

# Benchmarking Linked Open Data technology

SRbench: A Benchmark for Streaming RDF Storage Engines

# What is Database Benchmarking?

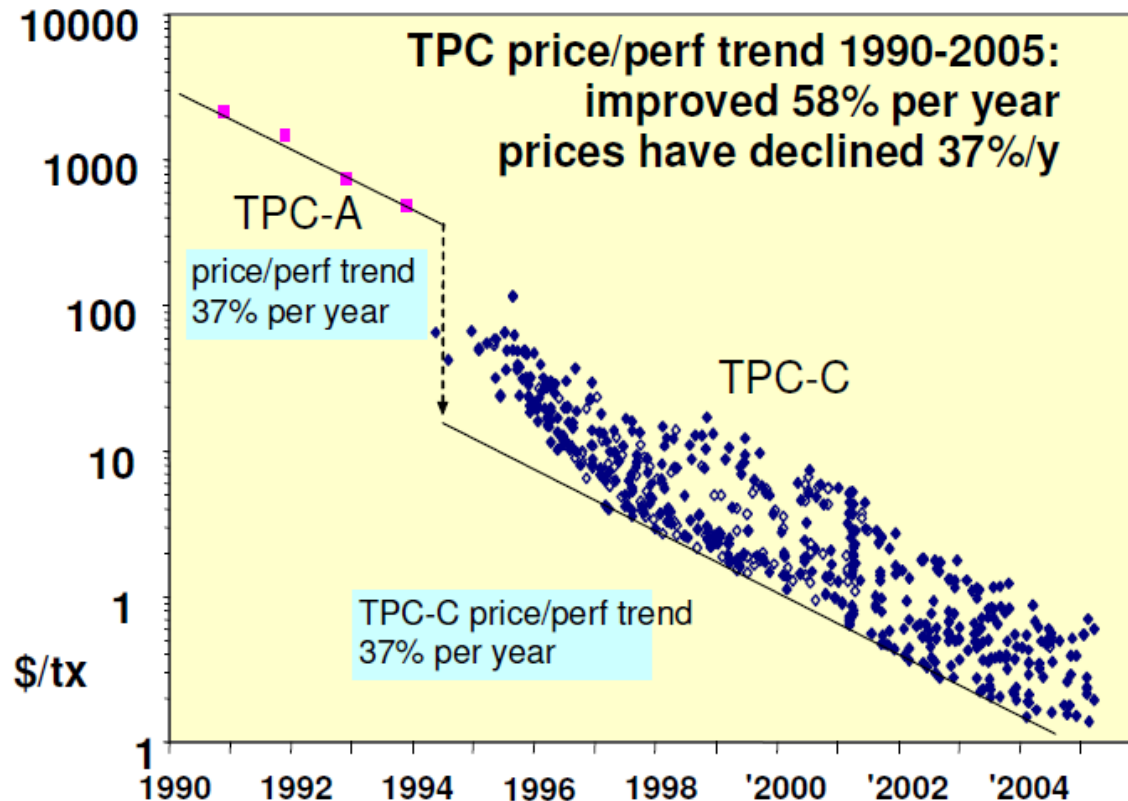
---

Standard test to measure and understand how technology performs

- ▶ **Dataset definition**
  - ▶ at various scales (100GB, 300GB, 1TB, 3TB, etc)
  - ▶ mimicks a recognizable relevant **usage scenario**
- ▶ **Database Queries**
  - ▶ often between 10-100 queries, with parameters
  - ▶ + rules/programs that specify how these queries are posed
- ▶ **Result Metrics**
  - ▶ a number to understand the result
  - ▶ tps = “transactions/second”
  - ▶ \$/QphH@size = “price per query per hour”
- ▶ **Audit Rules**
  - ▶ allow results to be checked by independent auditors
  - ▶ prevent/limit cheating

# Why Benchmarking?

- ▶ make competing products comparable
- ▶ accelerate progress, make technology viable

