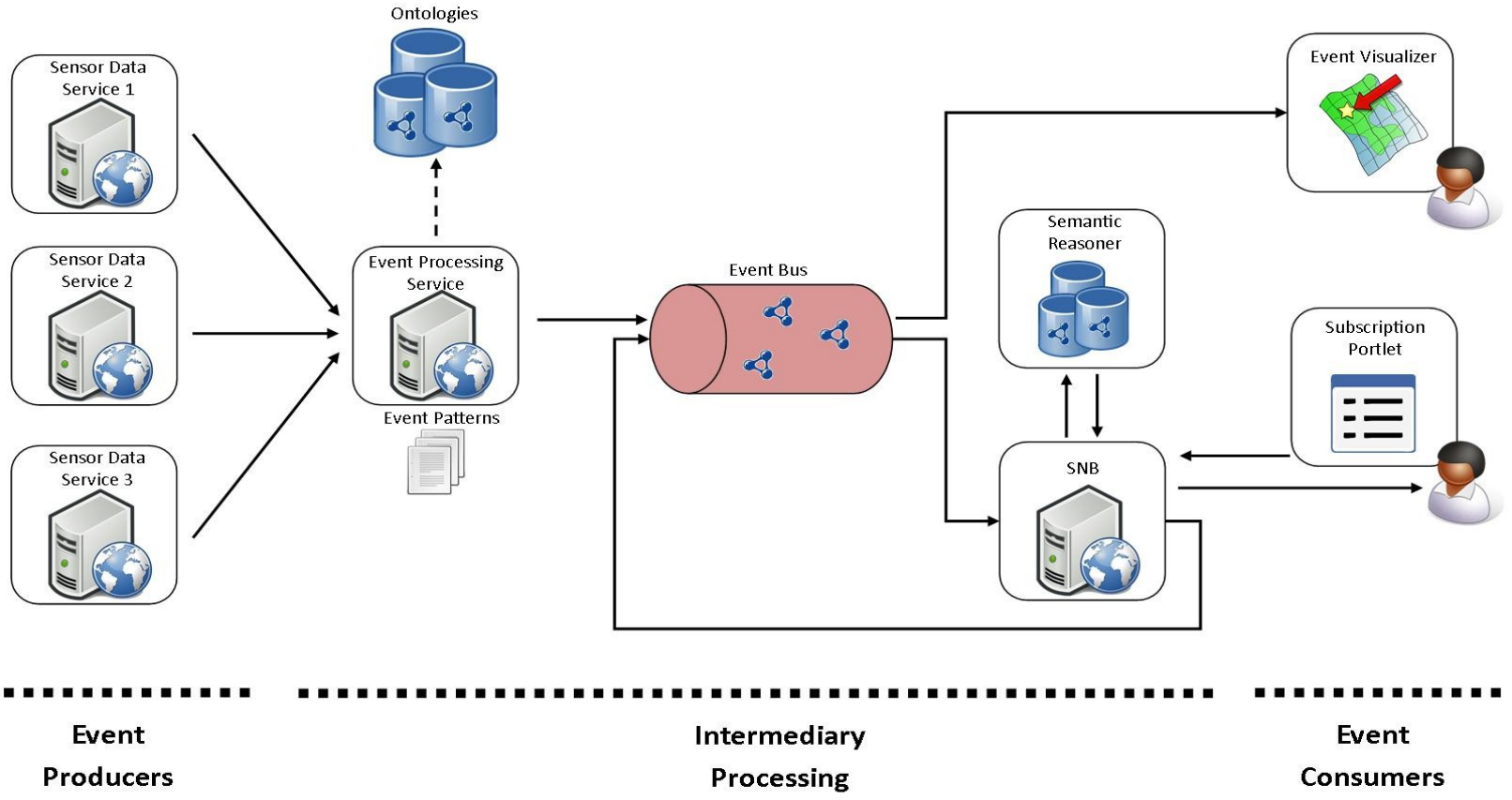
Alejandro Llaves
Ontology Engineering Group
March 13th

