AI for Smarter Cities. Hype or Reality?

Pascal Hitzler,
Raghava Mutharaju

Wright State University

Freddy Lecue, Jiewen Wu

(ex-IBM Research) accenture Technology Labs

Jeff Z. Pan

University of Aberdeen

IJCAI-16

Content (1)

Content (2)

Content (2)

Overview

Introduction of Speakers

Pascal Hitzler is a full professor at the Department of Computer Science, Wright State University, Ohio, USA. His research record lists over 300 publications in such diverse areas as semantic web, neural-symbolic integration, knowledge representation and reasoning. He is Editor-in-chief of the Semantic Web journal by IOS Press. He is on the editorial board of several journals and book series.

Freddy Lecue is a principal scientist and research manager in large scale reasoning systems in Accenture Technology Labs, Dublin - Ireland. He is also a research associate at INRIA, in WIMMICS, Sophia Antipolis - France. Before joining Accenture in January 2016, he was a research scientist and lead investigator in large scale reasoning systems at IBM Research - Ireland.

Raghava Mutharaju is a PhD student in the computer science department of Wright State University, USA and is supervised by Pascal Hitzler. His dissertation work involves investigating various approaches to distributed reasoning of OWL ontologies. His research interests are in knowledge representation and reasoning, distributed computing, semantic web and its applications, scalable SPARQL query processing.

Jeff Z. Pan is a Reader in the Department of Computing Science at University of Aberdeen. His research focuses on knowledge representation and reasoning, in particular scalable ontology reasoning, querying and reuse, and their applications. He is an editor of IJSWIS, an area editor of JWS and serves on the editorial board of the Journal of Big Data Research. He gave tutorials at top conferences such as AAAI and ISWC.

Jiewen Wu is a research scientist at Accenture Technology Labs, Dublin. He was a postdoctoral researcher at IBM Research, Smarter Cities Technology Center (SCTC) in Dublin, Ireland. His main research interests include knowledge representation and reasoning, the semantic web, and query processing over knowledge/data bases. He is currently working on optimization techniques for reasoning with large semantic data.

Overview

Round Table

Academia?
Industry?
PhD students?

PART I
Context

Introduction

Motivation – Today's Cities are Confronted with Serious Dilemmas

By 2050 over 6 billion people, two thirds of humanity, will be living in towns and cities

Water problems affect half of humanity!!!!

1.1 billion people in developing countries have inadequate access to water, and 2.6 billion lack basic sanitation

1.6 billion people — a quarter of humanity — live without electricity!

South Asia, Sub-Saharan Africa and East Asia have the greatest number of people living without electricity (as high as 706 million in South Asia)

Indoor air pollution resulting from the use of solid fuels [by poorer segments of society] is a major killer!

Claims the lives of 1.5 million people each year, more than half of them below the age of five (4000 deaths per day)

Source: 2008 UN Habitat; Smart Cities How will we manage our cities in the 21C?, Colin Harrison IBM Corporate Strategy

Introduction

Motivation – Regions have both Common and Unique Challenges

ASIA	N.A. & EUROPE	AFRICA	LATIN AMERICA
Rapid expansion	*Negative growth*	*Rural exodus increasing poverty*	*Decentralization*
			
▪ Over the next decade, Asia's urban areas will grow by more than 100,000 people a day ▪ Growth rates are more rapid than the investment in infrastructure ▪ Benefits of new infrastructure investments have not been distributed equally	▪ 46 countries (including Germany, Italy, most former Soviet states) are expected to be smaller in 2050 ▪ The number of shrinking cities has increased faster in the last 50 years than the number of expanding cities	▪ In 2008, more than 12M Africans left their rural homes to live in urban areas ▪ The projected increase in urban migration will exacerbate the problems of providing infrastructure, sanitation. health services, and food	▪ Large cities have incorporated nearby villages and towns – as a result, large urban areas developed sub-centers whose functions duplicated those of the central city ▪ Many large cities are competing with their outlying suburbs for people, revenue, and employment

Source: Various; IBM MI Analysis;

Introduction

Motivation – Rapidly Growing Interdependency and Complexity

We have built a world of massive complexity and interdependency….

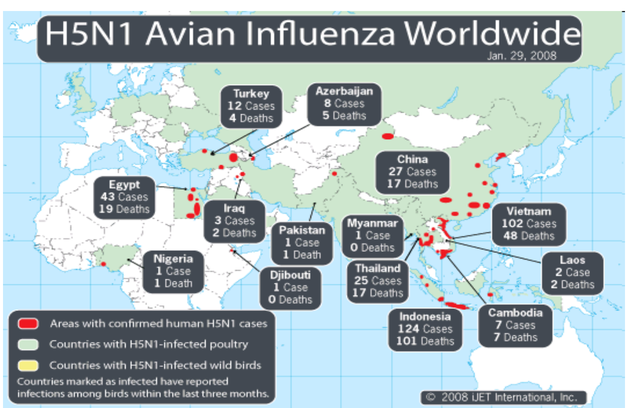

24 Hours of Air Travel


Global Trade


Global Financial Markets


Nuclear Technology

….and along with progress, we have brought on massive risks we don't manage well



Pandemics

Global Financial Crisis


Nuclear Disasters

Introduction

Motivation – Beyond the practical objectives, cities have ‘aspirations’

Source: Various; IBM MI Analysis;

Introduction

Motivation – Socio-Economic Context

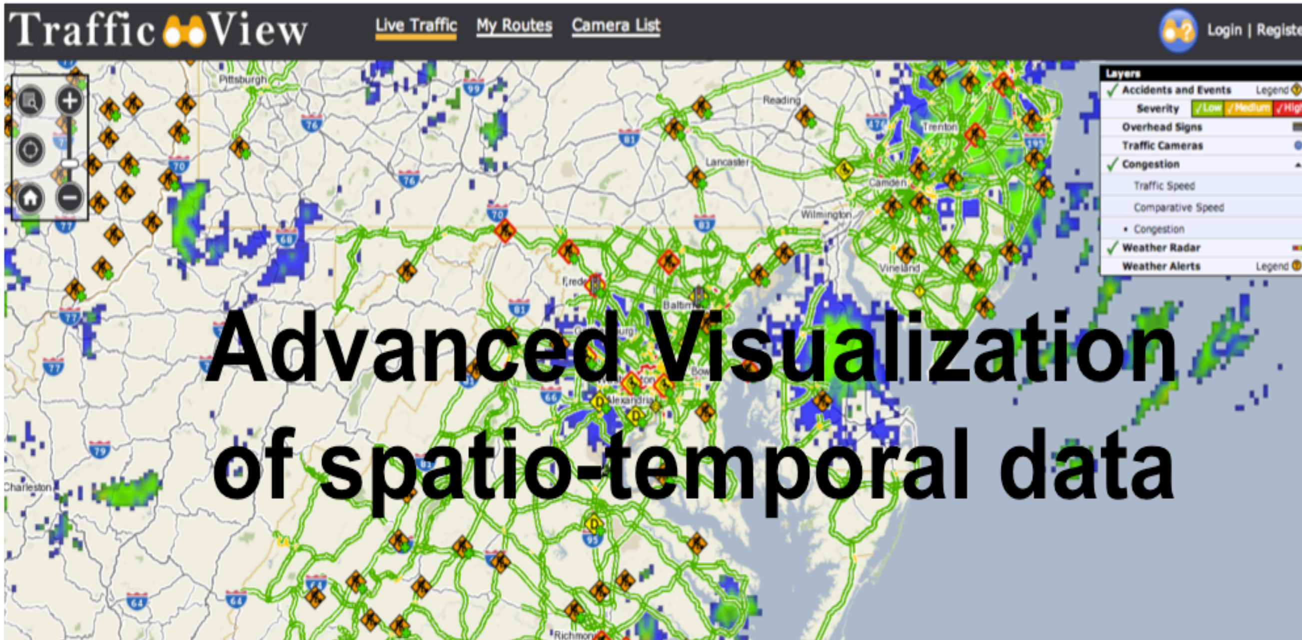
- **5.5 billion** **hours** of travel delay
- **2.9 billion** **gallons** of wasted fuel in the USA
- **$121 billion** / year
- **0.7% of USA GDP**

*5 over the past 30 years

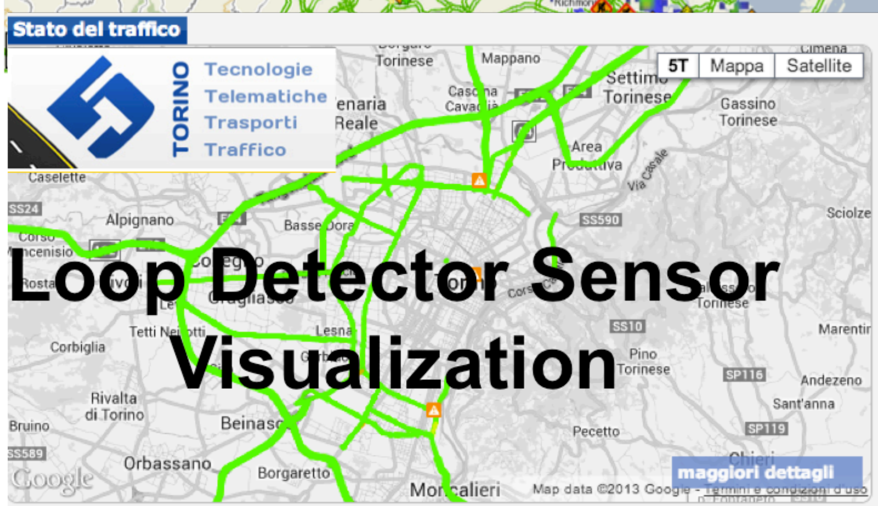
How to reduce traffic congestion ?

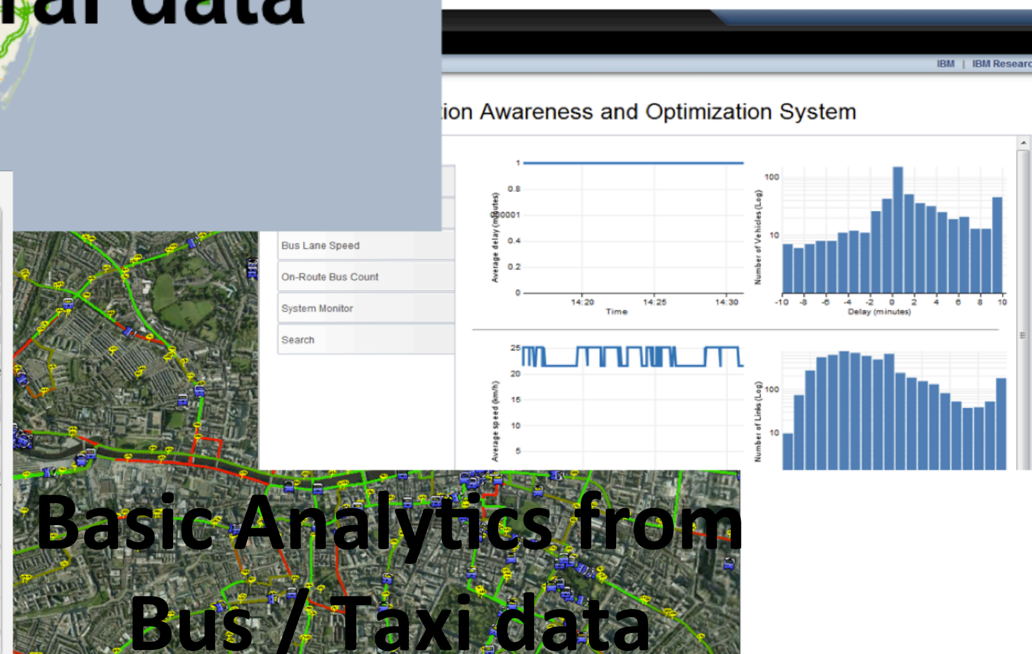
Introduction

Motivation – Limitation of Existing Systems (Example: Traffic) (1)



Most traffic systems **<u>already</u>** support Basic Analytics and Visualization!!!

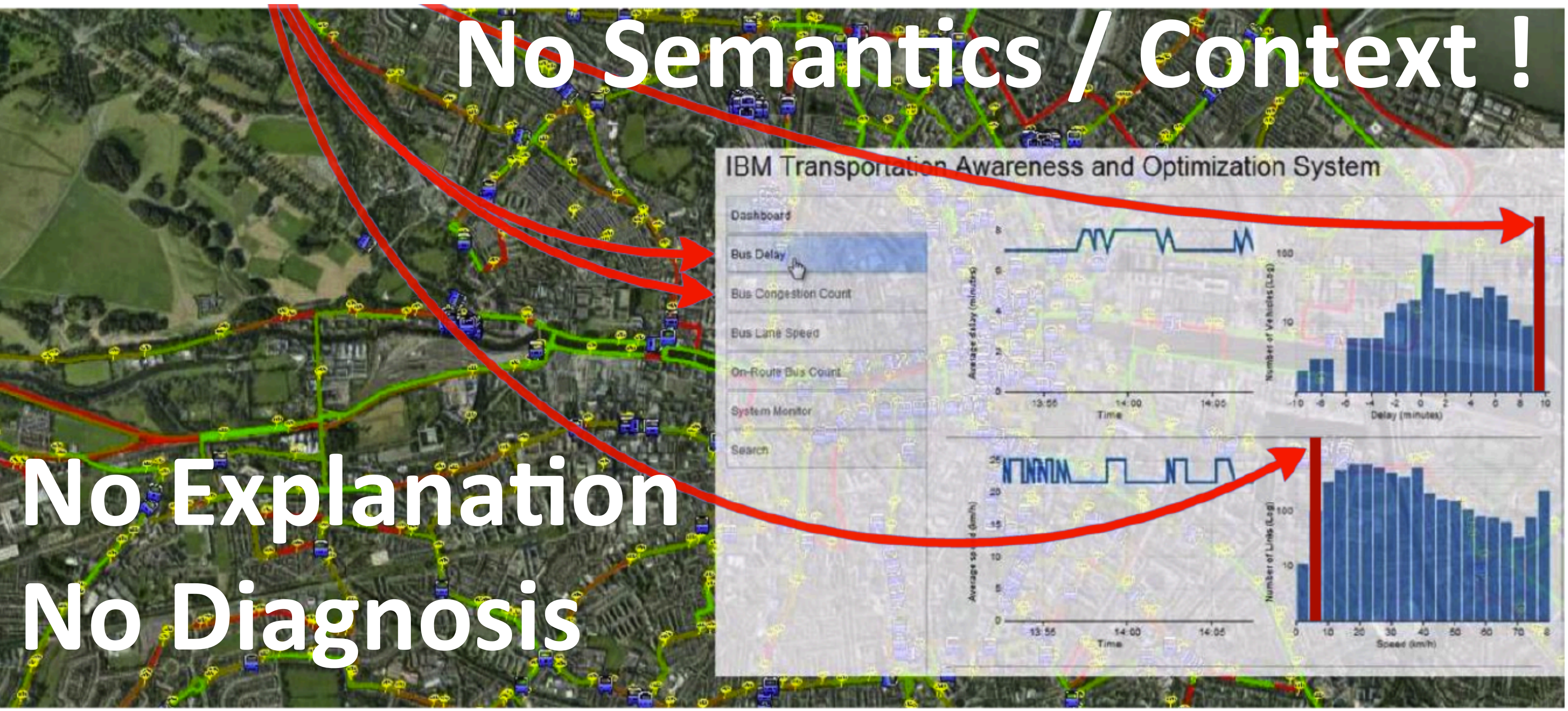




Introduction

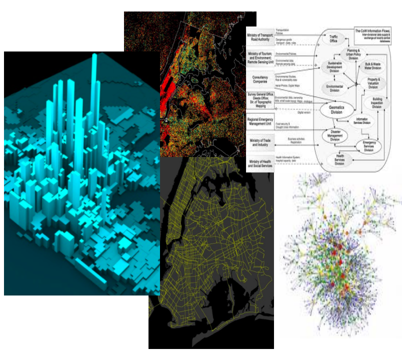
Motivation – Limitation of Existing Systems (Example: Traffic) (2)

