Semantic Integration of Streaming Data Sources

Jean-Paul Calbimonte and Oscar Corcho
Ontology Engineering Group. Departamento de Inteligencia Artificial.
Facultad de Informática, Universidad Politécnica de Madrid.
Campus de Montegancedo s/n.
28660 Boadilla del Monte. Madrid. Spain
{jpcalbimonte, ocorcho}@fi.upm.es

Date: 17/12/2009

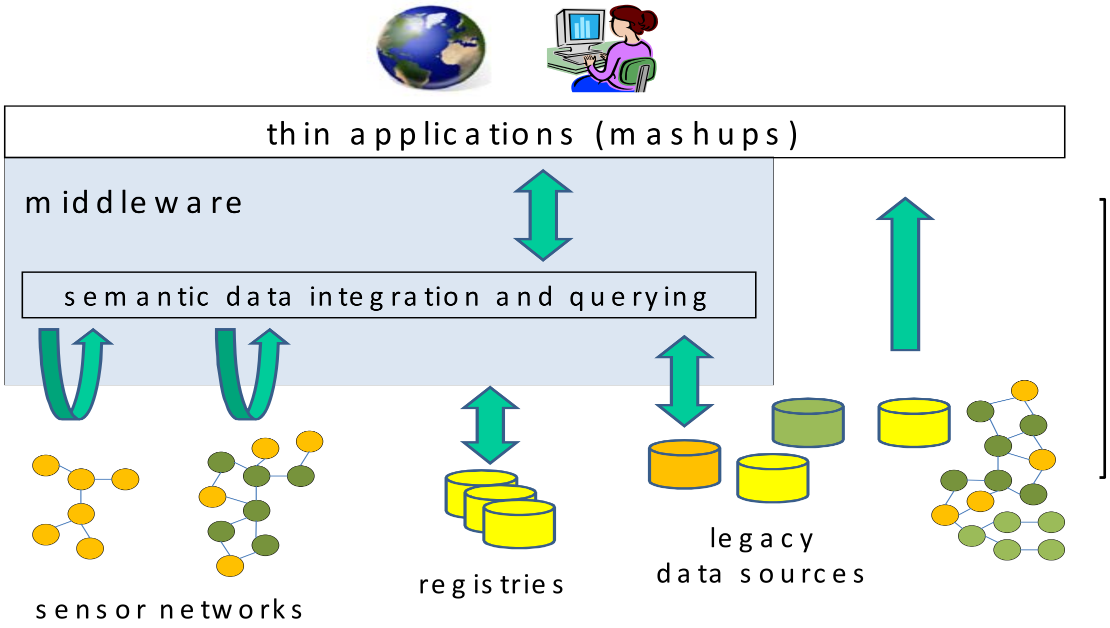
Index

- Introduction & Context: SSG4Env
- Architecure
- Background
- Semantic Integrator
- SPARQL extensions
- Mapping extensions
- Query translation & rewriting
- Distributed query processing
- Ongoing work