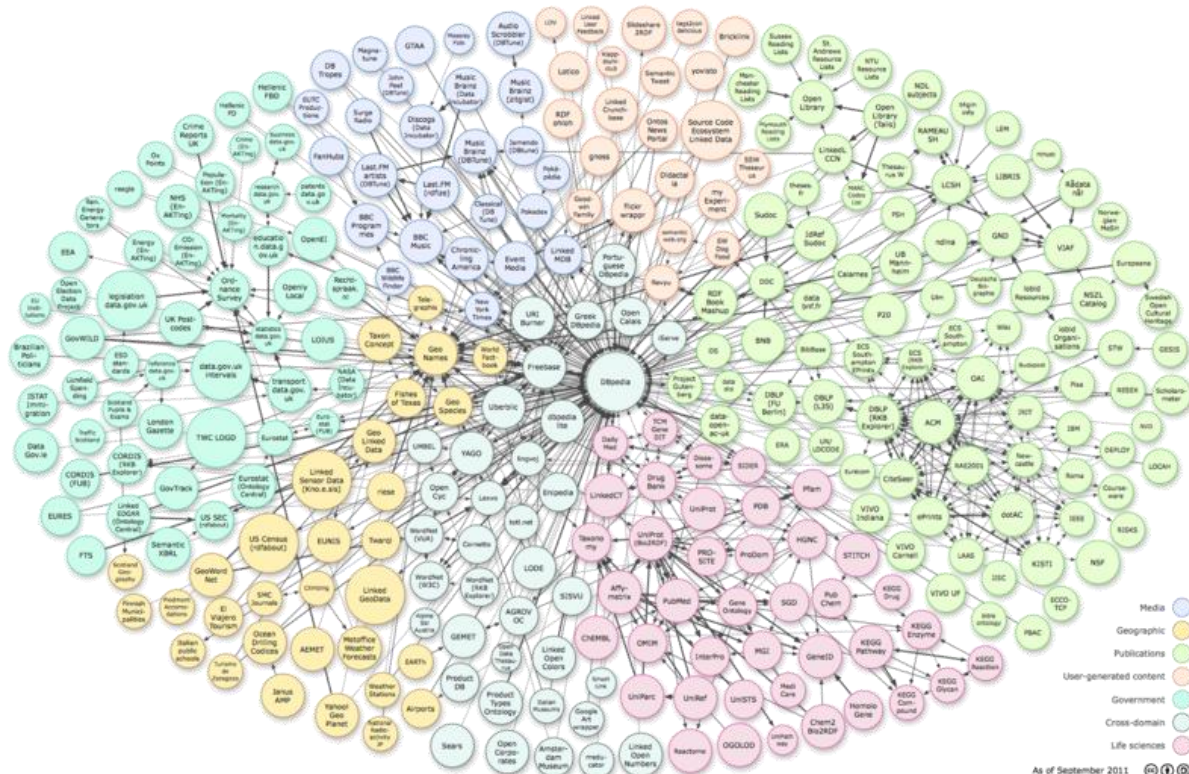


© Jim Gray, 2005

# Benchmarking LOD Technology

LOD = Linked Open Data

- ▶ web addressable data → RDF data format (W3C®)
- ▶ lots of useful data on the web (“LOD cloud”)



As of September 2011. © 100

# Benchmarking LOD Technology

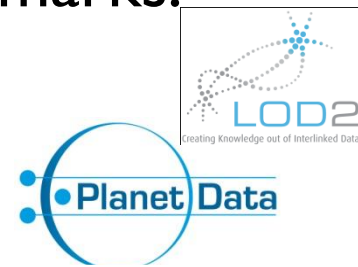
---

LOD = Linked Open Data

- ▶ web addressable data → RDF data format (W3C®)
- ▶ lots of useful data on the web (“LOD cloud”)

LOD technology (SPARQL) benchmarks:

- ▶ BSBM, DBpedia Benchmark, SIB
- ▶ SRbench ← topic of this talk
- ▶ New industry cooperation:



\* tentative/expected project



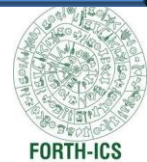
# LDDBC: FP7 2012-2015

vendor cooperation to establish accepted RDF/Graph database benchmarks and benchmark results