- **All** existing traffic management systems are based on **ONE signal / stream**
- **No possible interpretation of traffic Anomalies**
- **No Integration of Exogenous Data**



Introduction

AI for cities: How can AI help cities transform ?



- **Sensor data assimilation**
  - Data diversity, heterogeneity
  - Data accuracy, sparsity
  - Data volume
- **Modelling human demand**
  - Understand how people use the city infrastructure
  - Infer demand patterns
- **Factor in Uncertainty**
  - Operations and planning
  - Organise and open data and knowledge, to engage citizens, empower universities and enable business

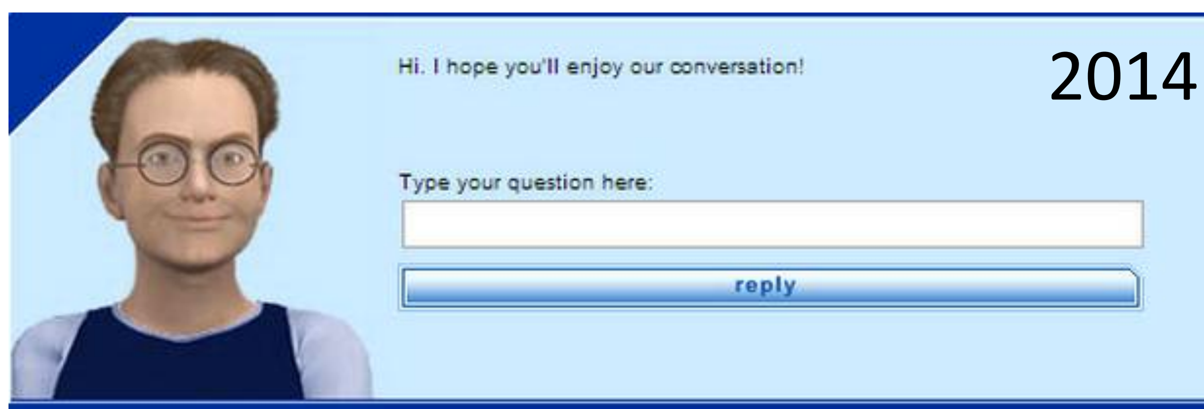
Introduction

AI for cities: why now? Open Data!

Introduction

AI for cities: why now? AI success!

Eugene Goostman, a computer programme pretending to be a young Ukrainian boy, successfully duped enough humans to pass the iconic Turing test



Introduction

AI for cities: why now? Cities want to be Smarter Cities!

Nowadays:
Issues of Urban Quality such as housing, economy, culture, social and environmental conditions.