SemSorGrid4Env

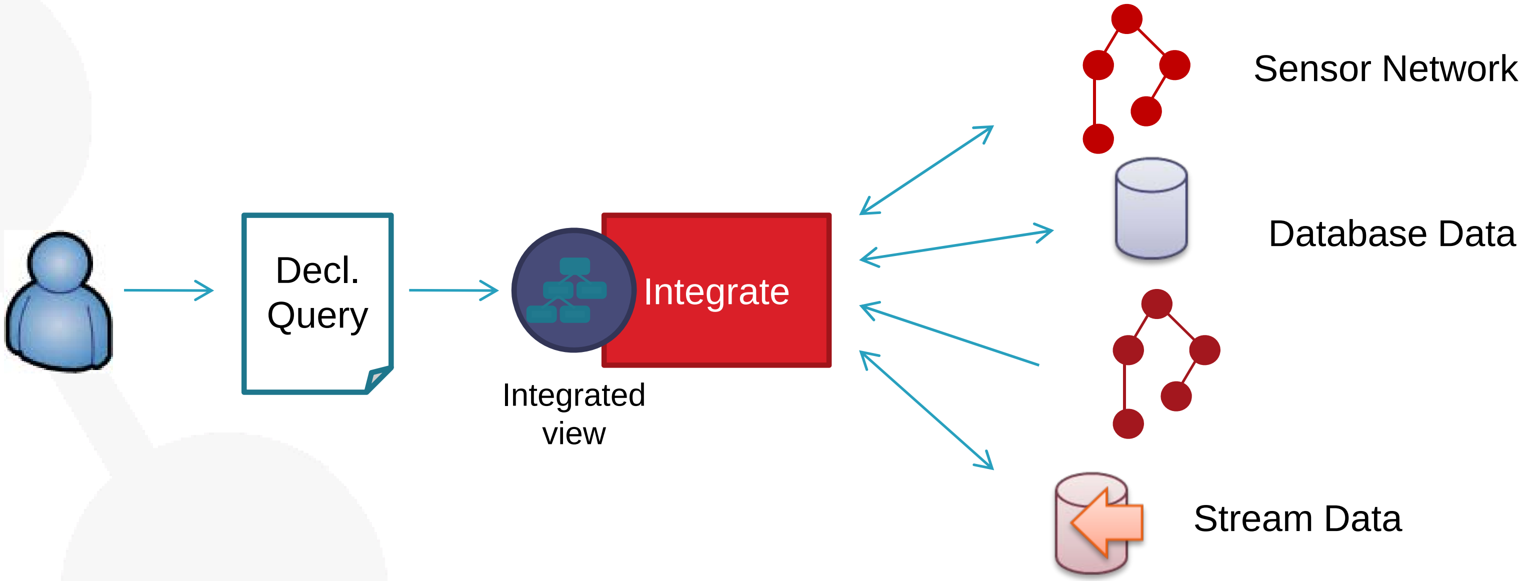
Development of an integrated information space where new sensor networks can be easily discovered and integrated with existing ones and possibly other data sources (e.g., historical databases)

Rapid development of flexible and user-centric decision support systems that use data from multiple autonomous independently deployed sensor networks and other applications.



Integration Requirements

- Based on use-cases
- Integrate stored and streaming data sources
- Integrate sources through unified view
- Pose declarative queries over integrated view



Background

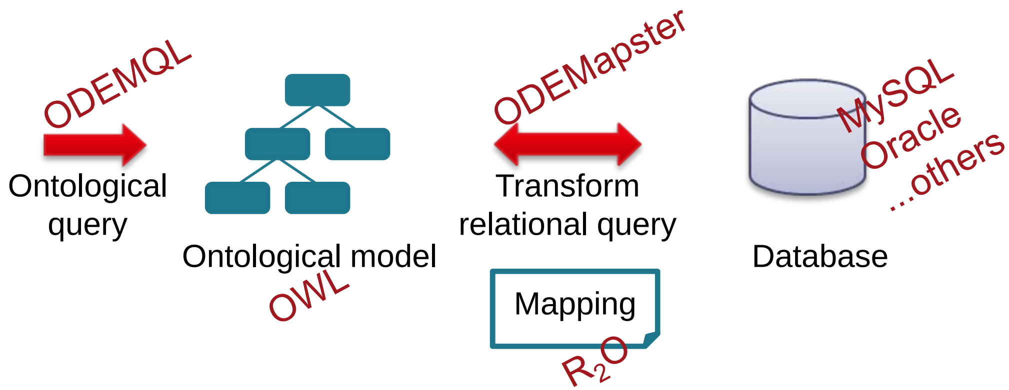
Research Areas

- Ontology-based data access
- Ontology-based data integration
- Streaming data access
- Distributed query processing