# LDBC Goals

1. Create the LDBC Foundation of graph and RDF DB vendors
2. Equip de LDBC Foundation with a good initial set of benchmarks, and benchmark results



# Benchmarking Linked Open Data technology



**SRbench: A Benchmark for Streaming RDF Storage Engines**



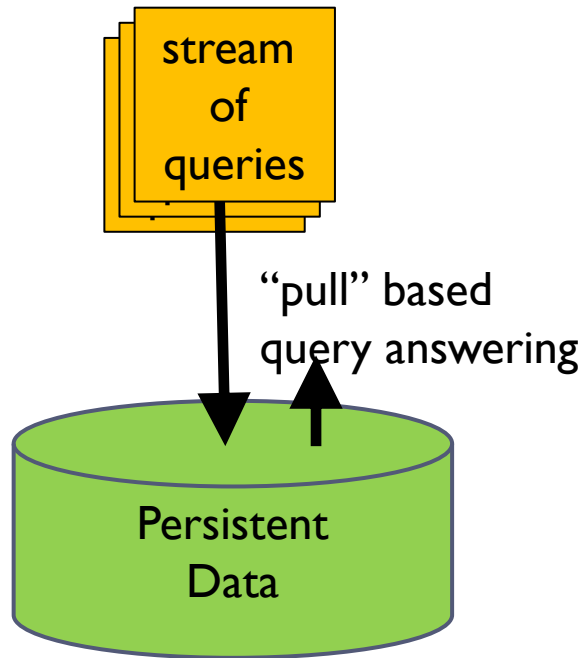
CWI

Ying Zhang, *Peter Boncz* (CWI, Amsterdam)

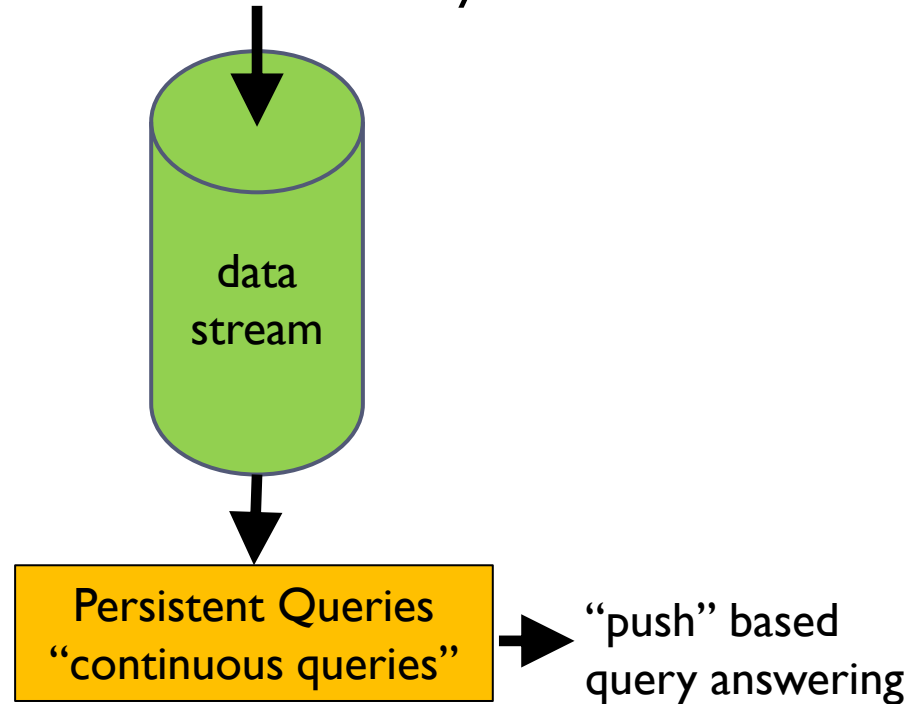


# SRbench: **Streaming** RDF Benchmark

**Traditional** Database System vs.



**Stream** Database System



# Data Streams (1 / 4): Stock Market

NASDAQ:FB: 26.71 -0.19 (-0.71%) - Facebook Inc - Mozilla Firefox

google.com <https://www.google.com/finance?client=ob&q=NASDAQ:FB>

NASDAQ:FB: ...