Introduction

AI for cities: why AI? Big Data

Introduction

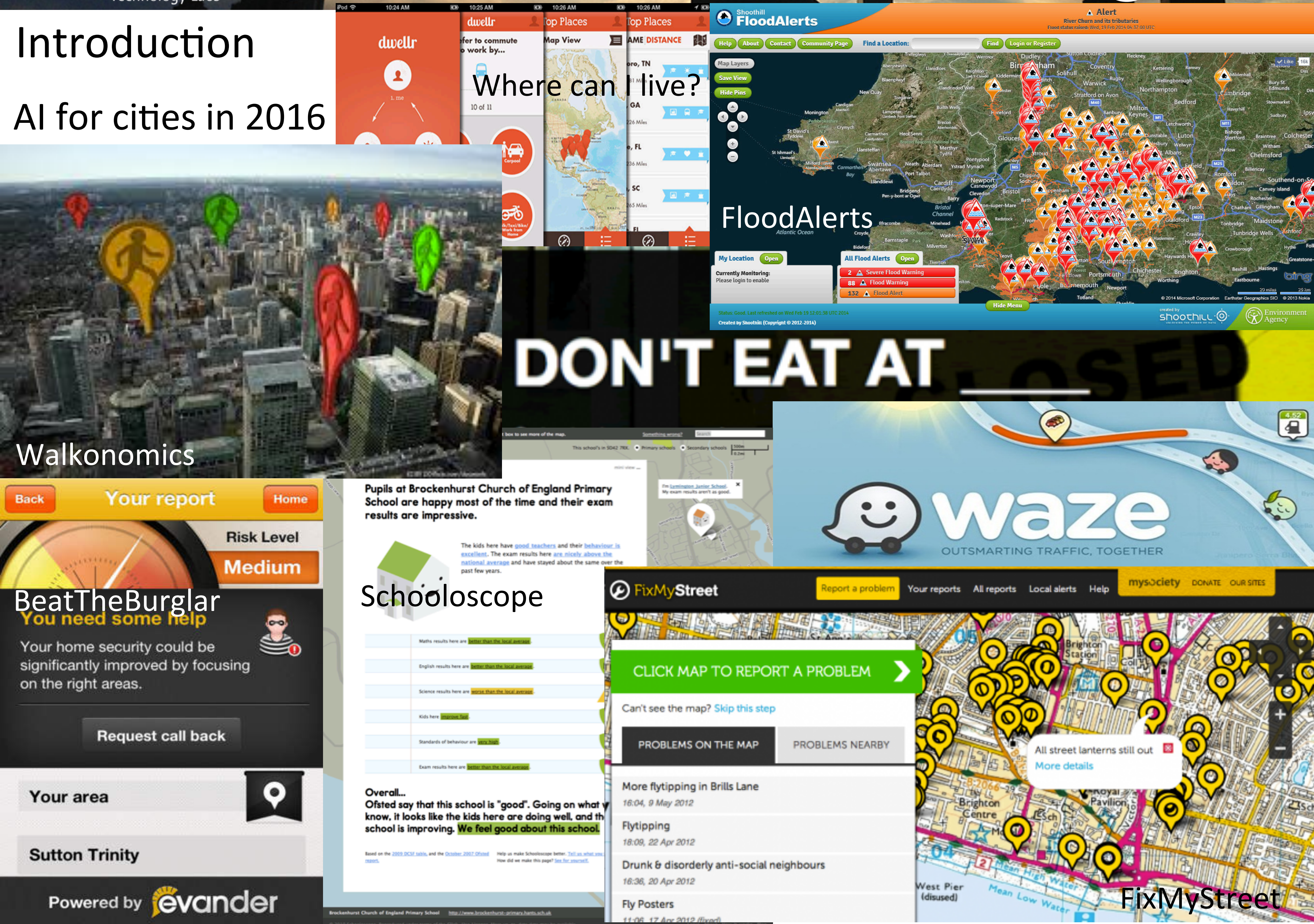
AI for cities: why AI? Smart Systems

accenture Technology Labs

IJCAI-16

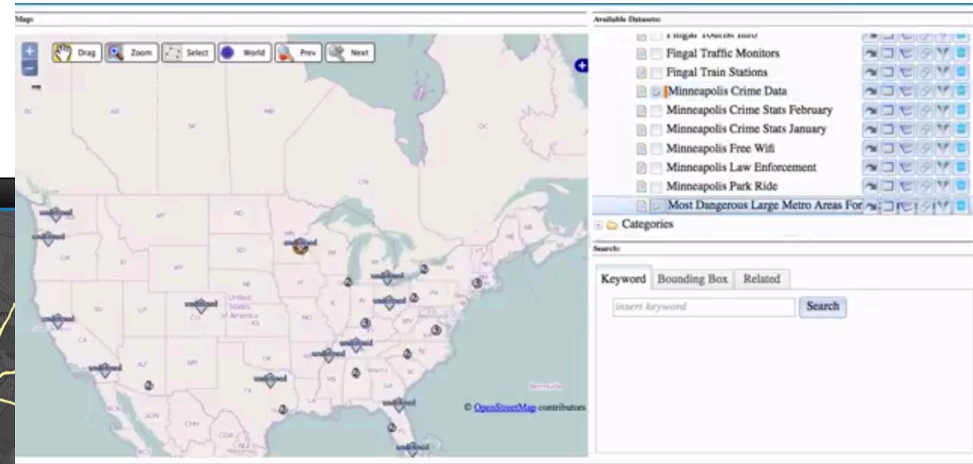
Introduction

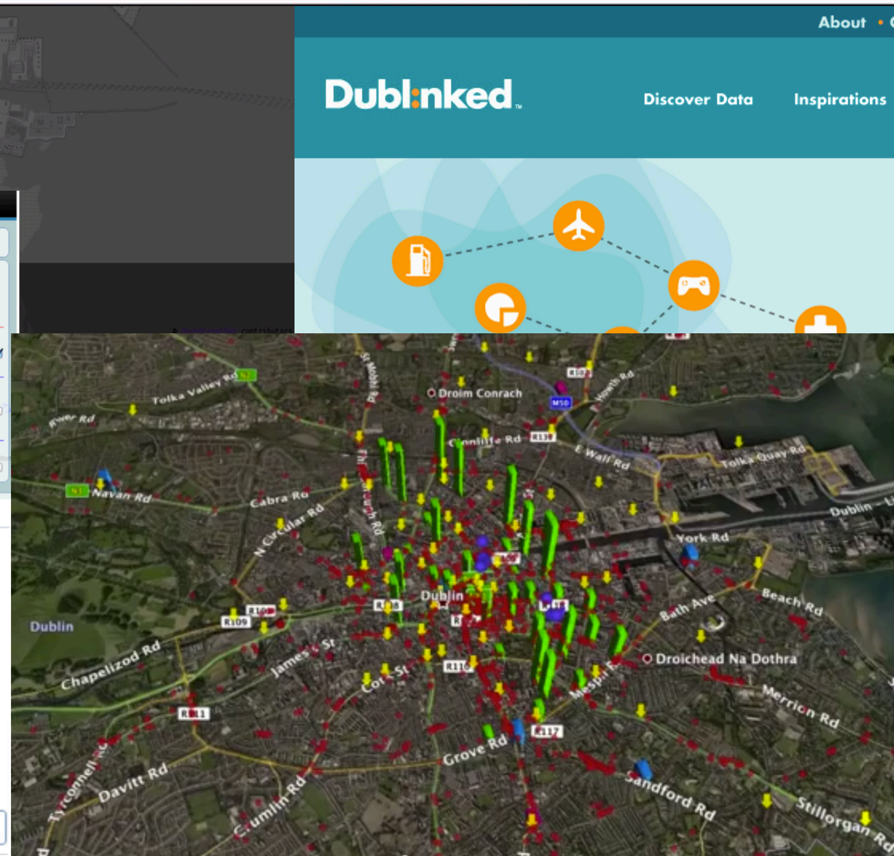
AI for cities in 2016



Introduction

AI for cities in 2016 – Our Experience

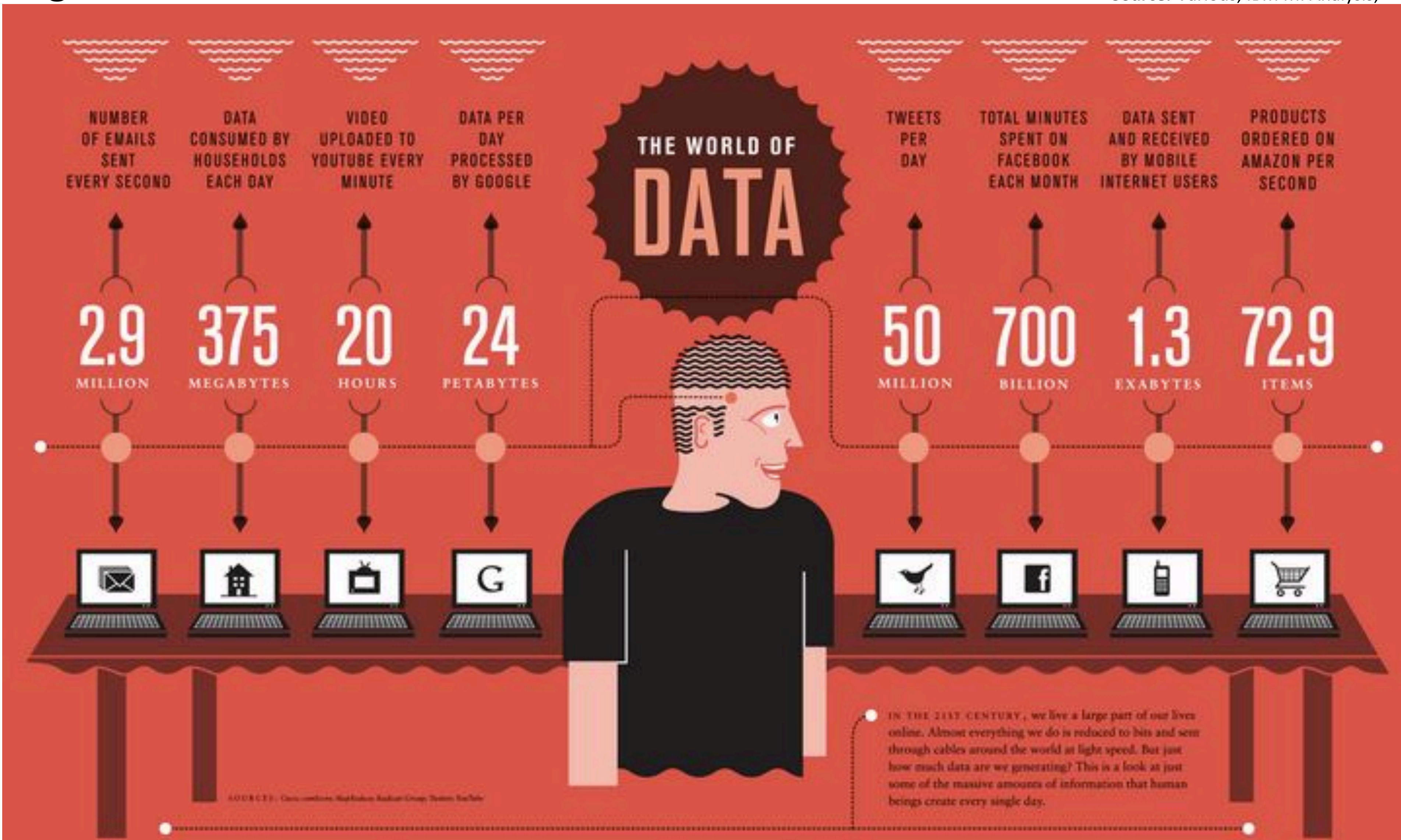




Data format and data access, collection, storage, transformation

Big Data – The World of Data

Source: Various; IBM MI Analysis;



Data format and data access, collection, storage, transformation

Big Data – The Speed of Data

Source: Various; IBM MI Analysis;

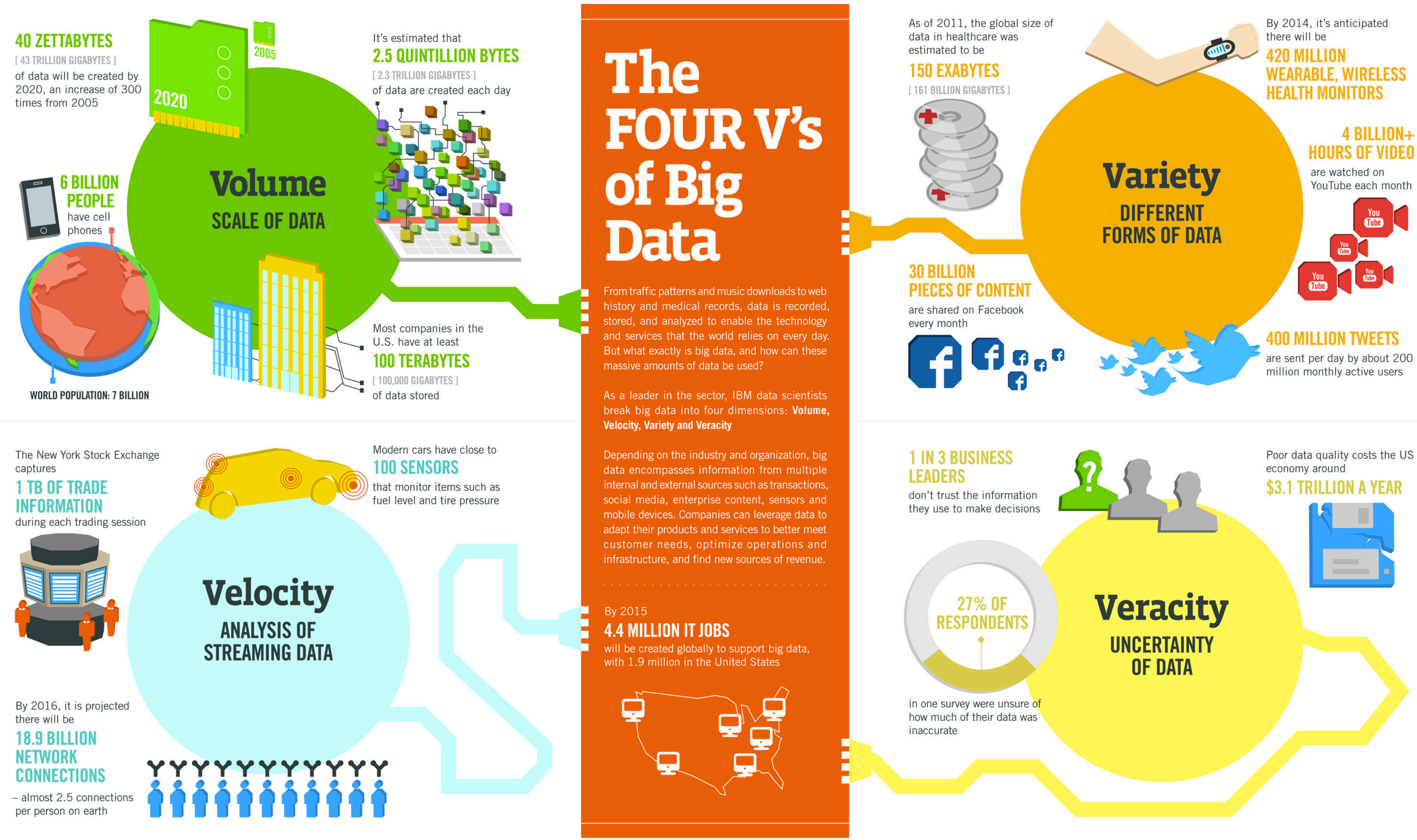
accenture Technology Labs

IJCAI-16

Data format and data access, collection, storage, transformation

Big Data – 4Vs

Source: Various; IBM MI Analysis;



Data format and data access, collection, storage, transformation

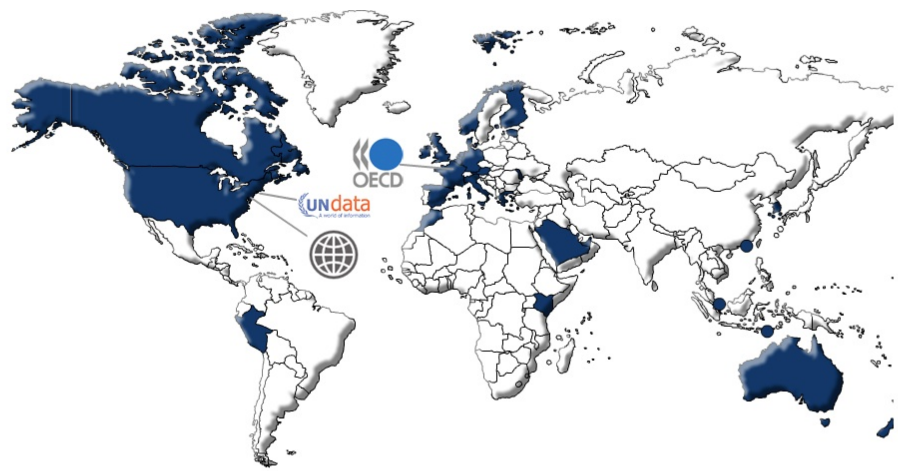
Data Format and Heterogeneity - City Data != Open Data (1)

DATA.GOV EMPOWERING PEOPLE

Think Big, Start Small, Innovate

Data.gov Quick Facts	May 2009	October 2011
Total datasets available	47	>400,000
Hits to Data.gov	0	>200 million
Apps and mash-ups by citizens and government	0	372 + 1113
RDF triples for semantic applications	0	6.7 billion
Dataset downloads	0	>2.0 million
Nations establishing open data sites	0	28
States offering open data sites	0	31
Cities in North America with open data sites	0	13
Open data contacts in Federal agencies	24	396
Agencies and subagencies participating	7	185
Communities	0	7
Community challenges	0	23

U.S. States	U.S. Cities	Agencies and Subagencies	International
31	13	172	28



A lot of relevant open data for city data analytics

(*) "Driving Innovation with Open Data", Jeanne Holm, Data.gov, February 9th, 2012 (Presentation to Ontology 2012)

Data format and data access, collection, storage, transformation

Data Format and Heterogeneity - City Data != Open Data (2)

Property management
Irish Life
CHARTERED LAND
CBRE
An tÚdarás Clárúcháin Maoine
The Property Registration Authority

Commercial valuations and rates
oifig luachála valuation office
bannon

Business & Retail
DUBLIN CITY BUSINESS IMPROVEMENT DISTRICT
DCBA

Housing
myhome.ie
Find your home at my home

Ordnance Survey Ireland
National Mapping Agency
Irish Spatial Data Exchange

Mapping

ALL-ISLAND RESEARCH OBSERVATORY
An Tlonad Breathnaithe um Thaighde Uile-Éireann
DECISIONMAP.IE

Demographics
Central Statistics Office Ireland

TrulyMadlyDeeply Dublin
Official Online Tourist Office for Dublin

Tourism
GAA

Events
Culturefox.ie

NATIONAL INVENTORY of ARCHITECTURAL HERITAGE
Dublin Civic Trust

Heritage
epa
Environmental Protection Agency

Environment

An Garda Síochána
Ireland's National Police Service

Crime

Údarás Náisiúnta Iompair
National Transport Authority

An Bord Pleanála

An Roinn Sláinte
DEPARTMENT OF HEALTH

Health

RPA
Railway Procurement Agency

Comhshaol, Pobal agus Rialtas Áitiúil
Environment, Community and Local Government

Transport & Access

Planning

FIX YOUR STREET

Fault Reporting

GREYHOUND.ie

Waste Collection

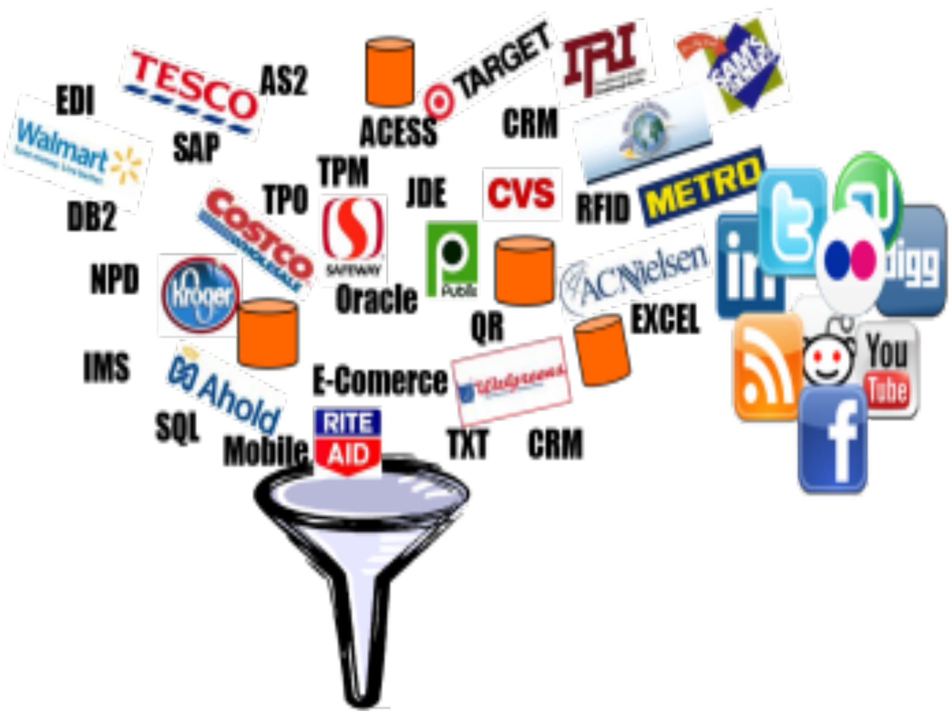
Marine Institute

Water
Dublin City
Baile Átha Cliath

Data format and data access, collection, storage, transformation

Data Format and Heterogeneity - Data Variety (1)

... In one city: **different information, different format**



Representation?