Outline

- A bit of background
- Research goals
- What have I done so far?
- What am I currently doing?
- Next steps

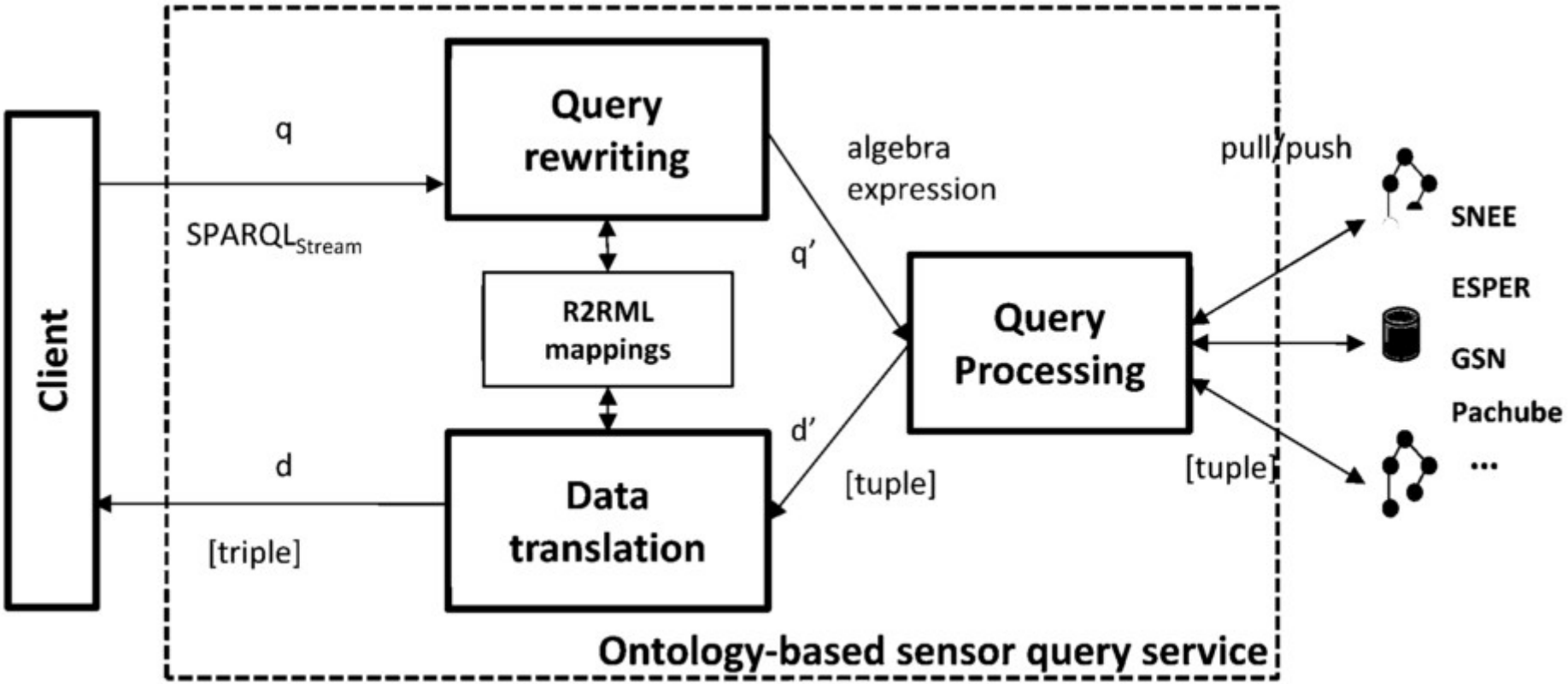
Background: my PhD thesis

- “Integration of Sensor Data by means of an Event Abstraction Layer” http://miami.uni-muenster.de/Record/9700aedd-4f44-4d1a-9884-e06b31703b27
- Keywords: (geo)sensor data streams; complex event processing; data integration; Semantic Sensor Network ontology



Background: morph-streams

...SPARQL$_{Stream}$ and similar stuff are being discussed at the W3C RDF Stream Processing Group - http://www.w3.org/community/rsp/



Background: PlanetData

- PlanetData: a *scalable* RDF streaming engine
- What is Scalability? (Hill 1990)

  "Scalability is a frequently-claimed attribute of multiprocessor systems. While the basic notion is intuitive, scalability has no generally-accepted definition. For this reason, current use of the term adds more to marketing potential than technical insight.