**Finance**

Company

Summary

News

Option chain

Related companies

Historical prices

Financials

Markets

News

Portfolios

Stock screener

Google Domestic Trends

Recent Quotes

You have no recent quotes

**26.90** -0.82 (-2.96%)

Pre-market: 26.71 -0.19 (-0.71%)

Jun 5, 8:14AM EDT

NASDAQ real-time data - Disclaimer

Currency in USD

Range 26.44 - 27.65

52 week 26.44 - 45.00

Open 27.20

Vol / Avg 8,916.00/130.10M

Mkt cap 57.51B

P/E 86.10

Div/yield -

EPS 0.31

Shares 2.14B

Beta -

inst. own 0%

7k

Compare:  Dow Jones  Nasdaq  LVWD  IMKI  THWI  CCLG  LNKD  TTGT  MRGN

Zoom: [1d](#) [5d](#) [1m](#) [3m](#) [6m](#) [YTD](#) [1y](#) [5y](#) [10y](#) [All](#)

May 31, 2012 - Jun 04, 2012 -1.29 (-4.58%)

Volume (mil / 2min)

Settings | Plot feeds | Technicals | [Link to this view](#)

Volume delayed by 15 mins.

Related companies

Show: [Most Recent Annual](#) [Add or remove columns](#)

	Company name	Price	Change	Chg %	d   m   y	Mkt Cap
<b>FB</b>	<b>Facebook Inc</b>	26.90	-0.82	-2.96%		57.51B
	LVWD LiveWorld, Inc.	0.210	0.000	0.00%		6.96M
	IMKI Immediatek Inc(NDA)	1.50	0.00	0.00%		23.80M
	THWI Thwapr Inc	0.0430	+0.0060	16.22%		2.44M
	CCLG CycleLogic, Inc.	0.0001	0.0000	0.00%		10.00
	LNKD LinkedIn Corporation	91.09	-0.42	-0.46%		9.41B
	TTGT TechTarget Inc	4.92	-0.13	-2.57%		193.47M
	MRGN Mergence Corporation	0.0020	0.0000	0.00%		18,550.00
	VEDO VillageEDOCs Inc.	0.0035	0.0000	0.00%		1.58M
	CLKZ Clicker Inc	0.300	0.000	0.00%		139,455.00
	YTNO Yatinoo, Inc.					

Sector: [Technology](#) > Industry: [Computer Services](#) [More from Revere Data](#)

**More results**

Dow Jones 12,101.46 -0.14%

Nasdaq 2,760.01 0.46%

Technology 0.19%

FB 26.90 -2.96%

**All news for Facebook Inc** [Subscribe](#)

Advertisement

**Key stats and ratios**

	Q1 (Mar '12)	2011
Net profit margin	19.38%	26.95%
Operating margin	36.01%	47.32%
EBITD margin		-56.02%
Return on average assets	12.47%	21.46%
Return on average equity	12.29%	22.91%
Employees	3,539	-
Carbon Disclosure Rating	-	-

[Screen stocks with similar metrics](#)