Data format and data access, collection, storage, transformation

Data Format and Heterogeneity - Data Variety (2)

... In one city: **a lot of information, no structure – go figure!**

An example from Dublin

No common schema

No common reference

No common vocabulary

Structure is not declared

No explicit semantics

	A	B	C	D
1	CSO_Code	Name	Census_2006	Census_2002
2	4001	AIRPORT	567	35
3	40[illegible]	BALBRIGGAN RURAL	3579	1335
4	4003	BALBRIGGAN URBAN	2426	2334
5	4004	BALDOYLE	1971	1990
6	4005	BALGRIFFIN	277	177
7	4006	BALLYBOGHIL	279	251
8	4007	BALSCADDEN	197	174
9	4008	BLANCHARDSTOWN-ABBOTSTOWN	1391	702
10	4009	BLANCHARDSTOWN-BLAKESTOWN	10581	7859
11	4010	BLANCHARDSTOWN-COOLMINE	3326	2629
12	4011	BLANCHARDSTOWN-CORDUFF	1520	1216
13	40[illegible]	BLANCHARDSTOWN-DELWOOD	1689	1405
14	4013	BLANCHARDSTOWN-MULHUDDART	905	524
15	4014	BLANCHARDSTOWN-ROSELAWN	615	622
16	4015	BLANCHARDSTOWN-TYRRELSTOWN	443	428
17	4016	CASTLEKNOCK-KNOCKMAROON	5629	4701
18	4017	CASTLEKNOCK-PARK	1372	1279
19	4018	CLONMETHAN	192	182

100's times

PLUS:

- No linking to authoritative sources
- Various file formats (including binary)
- Different representations for the same thing (e.g. easting/northing)
- No relations (datasets in isolation)

Data format and data access, collection, storage, transformation

Data Format and Heterogeneity - Data Variety (3)

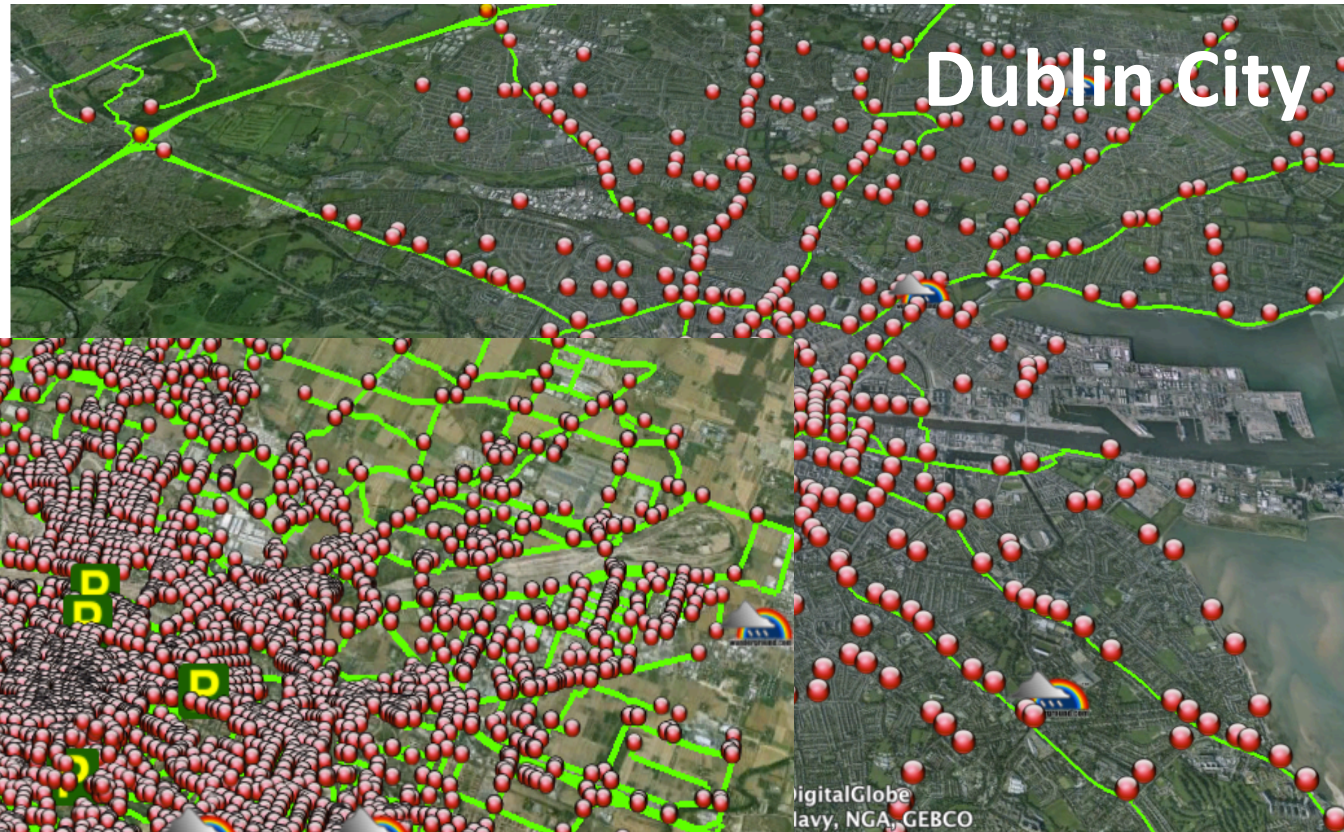
... In one application domain: **similar information, different format**

Source Type	Data Source	Description	City: Dublin (Ireland)	City: Bologna (Italy)	City: Miami (USA)	City: Rio (Brazil)
Traffic Anomaly	Journey travel times across the city	Traffic Department's TRIPS system[a]	CSV format (47 routes, 732 sensors) 0.1 GB per day[b]	✗ (not available)		
Traffic Anomaly	Dublin Bus Dynamics	Vehicle activity (GPS location, line number, delay, stop flag)	✗ (not used)	SIRI: XML format[c] (596 buses, 80KB per update 11GB per day[d])	CSV format (893 buses, 225 KB per update 43 GB per day[e])	CSV format (1, 349 buses, 181 KB per update 14 GB per day[f])
Traffic Diagnosis	Social-Media Related Feeds	Reputable sources of road traffic conditions in Dublin City	"Tweet" format - Accessed through Twitter streaming API[g]			
			Approx. 150 tweets per day[h] (approx. 0.001 GB)	✗ (not available)	Approx. 500 tweets per day[i] (approx. 0.003 GB)	✗ (not available)
Traffic Diagnosis	Road Works and Maintenance		PDF format (approx. 0.003 GB per day[j])	XML format (approx. 0.001 GB per day[k])	HTML format (approx. 0.001 GB per day[l])	✗ (not available)
Traffic Diagnosis	Social events e.g., music event, political event	Planned events with small attendance	XML format - Accessed once a day through Eventbrite[m] APIs			
			Approx. 85 events per day (0.001 GB)	Approx. 35 events per day (0.001 GB)	Approx. 285 events per day (0.005 GB)	Approx. 232 events per day (0.01 GB)
Traffic Diagnosis	Social events e.g., music event, political event	Planned events with large attendance	XML format - Accessed once a day through Eventful[n] APIs			
			Approx. 180 events per day (0.05 GB)	Approx. 110 events per day (0.04 GB)	Approx. 425 events per day (0.1 GB)	Approx. 310 events per day (0.08 GB)
Traffic Diagnosis	Bus Passenger Loading / Unloading (information related to number of passenger getting in / out)		✗ (not available)	✗ (not available)	CSV format (approx. 0.8 GB per day[e])	CSV format (approx. 0.1 GB per day[e])

Data format and data access, collection, storage, transformation

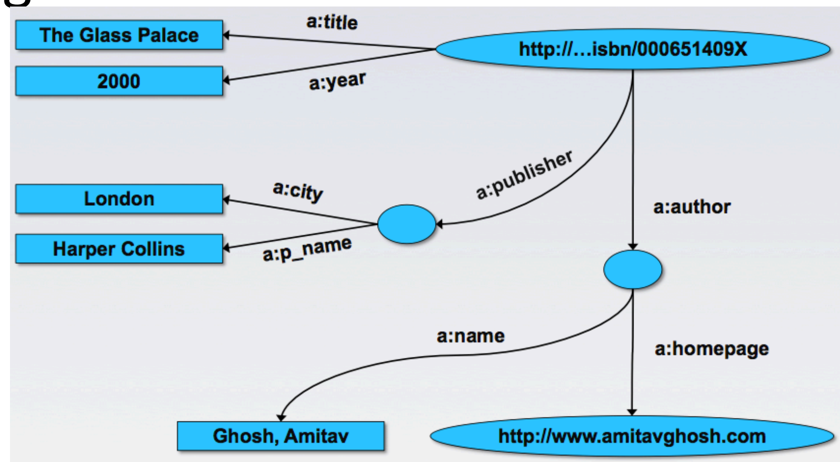
Data Format and Heterogeneity - Data Variety (4)

... In one application domain: **similar format, various sparsity**

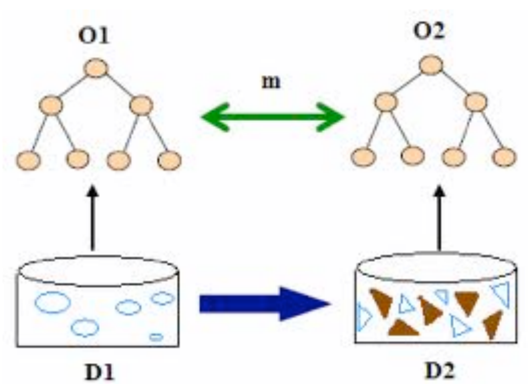


Data format and data access, collection, storage, transformation

Data Engineering – Semantic data integration



annotate
ontological
vocabulary



Map the data onto **RDF**

ID	Author	Title	Publisher	Year
ISBN 0-00-6511409-X	id_xyz	The Glass Palace	id_qpr	2000

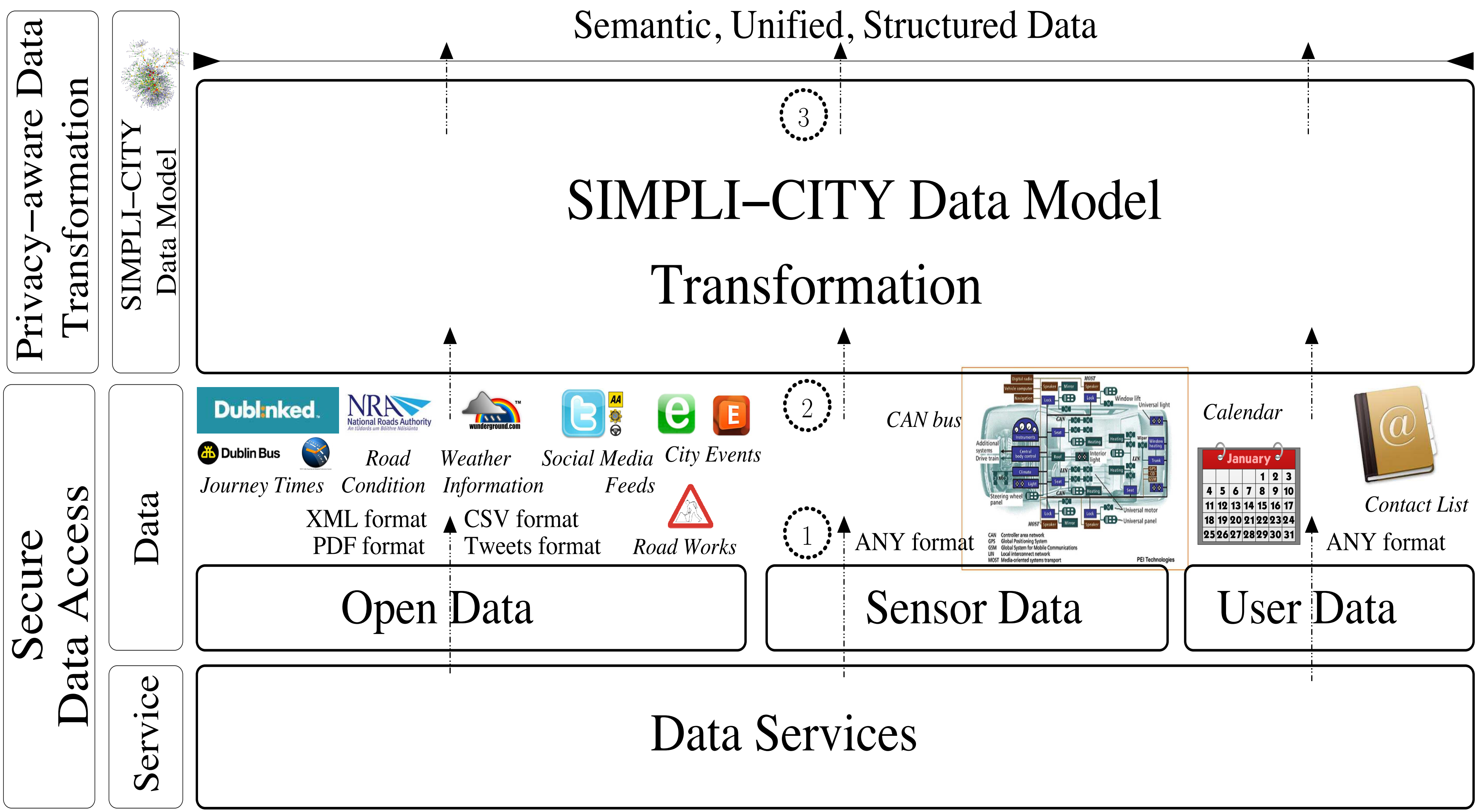
ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

ID	Publisher's name	City
id_qpr	Harper Collins	London

accenture Technology Labs IJCAI-16

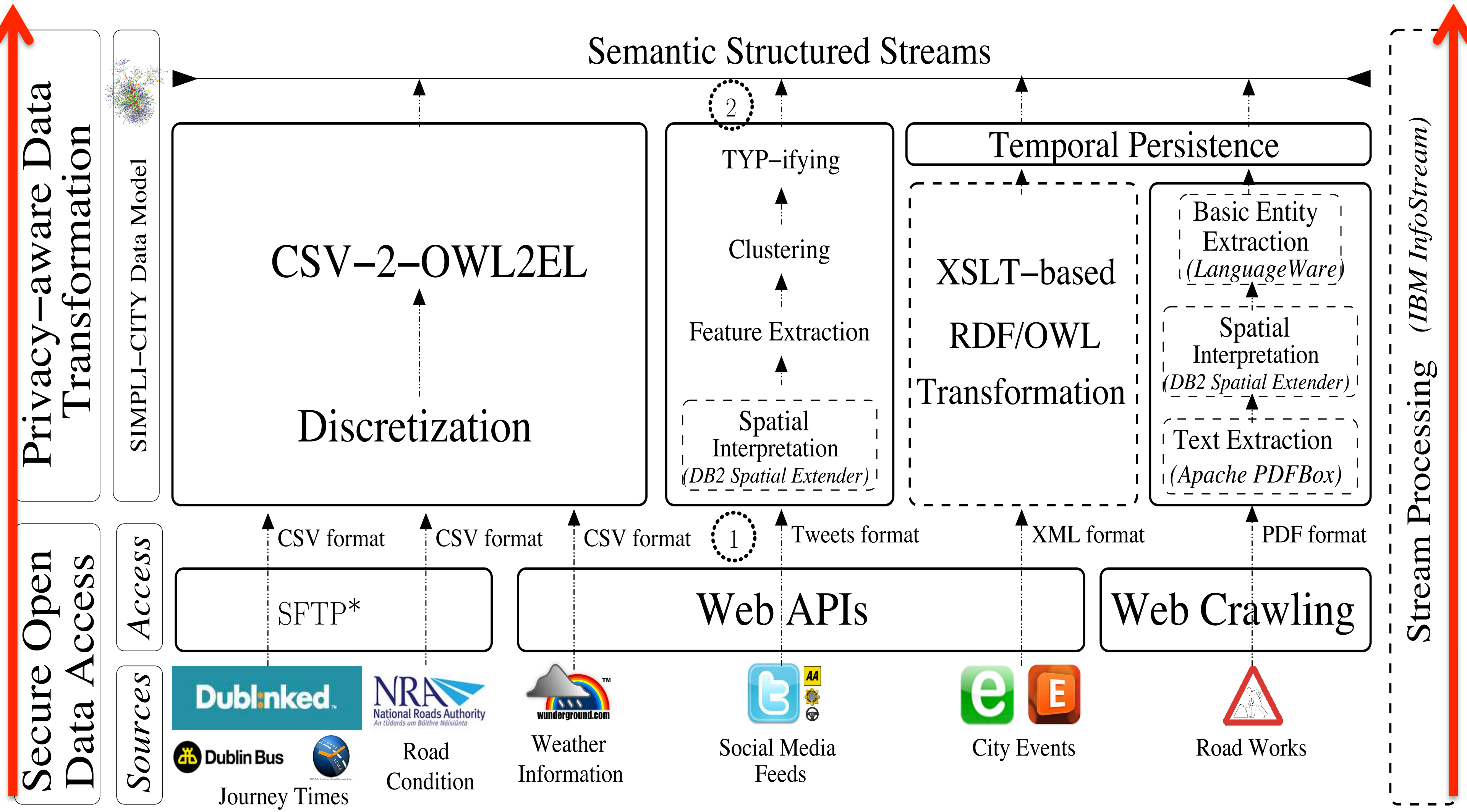
Data format and data access, collection, storage, transformation

Data Engineering – Data Access and Transformation – General



Data format and data access, collection, storage, transformation

Data Engineering – Data Access and Transformation – Dublin Instance



Data format and data access, collection, storage, transformation

Data Engineering – Data Access and Transformation – Travel Time Mapping

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ttr: <http://www.ibm.com/SCTC/ontology/TravelTimeOntology#> .