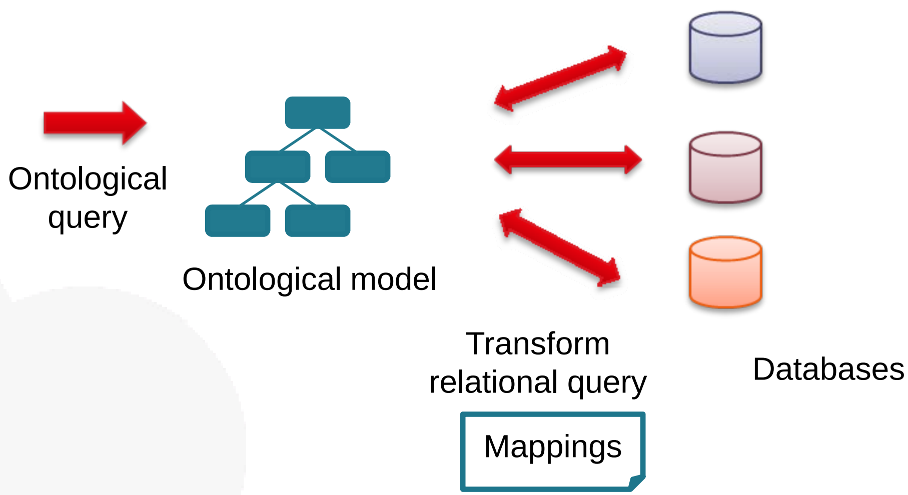
Technologies

- SNEE/SNEEql
- R_2O/ODEMapster
- SPARQL streaming extensions

Ontology-based data access & integration



SquirrelRDF
RDBToOnto
Relational.OWL
SPASQL
Virtuoso
D2RQ
MASTRO
R_2O +
ODEMapster



OBSERVER
SIMS
Carnot
DWQ
PICSEL
MOMIS

Streaming Data Access

Sensor

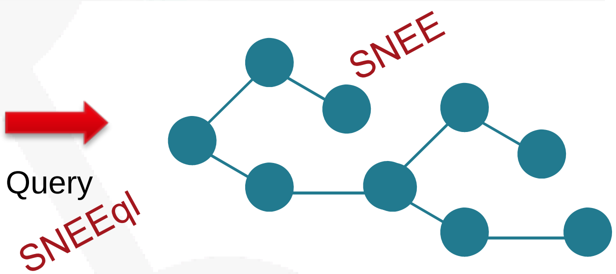
ueries
sed

Streaming
Data

(t9, a1, a2, ... , an)
(t8, a1, a2, ... , an)
(t7, a1, a2, ... , an)
...
...
(t1, a1, a2, ... , an)
...
...

Window
[t7 - t9]