Description Address

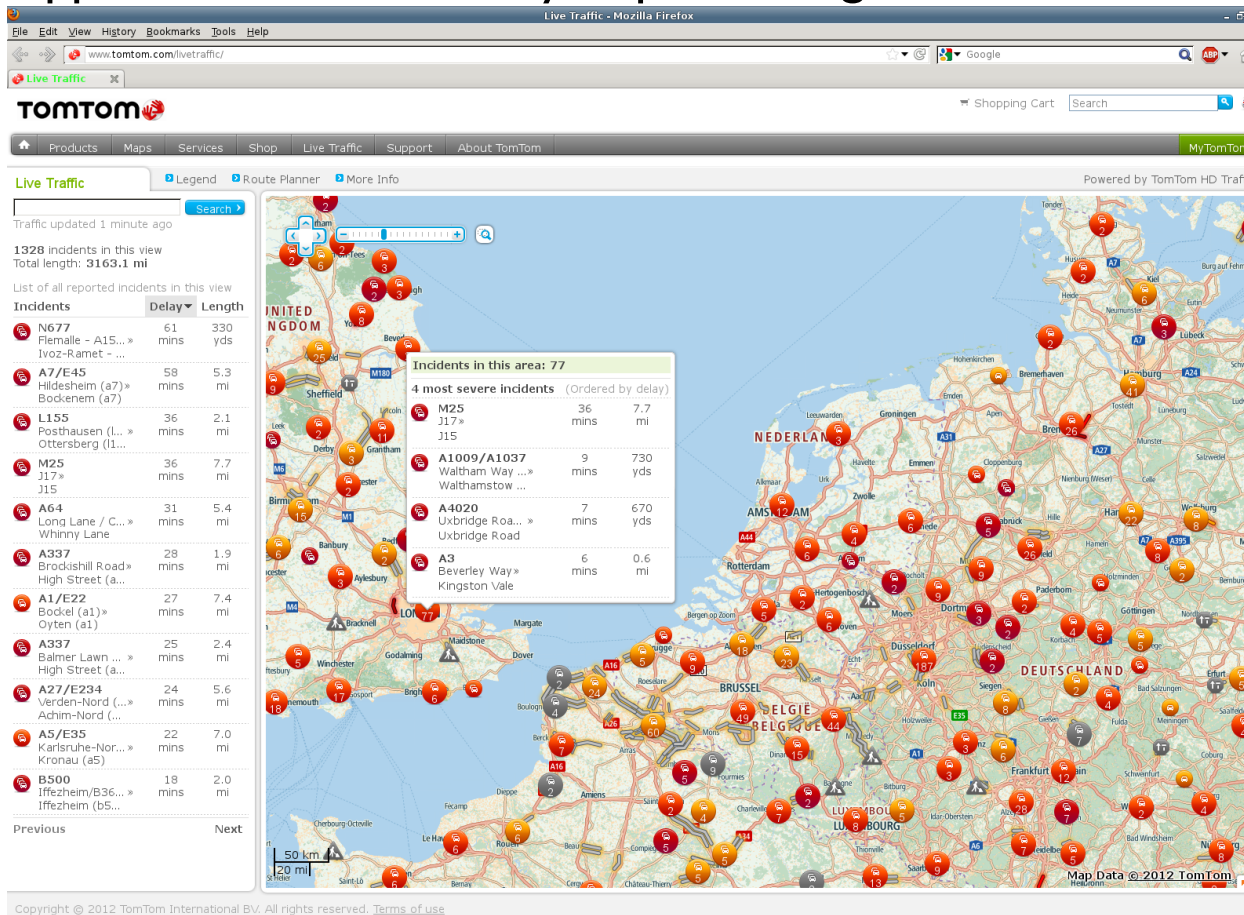
# Data Streams (2/4): Social Chatter

- ▶ Detect breaking news
- ▶ Analyze Marketing campaigns



# Data Streams (3/4): Car Traffic

- ▶ monitor positions and speeds of cars → **detect accidents**, traffic jams
- ▶ Applications: better safety, improved logistics