_:{$uuid}_0 rdf#type ttr#TravelTimeReport .   # $uuid: URI for new travel time report
_:{$uuid}_0 owl#intersectionOf _:{$uuid}_1 . # Each report: intersection of concepts (Fig.5)
_:{$uuid}_1 rdf#first _:{$uuid}_2 .          # Join to the first existential restriction in (9)
_:{$uuid}_2 rdf#type owl#Restriction .       # Existential restriction in (9)
_:{$uuid}_2 owl#onProperty ttr#hasSourceFrom # hasSourceFrom property in (9)
_:{$uuid}_2 owl#hasValue ttr#{$sourceFom} .  # Capture of $sourceFom variable in CSV
_:{$uuid}_1 rdf#rest _:{$uuid}_3 .           # Right part of the Intersection in (9)
_:{$uuid}_3 rdf#first _:{$uuid}_4 .          # Join to the second existential restriction in (9)
_:{$uuid}_4 rdf#type owl#Restriction .       # Another existential restriction in (9)
_:{$uuid}_4 owl#onProperty ttr#hasSourceTo   # hasSourceTo property in (9)
_:{$uuid}_4 owl#hasValue ttr#{$sourceTo} .   # Capture of $sourceTo variable in CSV
_:{$uuid}_3 rdf#rest _:{$uuid}_5 .           # Remaining parts of the Intersection for (7-10)
```

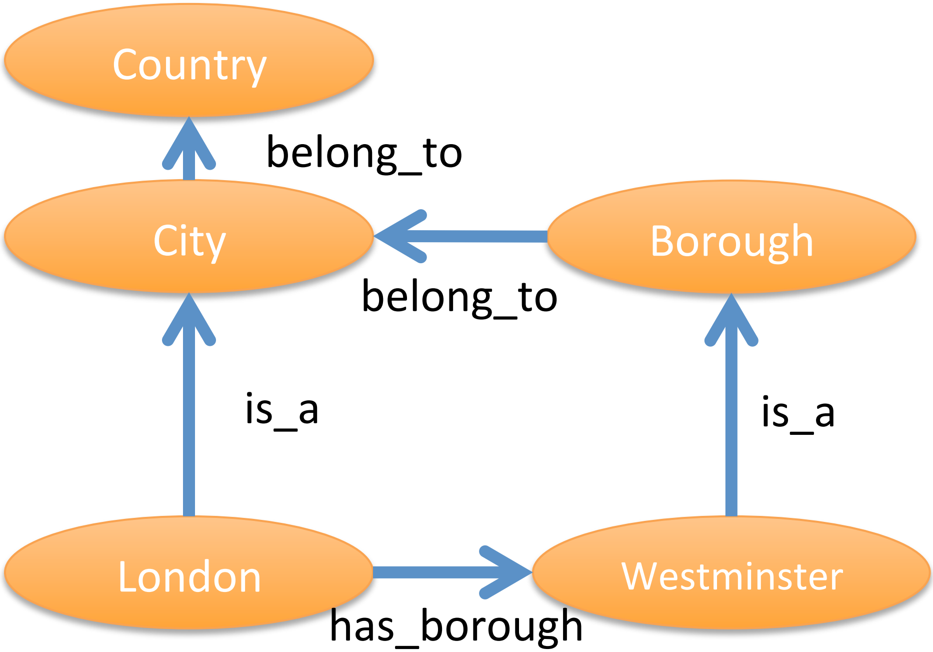
Travel time between 2 sensors

$$\mathcal{O}^{n}_{m}(t_1) : TravelTimeReport \sqcap \quad (6)$$
$$\exists createdAt.(TemporalEntity \sqcap (\exists inXSDDateTime.\{2013\text{-}04\text{-}22T23{:}01{:}00\})) \sqcap \quad (7)$$
$$\exists reportsForTimeInterval.(\exists hasDurationDescription.(\exists minutes.\{1\})) \sqcap \quad (8)$$
$$\exists hasSourceFrom.\{TRIPS\text{-}DCC\text{-}44\} \sqcap \exists hasSourceTo.\{TRIPS\text{-}DCC\text{-}351\} \sqcap \quad (9)$$
$$\exists reportsObservation.(\{r_1\} \sqcap \exists hasTravelTimeStatus.HeavyTrafficFlow) \quad (10)$$

Data format and data access, collection, storage, transformation

Knowledge Representation (KR)

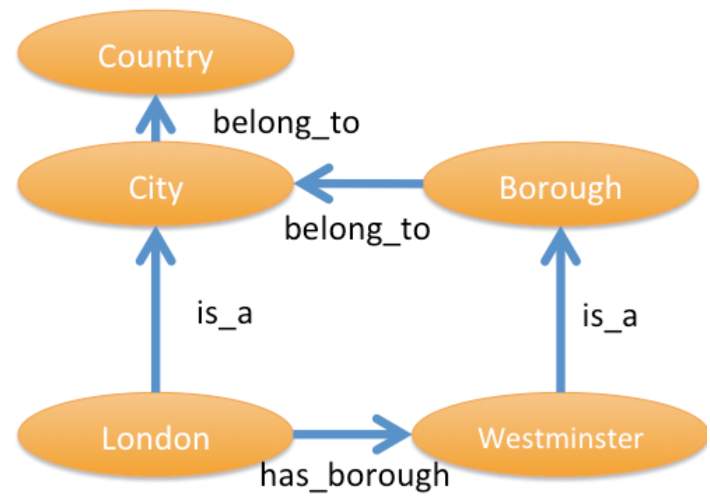
- How knowledge can be represented symbolically
    - Is City an entity or a collection of entities such as London?
    - Does each borough belong only to one city?
    - Does a city have to have some borough?
- How knowledge can be manipulated in an automated way by reasoning programs
    - trade-off between expressive power and efficiency of reasoning



Data format and data access, collection, storage, transformation

Knowledge Representation – Description Logics (DLs)

- A family of decidable sub-languages of FOL
  - Describe the domain in terms of classes/ concepts, properties/roles and individuals
- Offer constructors for class/property descriptions, such as
  - $C \sqcap D$, $C \sqcup D$, $\neg C$, $\exists R.C$, $\forall R.C$, $\leq n R.C$, $\geq n R.C$, A, $\top$, $\bot$, $\{a1,a2\}$