STREAM
Aurora/Borealis
Cougar
TinyDB

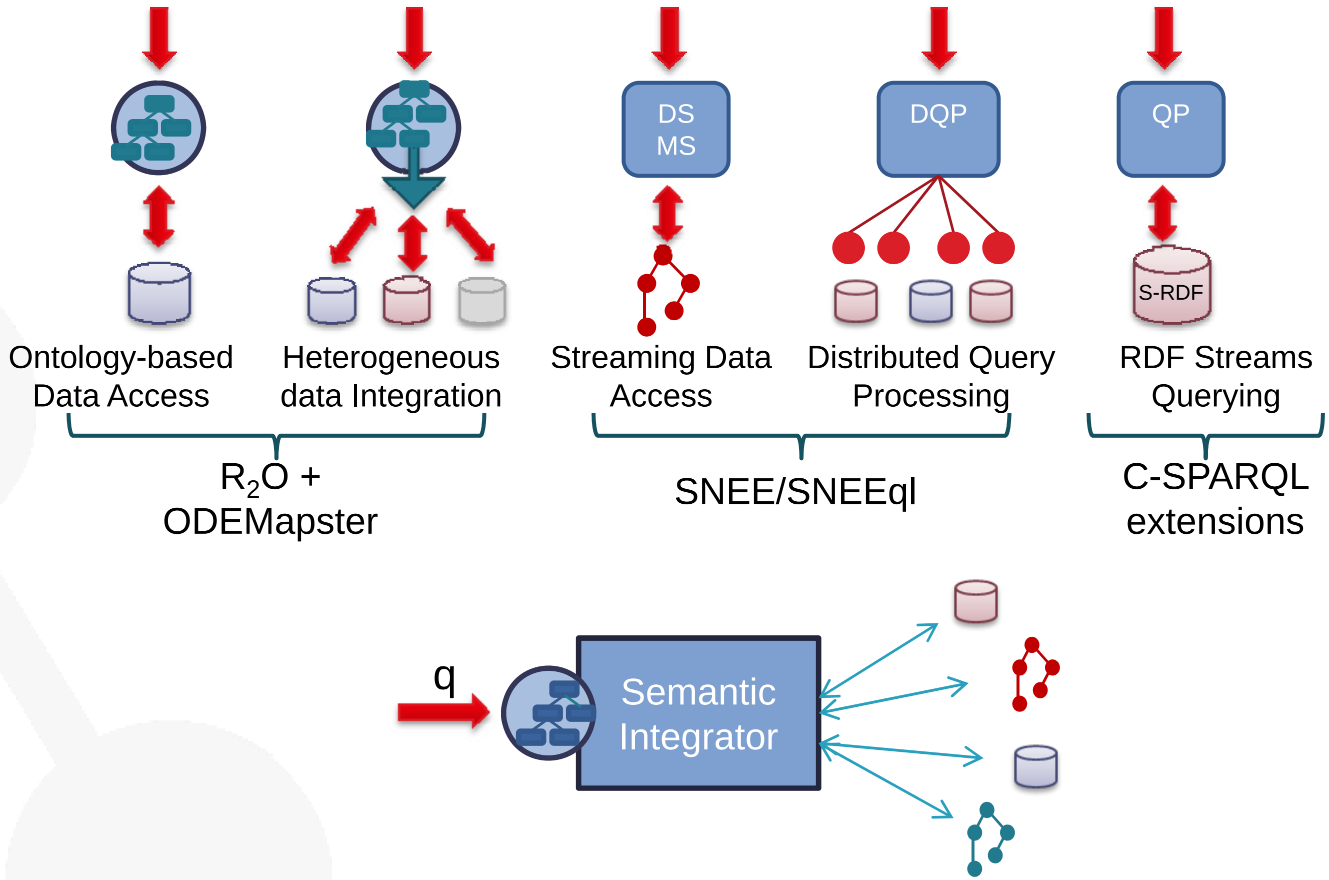


- Sensor Networks characteristics
- Cheap, Nosiy, Unreliable (depends)
- Low computational, power resources, storage
- Distributed query execution
- Routing, Optimization

SPARQL streaming extensions

- SPARQL RDF query language
- Language limitations for streams
  - Windows, time, tuple
  - Data model
  - Aggregates, stream operators
- Streaming SPARQL
- C-SPARQL