  In this paper, I first examine formal definitions of scalability, ***but I fail to find a useful, rigorous definition of it****. I then question whether scalability is useful and conclude by challenging the technical community to either (1) rigorously define scalability or (2) stop using it to describe systems."*

- Define scalability based on certain dimensions: data size, query complexity, number of concurrent queries, input rate...

Research goals

September 2014

- A RDF streaming engine that scales to higher input data rates and/or more complex queries and/or more concurrent queries
- Deliverable describing such engine and the research behind it for PlanetData

December 2014

- Improvements on the engine based on application purpose
- Journal paper

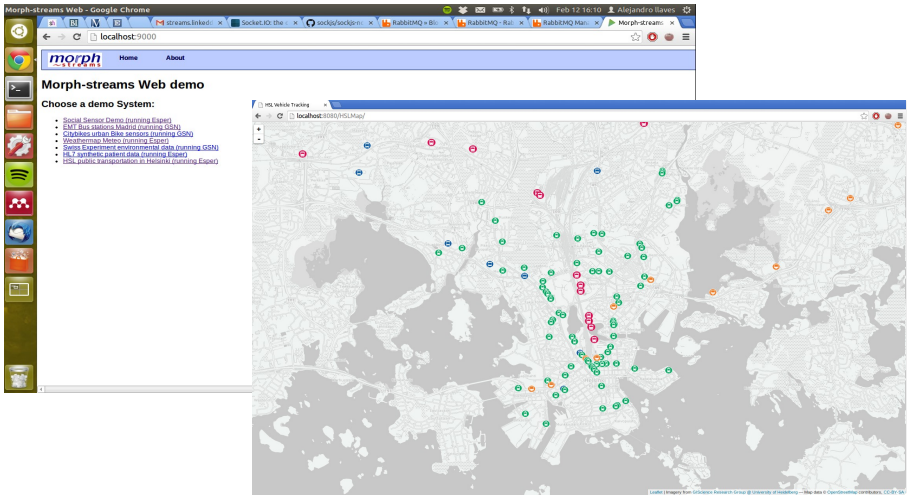
Preliminary hypothesis & RQs

- **PH1:** Given a RDF streaming engine, a SPARQL query, a set of RDF data streams, and finite computing resources, it is possible to define a processing strategy that optimizes the time and resources used to provide a response in near real-time.
- **PH2:** Using an adaptive query processing strategy, a RDF streaming engine offers better performance against varying input data rates, requests, and system conditions.

- **RQ1:** What technologies are suited to optimize query processing over RDF data streams?
- **RQ2:** Which set of query operators are optimizable in the context of Linked Stream Data?
- **Bonus Question:** Are there features of Linked Geospatial Data that make its processing more parallelizable?