  - P^{-}
- A DL KB (also known as ontology) contains
  - class axioms: $C \sqsubseteq D$
  - property axioms: $P1 \sqsubseteq P2$, $P3 \circ P4 \sqsubseteq P5$
  - individual axioms: $a: C$, $(a,b): P$, $a = b$, $a \neq b$



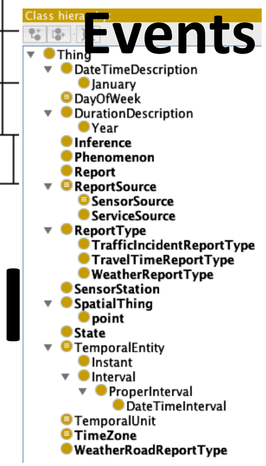
$Borough \sqsubseteq \exists belong\_to.City$
$belong\_to \circ belong\_to \sqsubseteq belong\_to$
$has\_borough^{-} \sqsubseteq belong\_to$
$London{:}City$
$Westminster{:}Borough$
$(London, Westminster){:}has\_borough$

Data format and data access, collection, storage, transformation

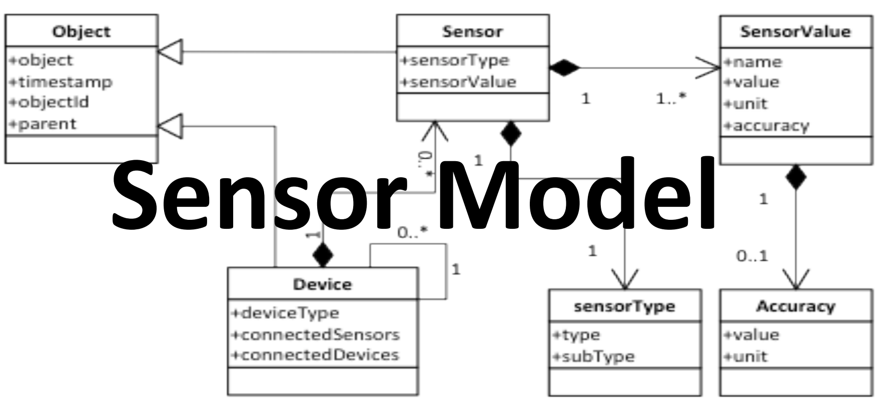
Knowledge Representation – Example DL ontologies

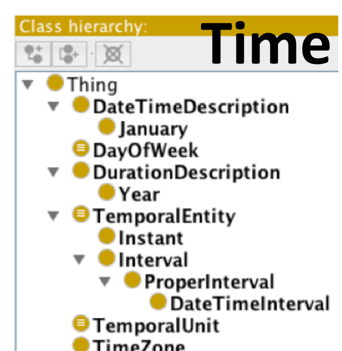
An ontology cannot rule them all … but reasoning needs an integrated one!!

Ontology	Size (KB)	#Concepts	#Object Properties	#Data Properties	#Individuals	Imported Ontologies
NASA SWEET[12] (IBM adaptation)	158.8	90	40	34	63	W3C Time, Geo
IBM Travel Time	4, 194	41	49	22	1, 429	
IBM SIRI-BUS	41.9	21	17	18	-	
W3C Time[9]	25.2	12	24	17	14	-
W3C Geo[10]	7.8	2	4	-	-	-
DBpedia	Only a subset is used for annotation i.e., 28 concepts, 9 data properties					



Data Model

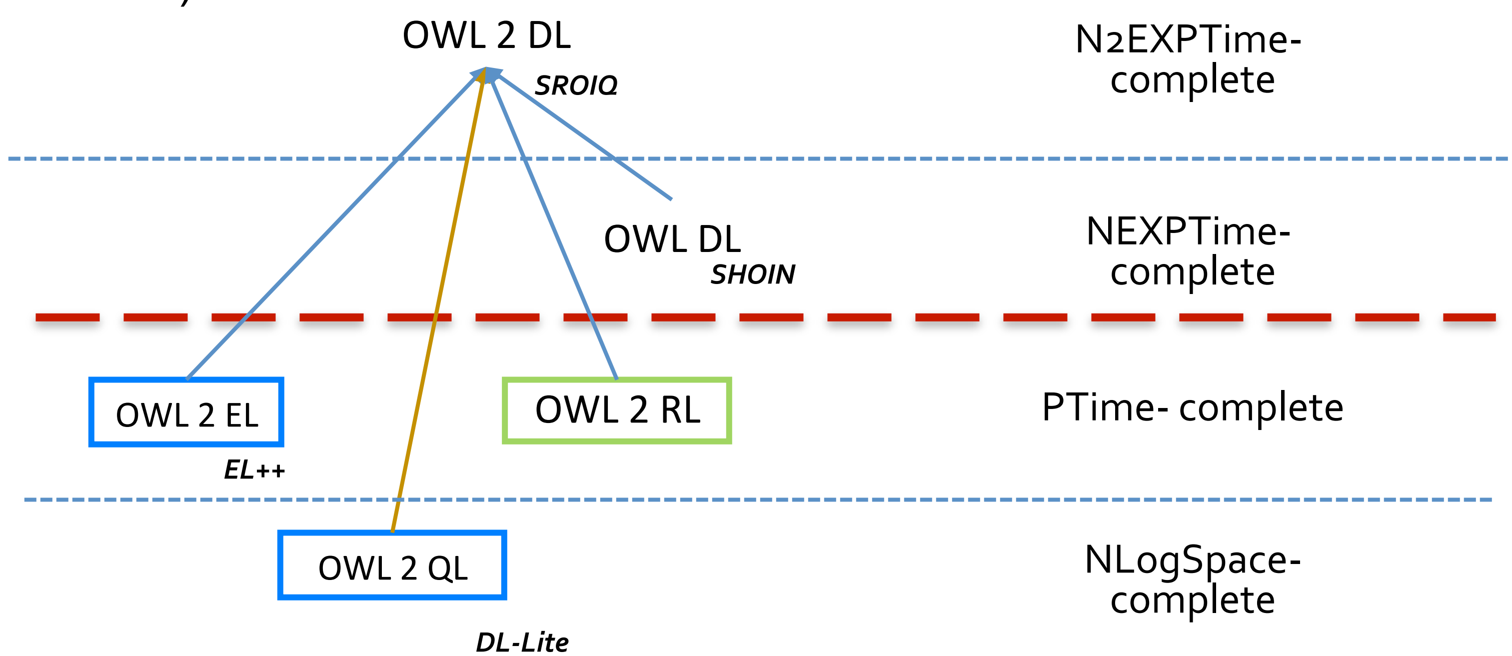




Data format and data access, collection, storage, transformation

OWL – W3C Standard Web Ontology Language

- OWL comes with different sub-languages, with different expressive power and reasoning complexity
- OWL is widely adopted by KR users
- Reasoning tasks: schema level (classification, satisfiability and subsumption checkings), schema and data (consistency and instance checkings, realisation)



Data format and data access, collection, storage, transformation

OWL 2 EL

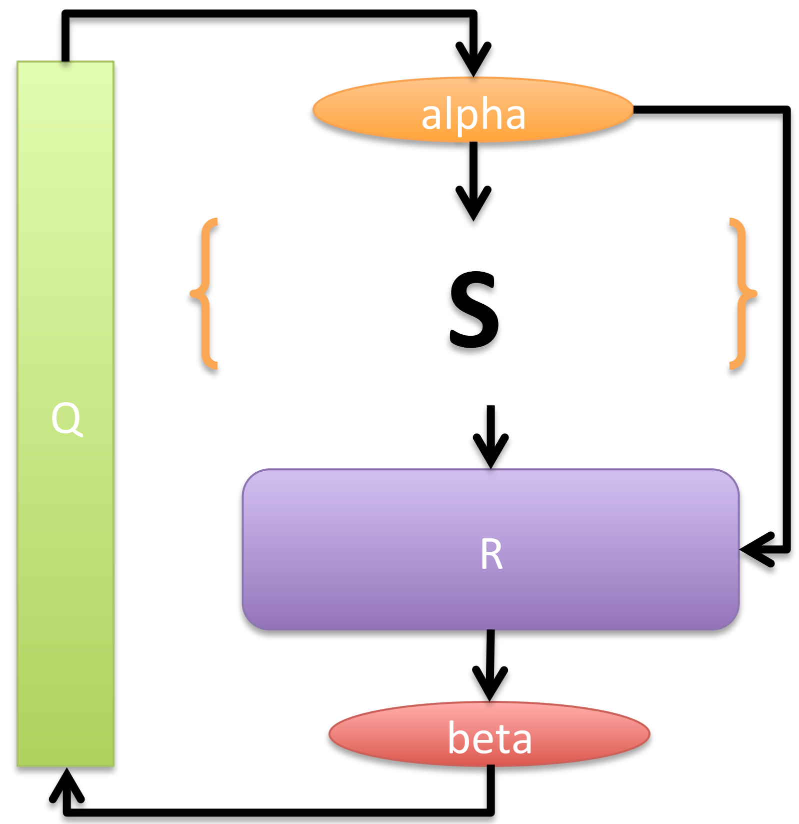
- A fragment of OWL 2 such that
  - Satisfiability checking is PTime-Complete
  - Data complexity of query answering also PTime-Complete
  - Widely used in many bio-medical ontologies
    - e.g. SNOMED CT, Gene Ontology
- Based on EL family of description logics [Baader et al. 2005]
  - supports limited class descriptions $C \sqcap D$, $\exists R.C$, A, $\top$, $\bot$
- There exist well-known effective approximate reasoning algorithms based on EL [Ren et al. 2010]
  - such as the ones implemented in the TrOWL reasoner.

Data format and data access, collection, storage, transformation

Reasoning in EL

- Can be realised by a consequence-based algorithm
  - A set of completion rules R, e.g.
    - If and , then
    - If and then
    - If , and , then