Background: Approaches & Technologies

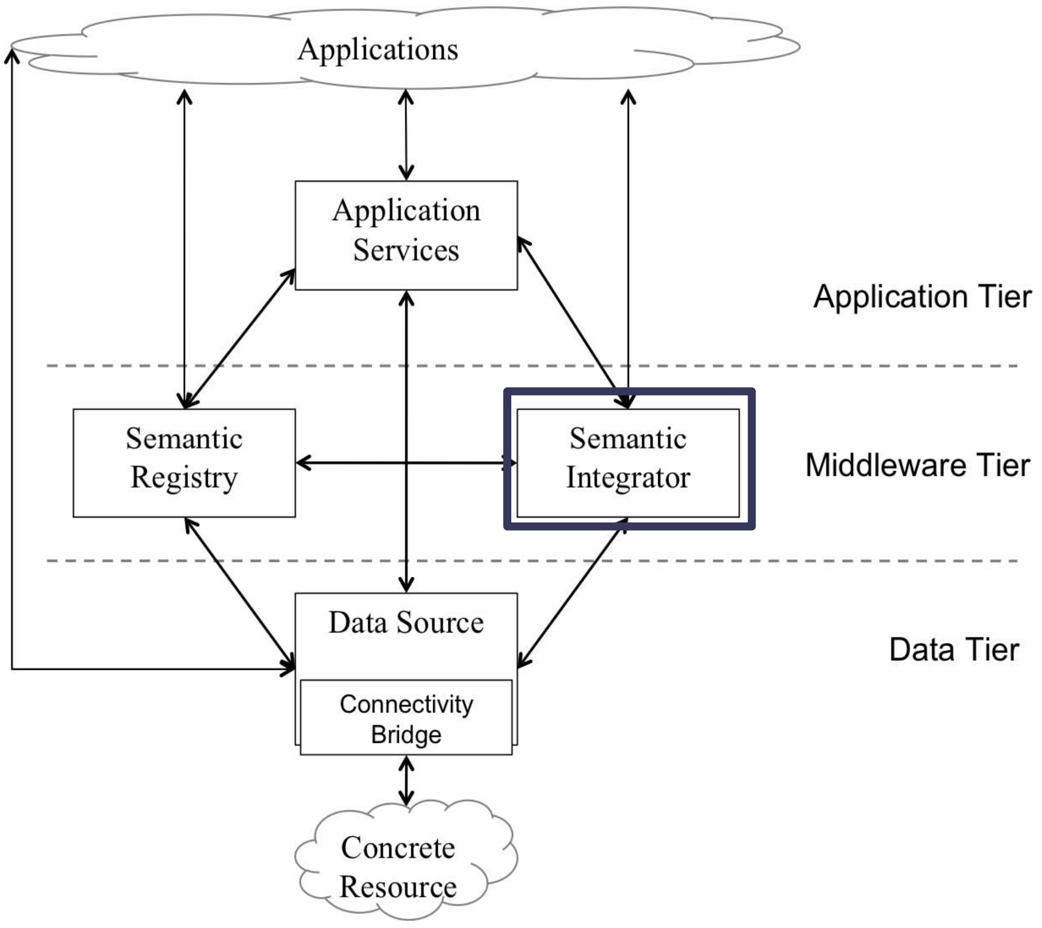


Architecture

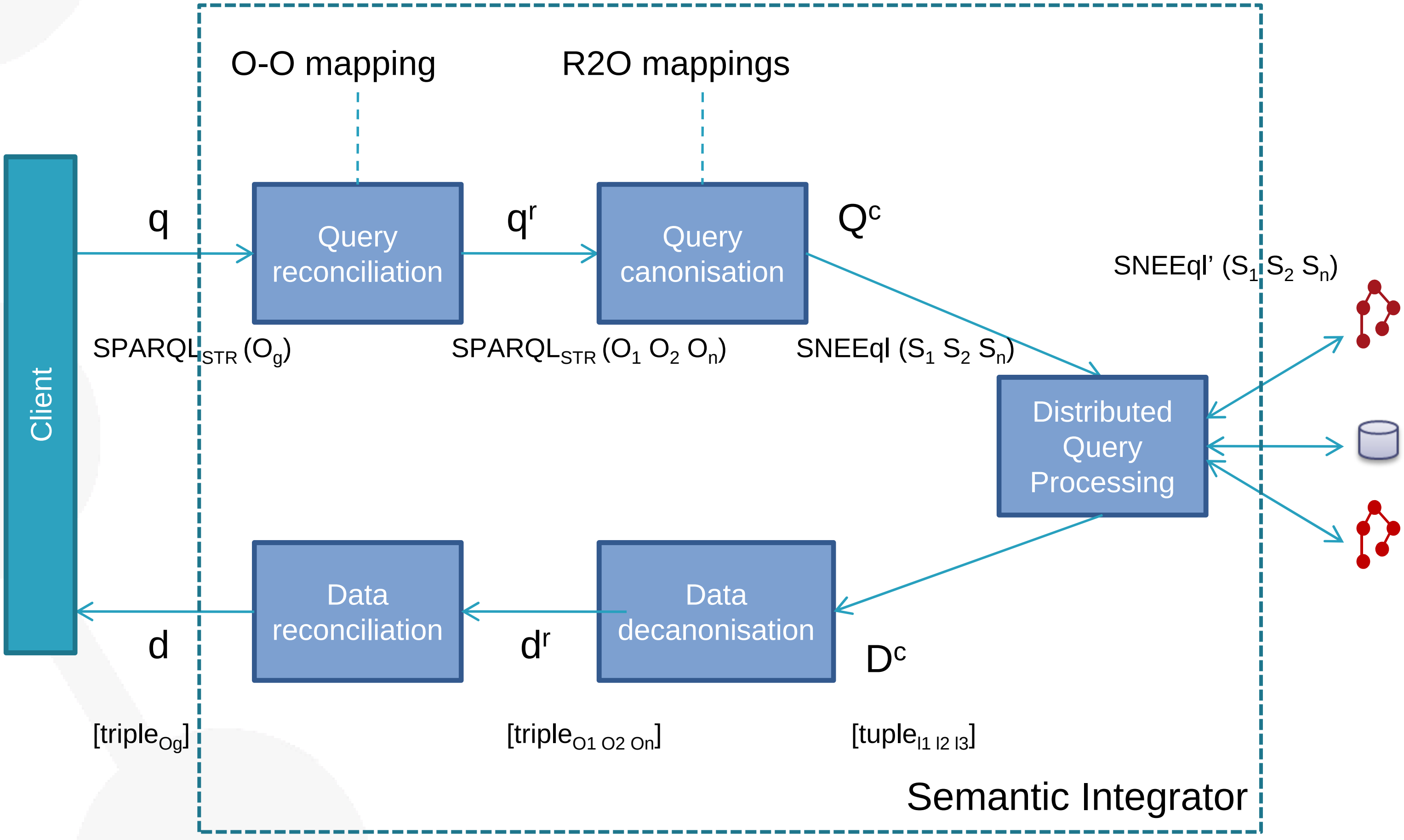
Semantic Integrator provides data access to an integrated virtual resource that integrates:

- Streaming data sources
- Stored data sources

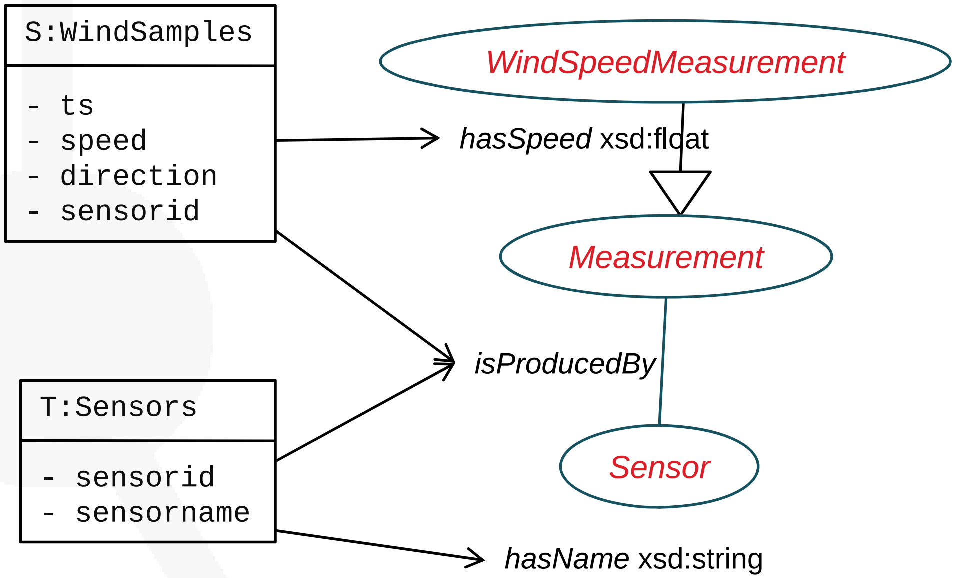
It provides a unified ontological view accessible through declarative queries.



Ontology-based Streaming Data Access



S2O: Mapping Streams to Ontologies



conceptmap-def *WindSpeedMeasurement*
  uri-as
    concat('ssg4env:WindSpeedMeasurement_',
        `windsamples.sensorid,windsamples.ts`)
  described-by
   attributemap-def *hasSpeed*
     operation "constant"
       has-column `windsamples.speed`
   dbrelationmap-def *isProducedBy* toConcept *Sensor*
     joins-via
       condition "equals"
         has-column `sensors.sensorid`
         has-column `windsamples.sensorid`

conceptmap-def *Sensor*
  uri-as
    concat('ssg4env:Sensor_',`sensors.sensorid`)
  described-by
    attributemap-def *hasName*
      operation "constant"
        has-column `sensors.sensorname`

Query Transformation Semantics

- Conjunctive Queries

$$q(\boldsymbol{x}) \leftarrow \varphi(\boldsymbol{x}, \boldsymbol{y})$$

$$\varphi(\boldsymbol{x}, \boldsymbol{y}) : \bigwedge_{i=1\ldots k} P_i, \text{ with } P_i \begin{cases} C_i(x), C \text{ is an atomic class.} \\ R_i(x, y), R \text{ is an atomic property.} \\ x = y \end{cases}$$

$$q_1(x) \leftarrow WindSpeedMeasurement(x) \wedge measuredBy(x, y) \wedge SeaSensor(y)$$

- Mapping

conjunctive query $\Psi \rightsquigarrow \Phi$ expression over streaming sources

$$\Phi_{S_1,\ldots,S_n}(\boldsymbol{x}) = \exists \boldsymbol{y}.p^{Proj}_{S_1,\ldots,S_n}(\boldsymbol{x}) \wedge p^{Join}_{S_1,\ldots,S_n}(\boldsymbol{v}) \wedge p^{Sel}_{S_1,\ldots,S_n}(\boldsymbol{v})$$

$$\lambda(\Phi_{S_1,\ldots,S_n}(\boldsymbol{x})[t_i, t_f, \delta]) = \pi_{p^{Proj}}(\bowtie_{p^{Join}} (\sigma_{p^{Sel}}(\omega_{t_i,t_f,\delta} S_1), \ldots, \sigma_{p^{Sel}}(\omega_{t_i,t_f,\delta} S_n)))$$