# Data Streams (4/4): Tele Health



Monitor health of elderly in their homes

**Who are the users?**

**Why?**

- Difficult to reach locations
- Make health care more affordable

**How?**

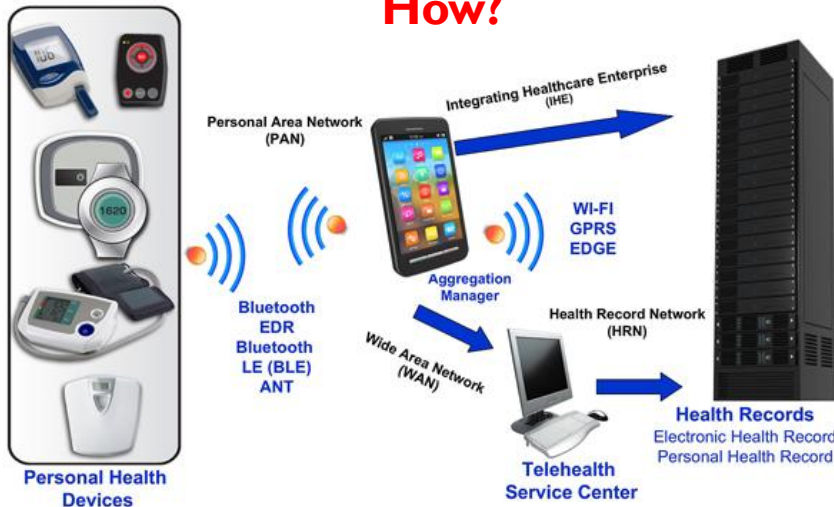


Figure 1. Basic telehealth system.



# SRbench: Streaming **RDF** Benchmark

---

Streaming RDF data **benefits**:

- ▶ apply Linked Open Data (LOD) principles to streaming data
    - ▶ Link streaming data to data on the web (**enrichment**)
    - ▶ Publish data streams on the web
  - ▶ support (simple) reasoning semantics in stream queries
- ➔ **Richer semantics** than relational streaming database systems