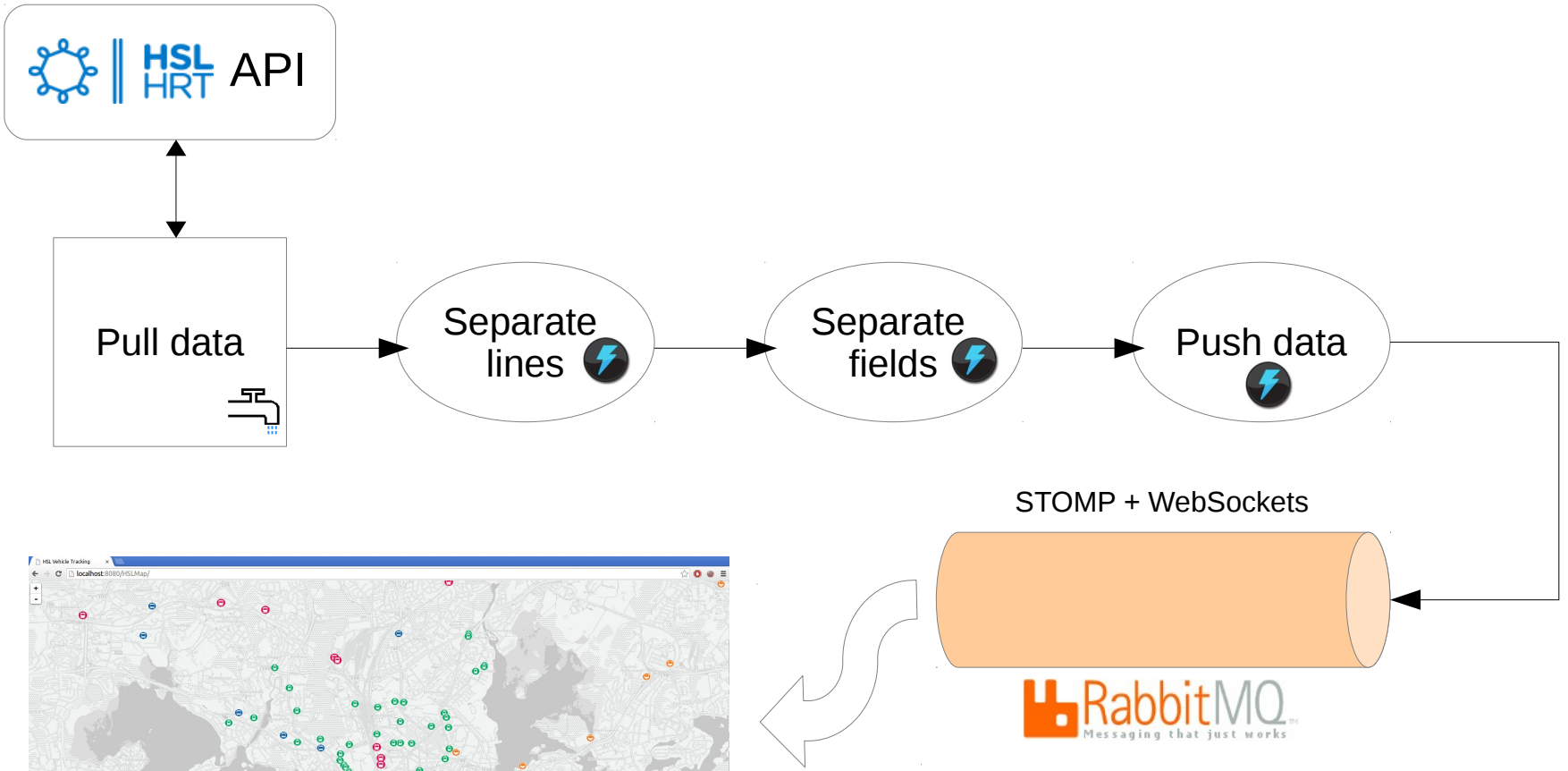
What have I done so far?

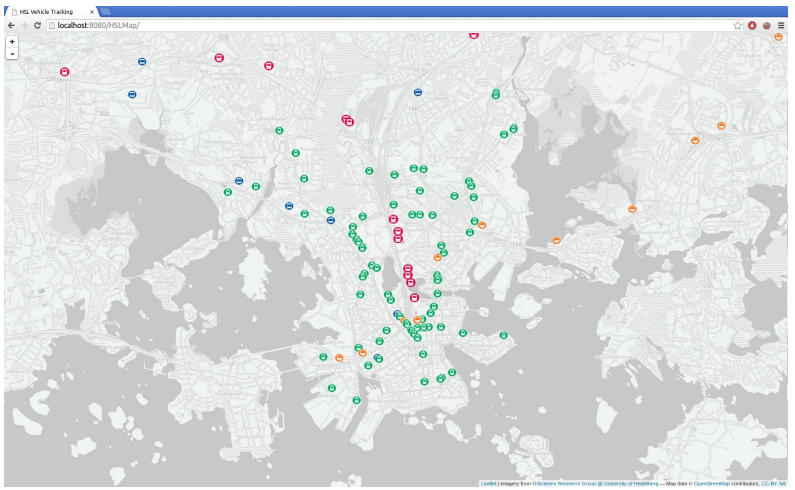
- Stream processing / Big Data technologies survey
- 2 demos
  - Morph-streams and HSL Live
  - Storm and HSL Live