    - Etc.
  - A forward chaining mechanism FCC to apply the rules
    - Starting from the original ontology
    - Repeated apply the rules on the original and inferred axioms until no more axioms can be inferred
  - Result is the closure of the ontology R*(O), including
    - Inferred subsumptions between concept names
    - Named types of individuals
    - Named relations between individuals
- R*(O) = FCC(O,Ø,R)

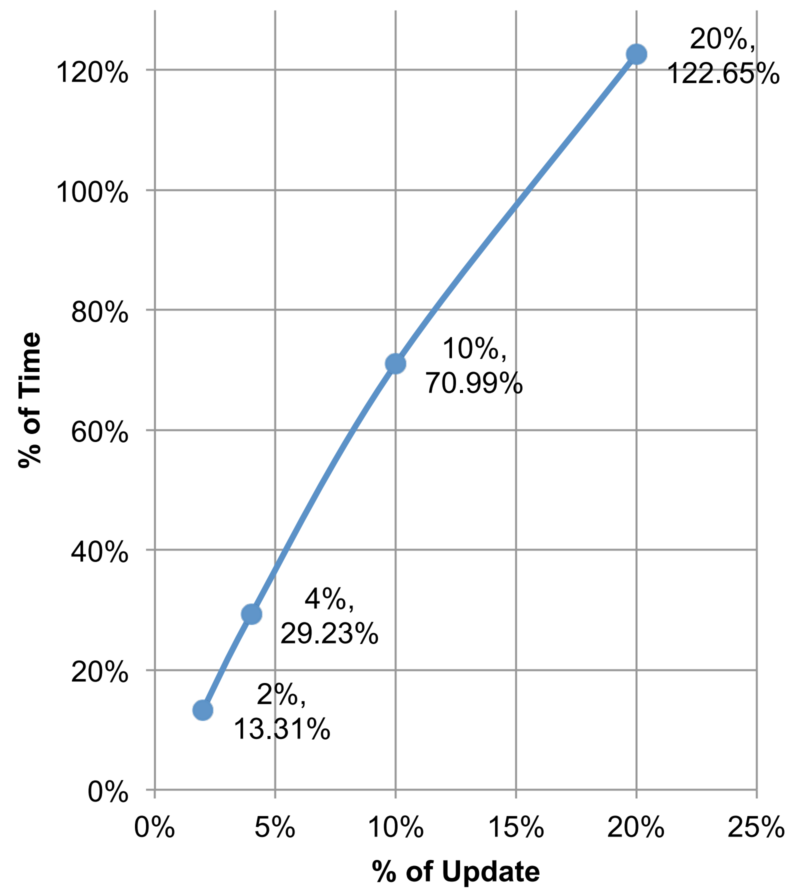
- FCC(Q,S,R):



Data format and data access, collection, storage, transformation

Incremental Reasoning in EL

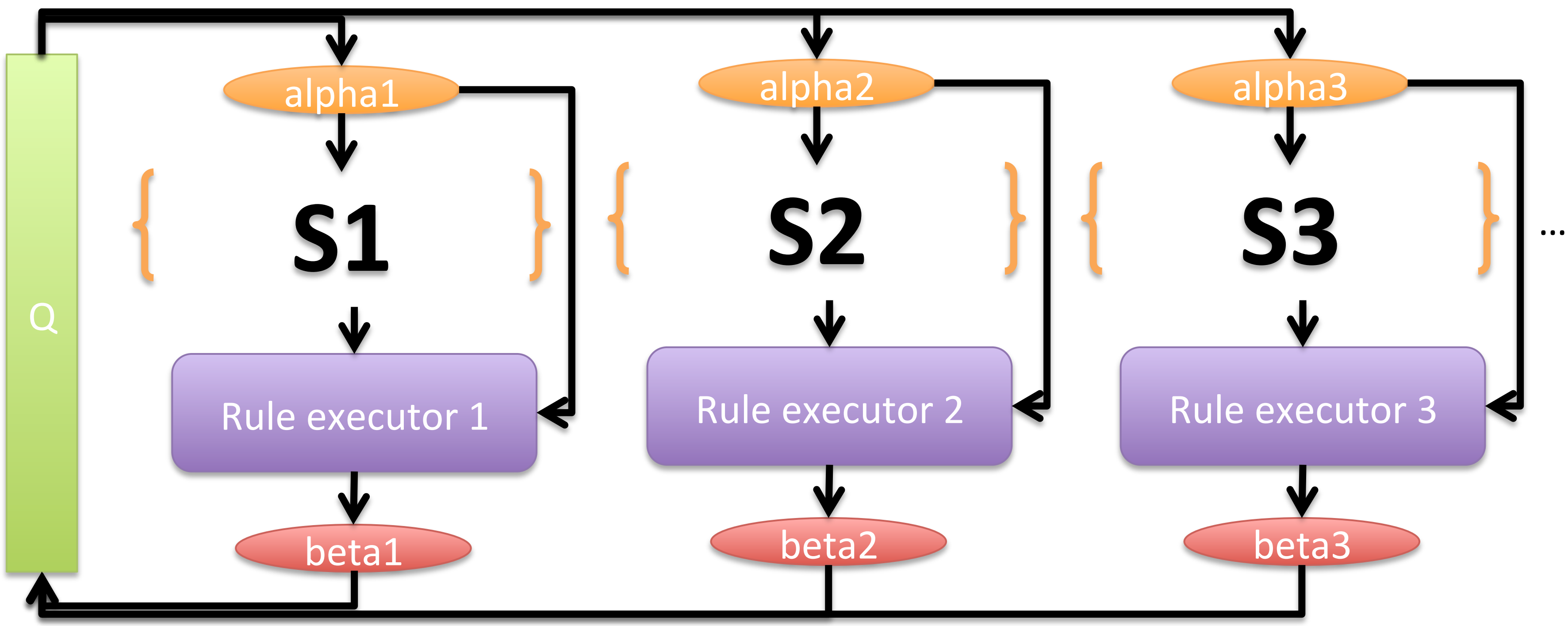
- Adding new axioms does not require complete re-computation
  - R*(O+Add) = FCC(Add, R*(O),R)
  - The original closure R*(O) can be used to compute the extended closure
- Removing small amount of original axioms can also be addressed efficiently
  - Delete-and-Rederive



Data format and data access, collection, storage, transformation

Parallel Reasoning in EL

- Multiple rule executions can be performed at the same time
- Multiple rule executors can process axioms with different context to avoid locking



Concurrent Classification of EL Ontologies, Y. Kazakov et al, ISWC2011

PART II
(Some) AI Techniques for Cities

accenture Technology Labs

IJCAI-16

Explaining Traffic Conditions with Diagnosis Reasoning

Challenge:

Logical correlation of anomalies and diagnosis in dynamic settings

Core Areas / Problems:

- Knowledge Representation and Reasoning
- Machine Learning / AI Diagnosis
- Database: Large scale data integration
- Signal Processing / Stream Reasoning

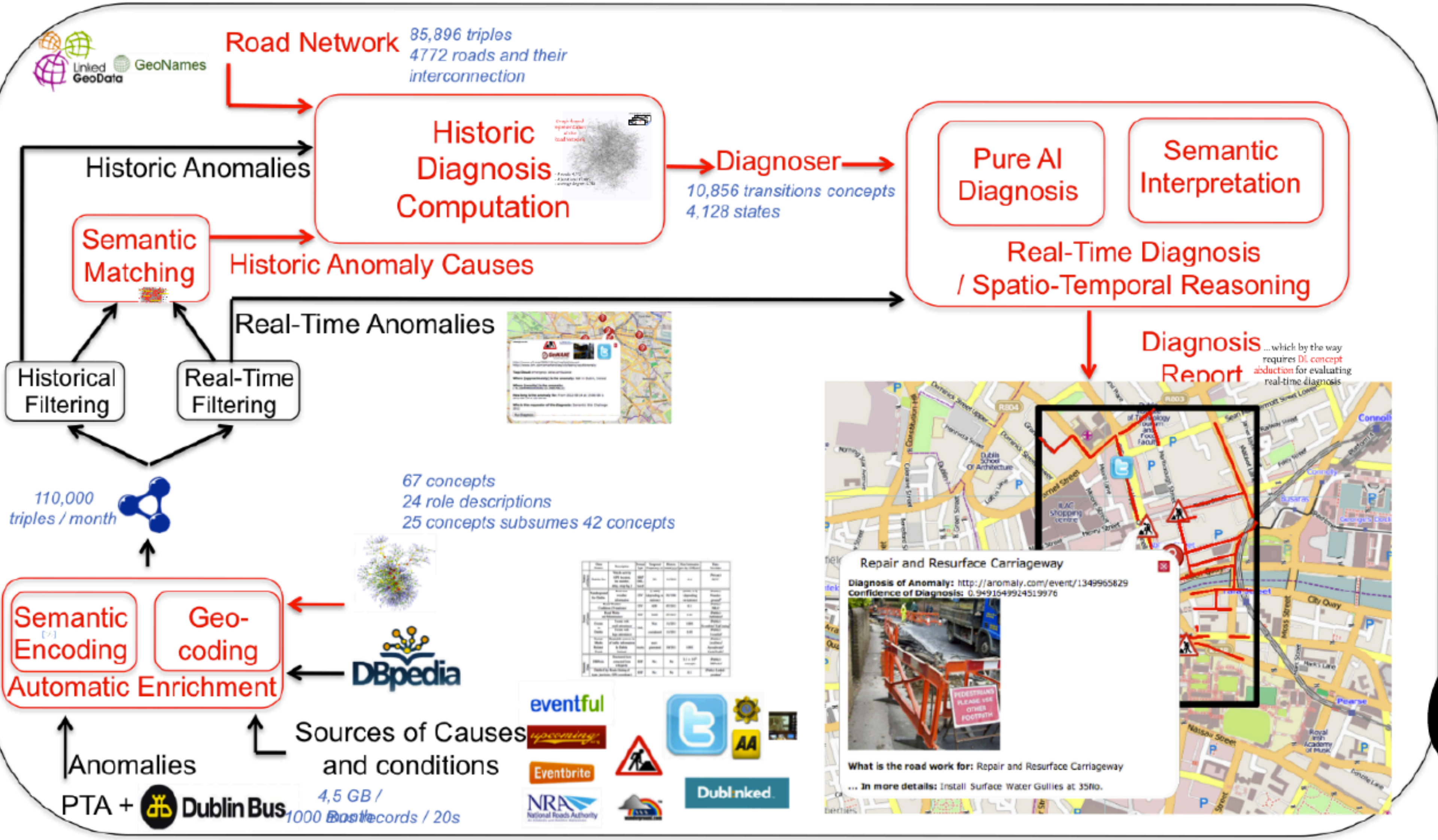
Source of Anomaly

Source of Diagnosis

Source Type	Data Source	Description	City: Dublin (Ireland)	City: Bologna (Italy)	City: Miami (USA)	City: Rio (Brazil)
Traffic Anomaly	Journey travel times across the city	Traffic Department's TRIPS system[a]	CSV format (47 routes, 732 sensors) 0.1 GB per day[b]	✗ (not available)		
Traffic Anomaly	Dublin Bus Dynamics	Vehicle activity (GPS location, line number, delay, stop flag,)	✗ (not used)	SIRI: XML format[c] (596 buses, 80KB per update 11GB per day[d])	CSV format (893 buses, 225 KB per update 43 GB per day[e])	CSV format (1, 349 buses, 181 KB per update 14 GB per day[f])
Traffic Diagnosis	Social-Media Related Feeds	Reputable sources of road traffic conditions in Dublin City	"Tweet" format - Accessed through Twitter streaming API[g]			
			Approx. 150 tweets per day[h] (approx. 0.001 GB)	✗ (not available)	Approx. 500 tweets per day[i] (approx. 0.003 GB)	✗ (not available)