# SRbench: Streaming **RDF** Benchmark

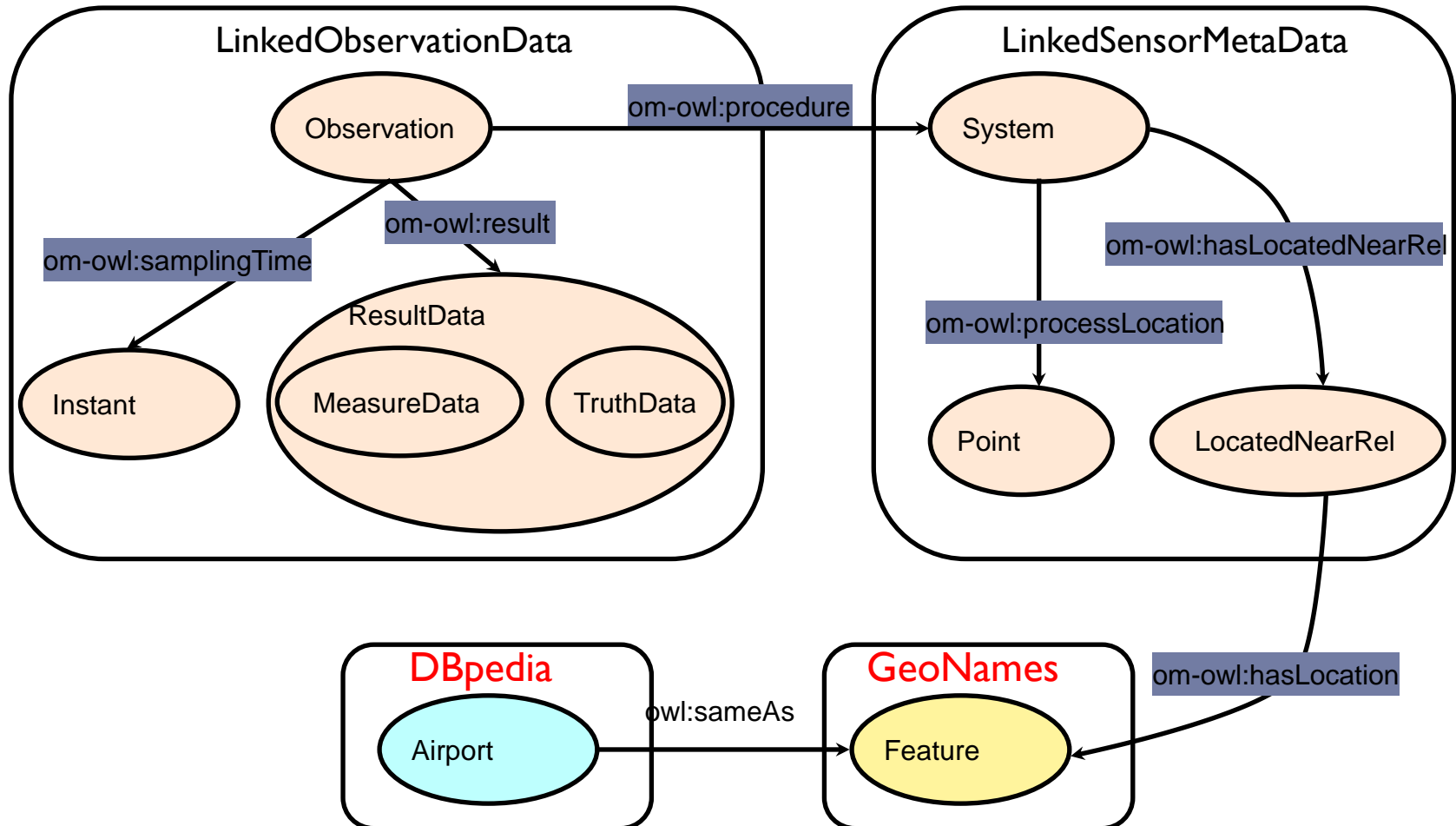
---

Streaming RDF data **challenges**:

- ▶ Proper benchmark dataset
  - ➔ use real-world datasets from LOD
- ▶ No standard query language
  - ➔ natural language query definition +  
three implementations (SPARQLStream, CQELS, C-SPARQL)
- ▶ Limited systems support
  - ➔ evaluate on the strRS system (UPM)

# SRbench: used Datasets

## LinkedSensorData





# SRBench Queries

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17
1 patter matching	A	A,F,O	A	A,F	A	A,F,U	A	A	A	A	A,F	A,F,U	A,F	A,F,U	A,F	A,F	A,F
2 solution modifier	P,D	P,D	P	P	P	P	P,D	P	P	P,D	P,D	P	P	P,D	P	P	P
3 query form																	
4 SPARQL 1.1		F,P	A	A,E,M,F	A,S		N	A,E,M	A,E,M		A,S,M,F	A,S,E,M,F,P	A,E,M,F,P	F,P	A,E,M,P	P	P
5 reasoing			R												C	A	C
6 CQL feature	T	T	T	T	T	T	T,	T	T	T	T	T	T	T	T		
7 data access	O	O	O	O	O	O	O	O,S	O,S	O,S	O,S	O,S,F	O,S,G	O,S,G	O,S,D	O,S,G,D	S

**Table 2.** Addressed features per query. Operators are abbreviated in per row unique capital letters, defined as: 1. **A**nd, **F**ilter, **U**nion, **O**ptional; 2. **P**rojection, **D**istinct, **L**imit; 3. **S**elect, **C**onstruct, **A**sk; 4. **A**ggregate, **S**ubquery, **N**egation, **E**xpr in SELECT, assign**M**ent, **F**unctions&operators, **P**roperty path; 5. sub**C**lassOf, sub**P**ropertyOf, owl:sameAs; 6. **T**ime-based window, tu**P**le-based window, **I**stream, **D**stream, **R**stream; 7. Linked**O**bservationData, Linked**S**ensorMetadata, **G**eoNames, **D**bpedia.

# Summary

---

- ▶ the importance of

- ▶ **Database System Benchmarking**

- ▶ **RDF Database System Benchmarking**

- ▶ **Streaming RDF Database System Benchmarking**



- ▶ SRbench

- ▶ Developed in PlanetData (CWI, UPM)

- ▶ **First dedicated streaming RDF/SPARQL benchmark**



- ▶ SRbench future work:

- ▶ performance evaluation

- ▶ results verification (not easy!)

# Thank You!

---

## Questions?

- ▶ Ying Zhang ([zhang@cwi.nl](mailto:zhang@cwi.nl))
- ▶ Peter Boncz ([boncz@cwi.nl](mailto:boncz@cwi.nl))