- Linking Geospatial Data short paper & presentation (London, March 5-6th) http://www.w3.org/2014/03/lgd/
- Project proposal: final phase!

Storm demo overview

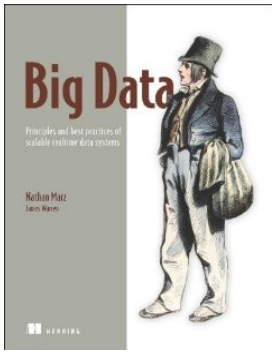




HTML + Javascript

What am I currently doing?

- Adaptive Query Processing (AQP) literature
- Analyzing SPARQL query operators
- Studying Lambda architectures
- Discovering RDF compression (ask Javi)
- Preparing a short paper for AGILE 2014



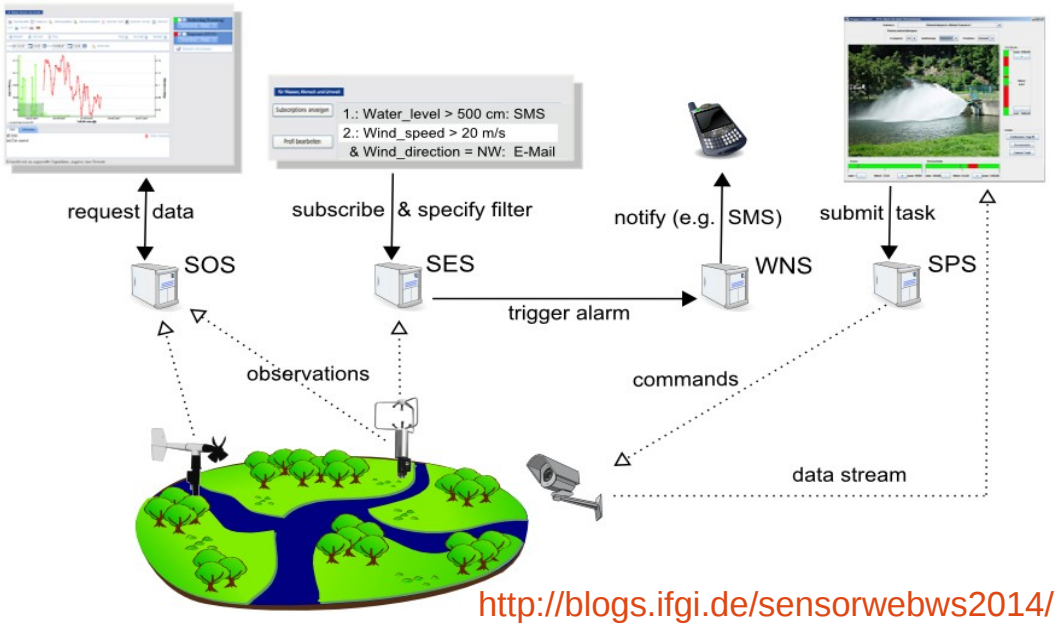
Next steps

- Test CQELS Cloud and analyze the code
- Focus on query operators (JOIN, FILTER, OPTIONAL...)
- Design algorithms that optimize different types of SPARQL queries
- Develop and evaluate a prototype that implements AQP

Some advertising: AGILE workshops

Sensor Web for Environmental Research

- Extended abstracts (1000-3000 w.) describing current projects and experiences with the application of Sensor Web technologies
- June 3rd, Castellón
- Deadline: April 15th



http://blogs.ifgi.de/sensorwebws2014/