From SPARQL$_{STR}$ to SNEEql

- **PREFIX** fire: http://www.semsorgrid4env.eu#
- **PREFIX** rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
- **SELECT** ?speed ?name
- **FROM STREAM** <http://www.ssg4env.eu/Readings.srdf>
- [**RANGE** 10 MINUTE **STEP** 1 MINUTE]
- **WHERE** {
- ?WindSpeed a *fire:WindSpeedMeasurement*;
- *fire:hasSpeed* ?speed;
- *fire:isProducedBy* ?sensor;
- *fire:hasTimestamp* ?time.
- ?sensor *a fire:Sensor*;
- *fire:hasName* ?name.
- }

SELECT concat(‘ssg4env.eu#Sensor' , sensors.sensorid) as a1 ,
(sensors.sensorname) as name
FROM sensors

SELECT concat(‘ssg4env.eu#WindSpeedMeasurement' ,
windsensor.id , windsensor.ts) as a1 ,
(windsensor.speed) as speed
FROM windsensor[**FROM** NOW - 10 **TO** NOW MIN]

SELECT concat(‘ssg4env.eu#WindSpeedMeasurement' ,
windsensor.id, windsensor.ts) as a1 ,
concat(‘ssg4env.eu#Sensor' , sensors.sensorid) as a2
FROM sensors, windsensor[**FROM** NOW - 10 **TO** NOW MIN]
WHERE (sensors.sensorid = windsensor.id)

Work in progress: removing redundant queries, basic optimisations, more complex scenarios

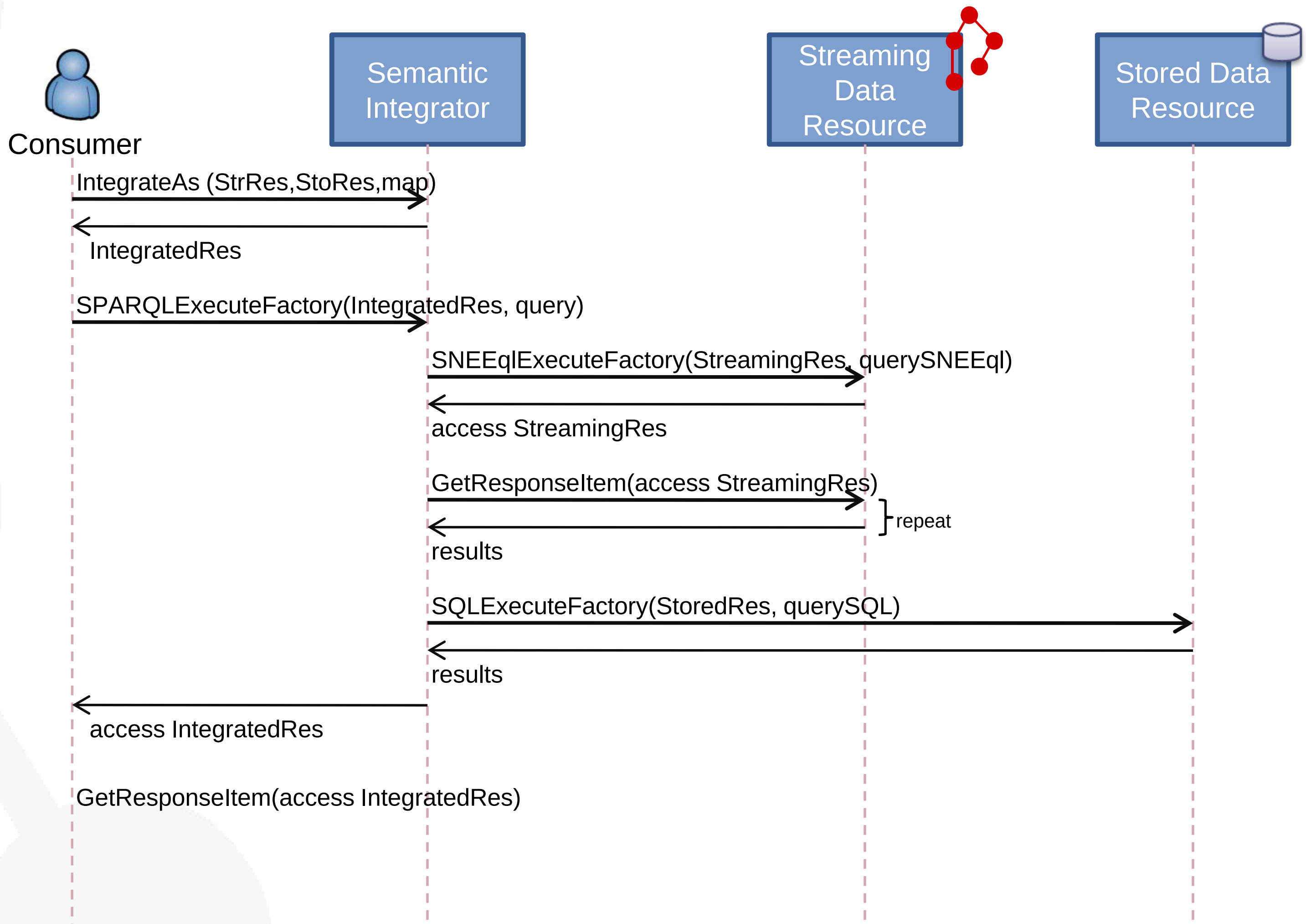
So Far…

- Starting with SPARQL support
- Define «S2O » extensions for R2O
- Define $\text{SPARQL}_{\text{STR}}$ language syntax and semantics
- Engine support for « S2O » documents, $\text{SPARQL}_{\text{STR}}$ queries
- Engine support for SNEEql translation and connection
- Limited to non-distributed scenario initially
- Initial implementation-> prototype D4.2v1

Future Works

- Ontology-based streaming data access
  - Query Rerwiting
  - Continue implementation of SPARQL streaming extensions
    - Multiple graph pattern matching, aggregates, projections, construct…
  - Continue implementation of SNEEql translation
    - Eliminate redundant queries, basic optimizations
    - Support for tuple-based windows, slide parameters
  - Provision as a Service in the SemSorGrid4Env Architecture
- Ontology-based streaming data integration
  - Continue implementation of R2O support for streams
    - Support for push and pull based streams
    - Add mapping of multiple virtual RDF streams
    - Integration mapping support (Planning)
    - DQP Integration
  - Support for quality of service parameters

Semantic Integration Interactions



References

•Arasu, A., Babcock, B., Babu, S., Cieslewicz, J., Datar, M., Ito, K., Motwani, R., Srivastava, U., Widom, J.: Stream: The stanford data stream management system. In Garofalakis, M., Gehrke, J., Rastogi, R., eds.: Data Stream Management. (2006)

•Sahoo, S.S., Halb, W., Hellmann, S., Idehen, K., Jr, T.T., Auer, S., Sequeda, J., Ezzat, A.: A survey of current approaches for mapping of relational databases to RDF. W3C (January 2009)

•Arasu, A., Babu, S., Widom, J.: The cql continuous query language: semantic foundations and query execution. The VLDB Journal 15(2) (June 2006) 121-142

•Brenninkmeijer, C.Y., Galpin, I., Fernandes, A.A., Paton, N.W.: A semantics for a query language over sensors, streams and relations. In: BNCOD '08. (2008) 87-99

•Barrasa, J., Oscar Corcho, Gomez-Perez, A.: R2O, an extensible and semantically based database-to-ontology mapping language. In: SWDB2004. (2004) 1069-1070

•Lenzerini, M.: Data integration: a theoretical perspective. In: PODS '02. (2002) 233-246

•Barrasa Rodriguez, J., Gomez-Perez, A.: Upgrading relational legacy data to the semantic web. In: WWW '06. (2006) 1069-1070

•Barbieri, D.F., Braga, D., Ceri, S., Della Valle, E., Grossniklaus, M.: C-sparql: A continuous query language for rdf data streams (to appear). In: (IJSC). (2010)

•Bolles, A., Grawunder, M., Jacobi, J.: Streaming SPARQL - extending SPARQL to process data streams. In: ESWC 08. (2008) 448-462

•Kossmann, D.: The state of the art in distributed query processing. ACM Comput. Surv. 32(4) (2000) 422-469

•Perez, J., Arenas, M., Gutierrez, C.: Semantics and complexity of sparql. ACM Trans. Database Syst. 34(3) (2009) 1-45

•Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: DL-Lite: Tractable description logics for ontologies. In: AAAI 2005. (2005) 602-607

•Poggi, A., Lembo, D., Calvanese, D., Giacomo, G.D., Lenzerini, M., Rosati, R.: Linking data to ontologies. J. Data Semantics 10 (2008) 133-173

•Perez-Urbina, H., Horrocks, I., Motik, B.: Ecient query answering for owl 2. In: ISWC 2009. (2009) 489-504