	Road Works and Maintenance		PDF format (approx. 0.003 GB per day[j])	XML format (approx. 0.001 GB per day[k])	HTML format (approx. 0.001 GB per day[l])	✗ (not available)
	Social events e.g., music event, political event	Planned events with small attendance	XML format - Accessed once a day through Eventbrite[m] APIs			
			Approx. 85 events per day (0.001 GB)	Approx. 35 events per day (0.001 GB)	Approx. 285 events per day (0.005 GB)	Approx. 232 events per day (0.01 GB)
		Planned events with large attendance	XML format - Accessed once a day through Eventful[n] APIs			
			Approx. 180 events per day (0.05 GB)	Approx. 110 events per day (0.04 GB)	Approx. 425 events per day (0.1 GB)	Approx. 310 events per day (0.08 GB)
	Bus Passenger Loading / Unloading (information related to number of passenger getting in / out)		✗ (not available)	✗ (not available)	CSV format (approx. 0.8 GB per day[e])	CSV format (approx. 0.1 GB per day[e])

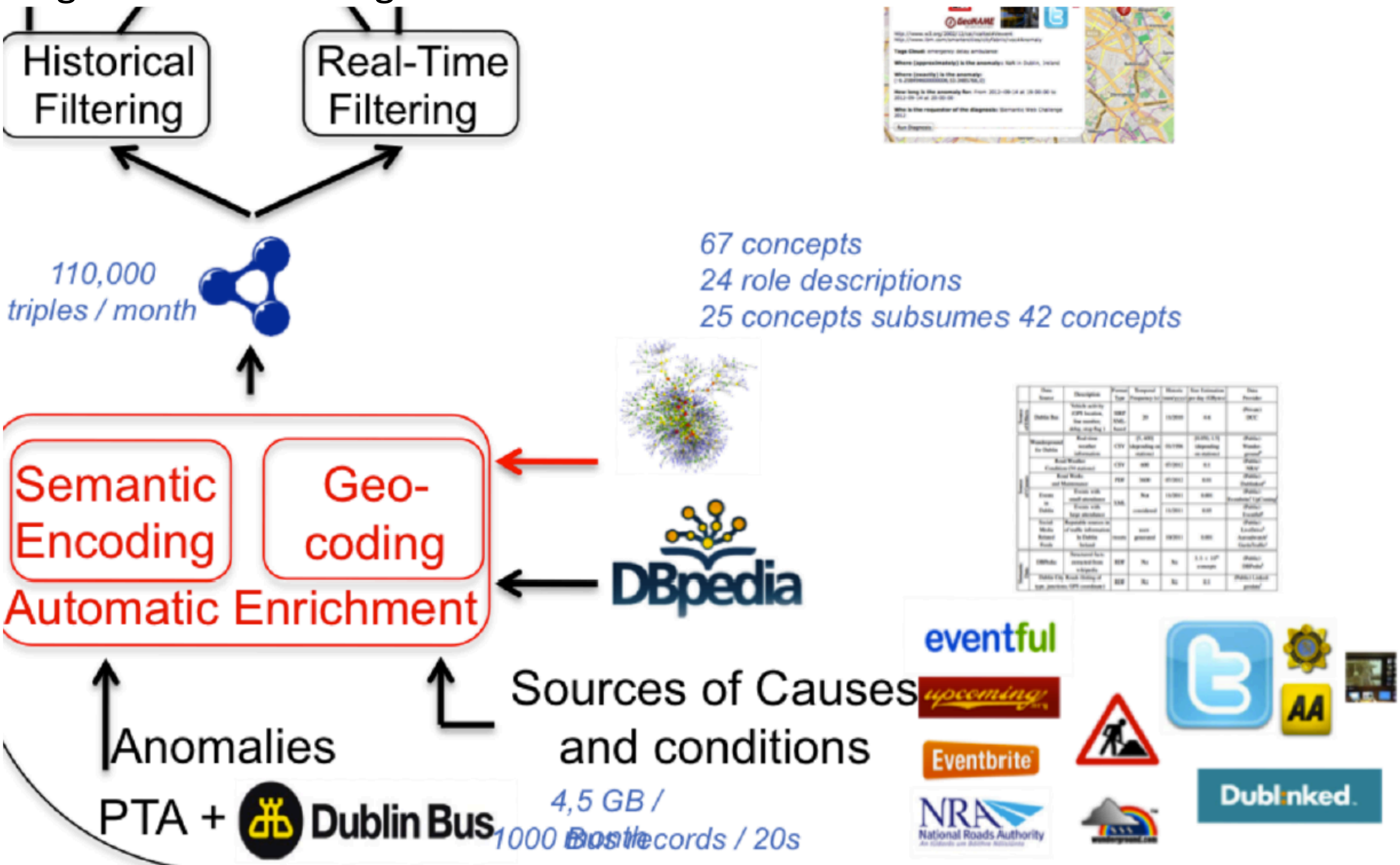
Explaining Traffic Conditions with Diagnosis Reasoning

Architecture with AI Components

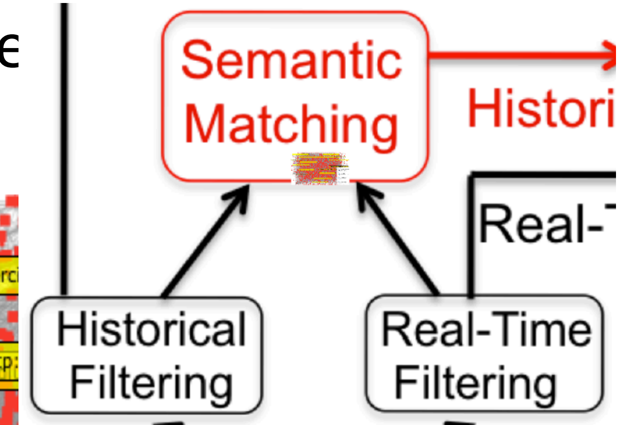


Explaining Traffic Conditions with Diagnosis Reasoning

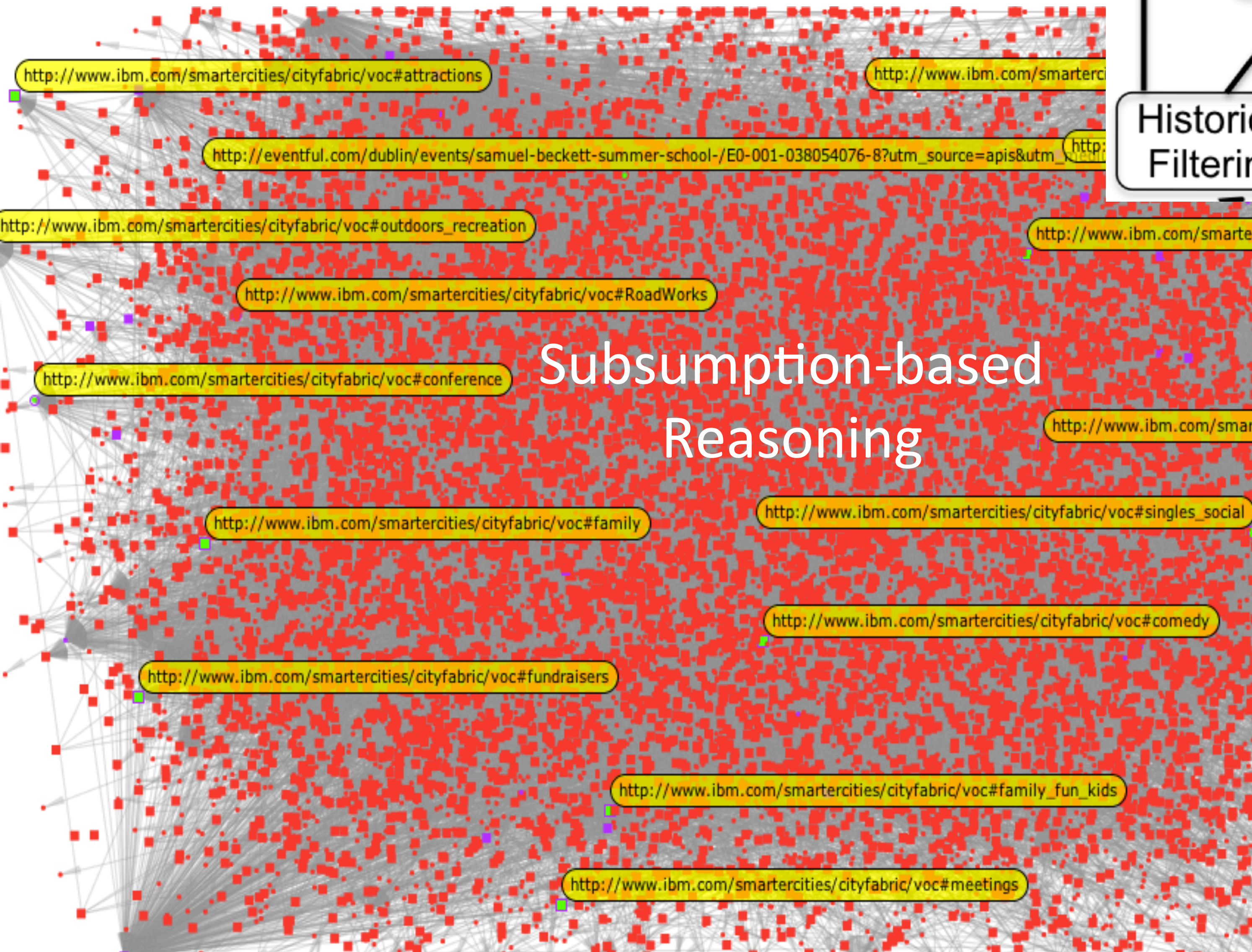
Large Scale Data Integration

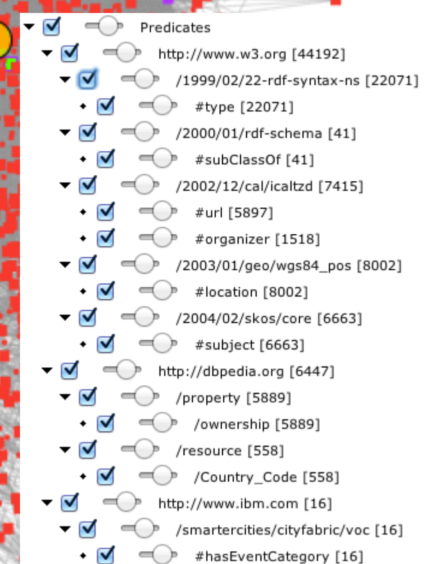


Explaining Traffic Conditions with Diagnosis Re



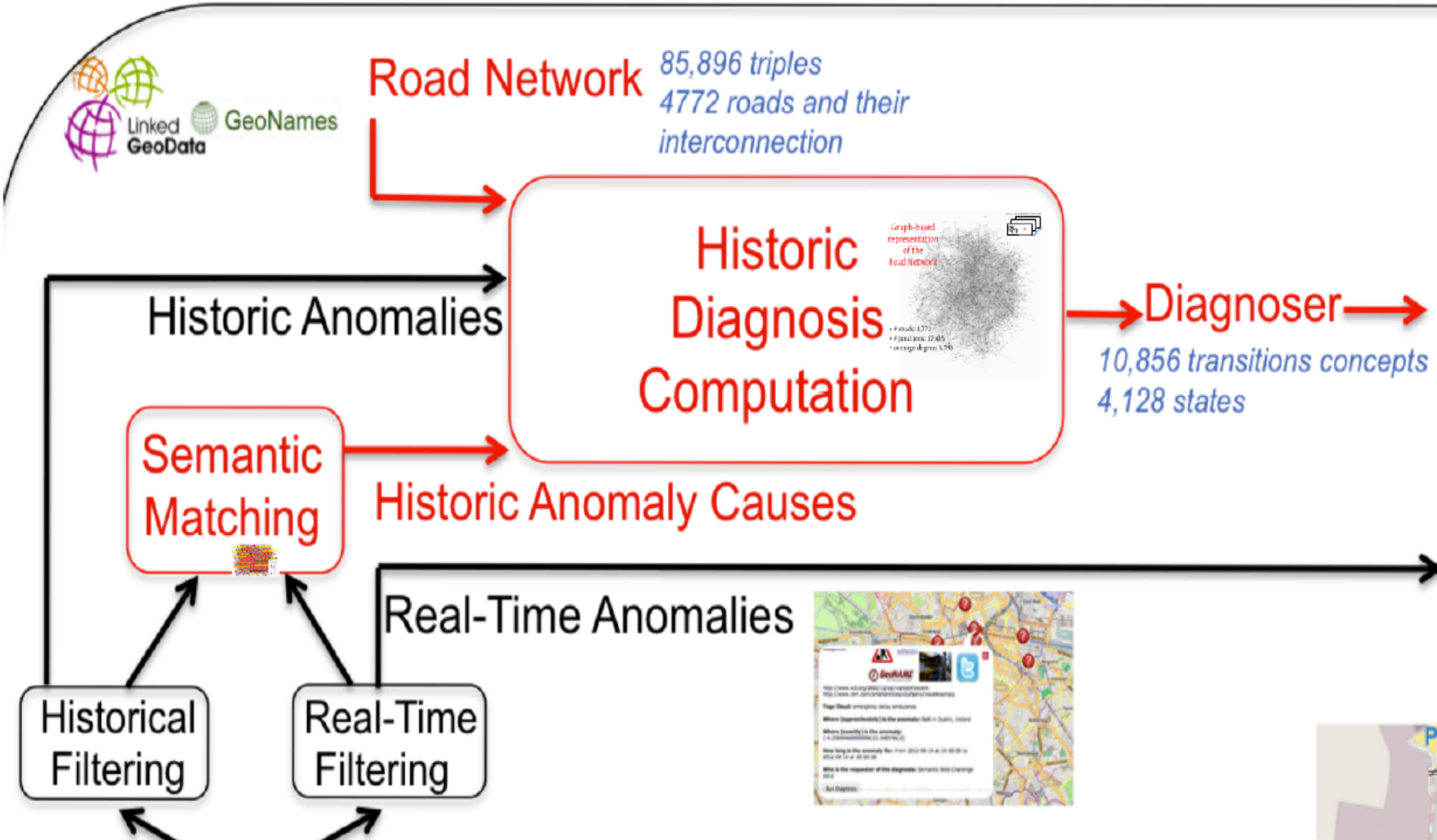
Deductive Reasoning – Semantic Matching





Explaining Traffic Conditions with Diagnosis Reasoning

Diagnosis Reasoning (1)

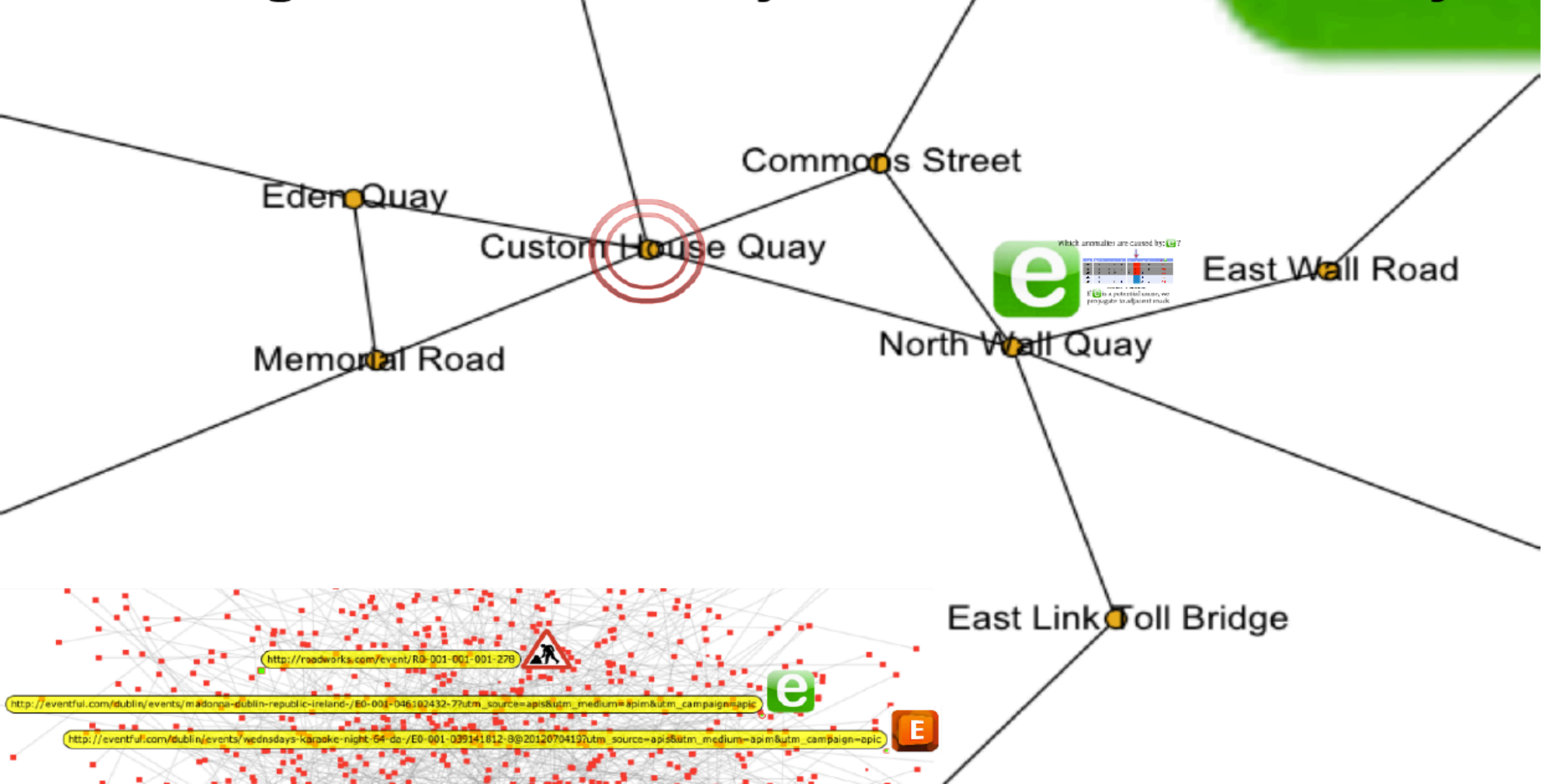


Explaining Traffic Conditions with Diagnosis Reasoning

Diagnosis Reasoning (2)

Graph-based Representation
of Road Network
4,772 roads, 17,485 junctions, 5.792: average degree

Let's diagnose one anomaly on "Custom House Quay"



Explaining Traffic Conditions with Diagnosis Reasoning

Diagnosis Reasoning (3)

Which anomalies are caused by: ?

Day	Events	r_1	r_2	r_3	r_4	r_5	r_6	r_7	sim ( )
d1	E_1	F	F	F	F	F	F	F	yes
d2	E_2	J	J	J	J	J	F	F	yes
d3	E_3	J	F	J	F	J	J	J	yes
d4	E_4	F	J	J	J	F	J	J	no
d5	E_5	F	F	F	F	F	J	J	no

F: Free Road J: Jammed Road

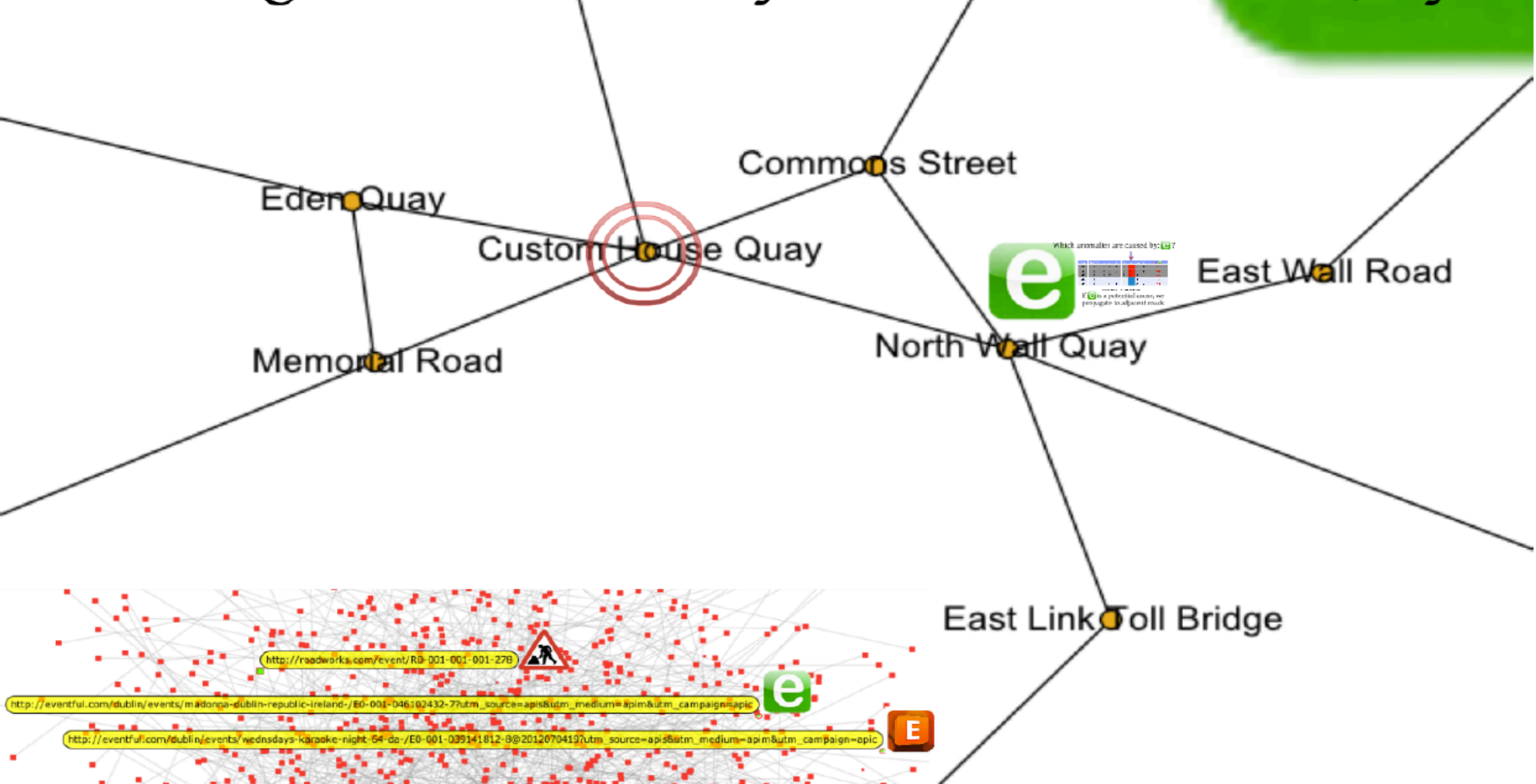
If is a potential cause, we propagate to adjacent roads

Explaining Traffic Conditions with Diagnosis Reasoning

Diagnosis Reasoning (4)

Graph-based Representation
of Road Network
4,772 roads, 17,485 junctions, 5.792: average degree

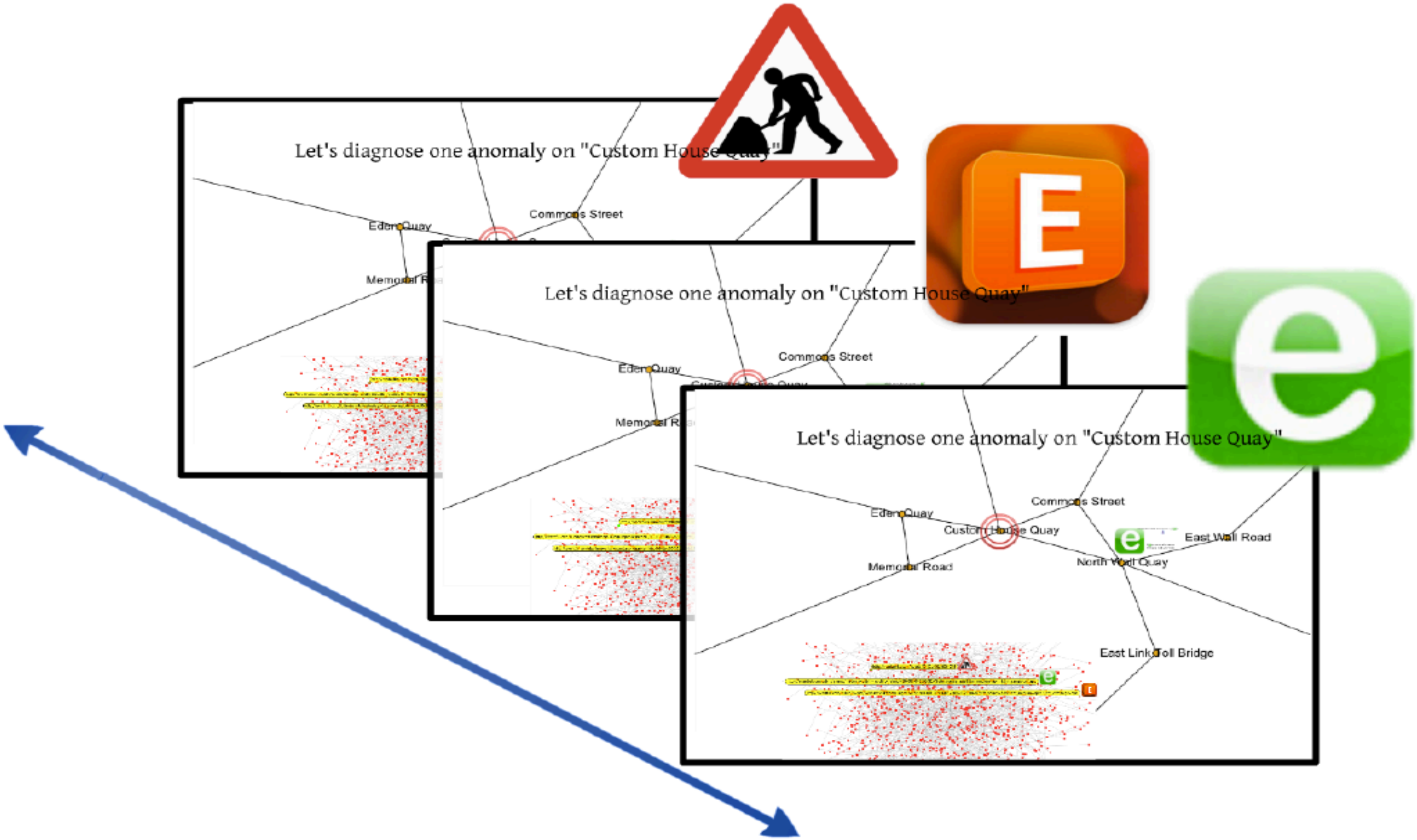
Let's diagnose one anomaly on "Custom House Quay"



Explaining Traffic Conditions with Diagnosis Reasoning

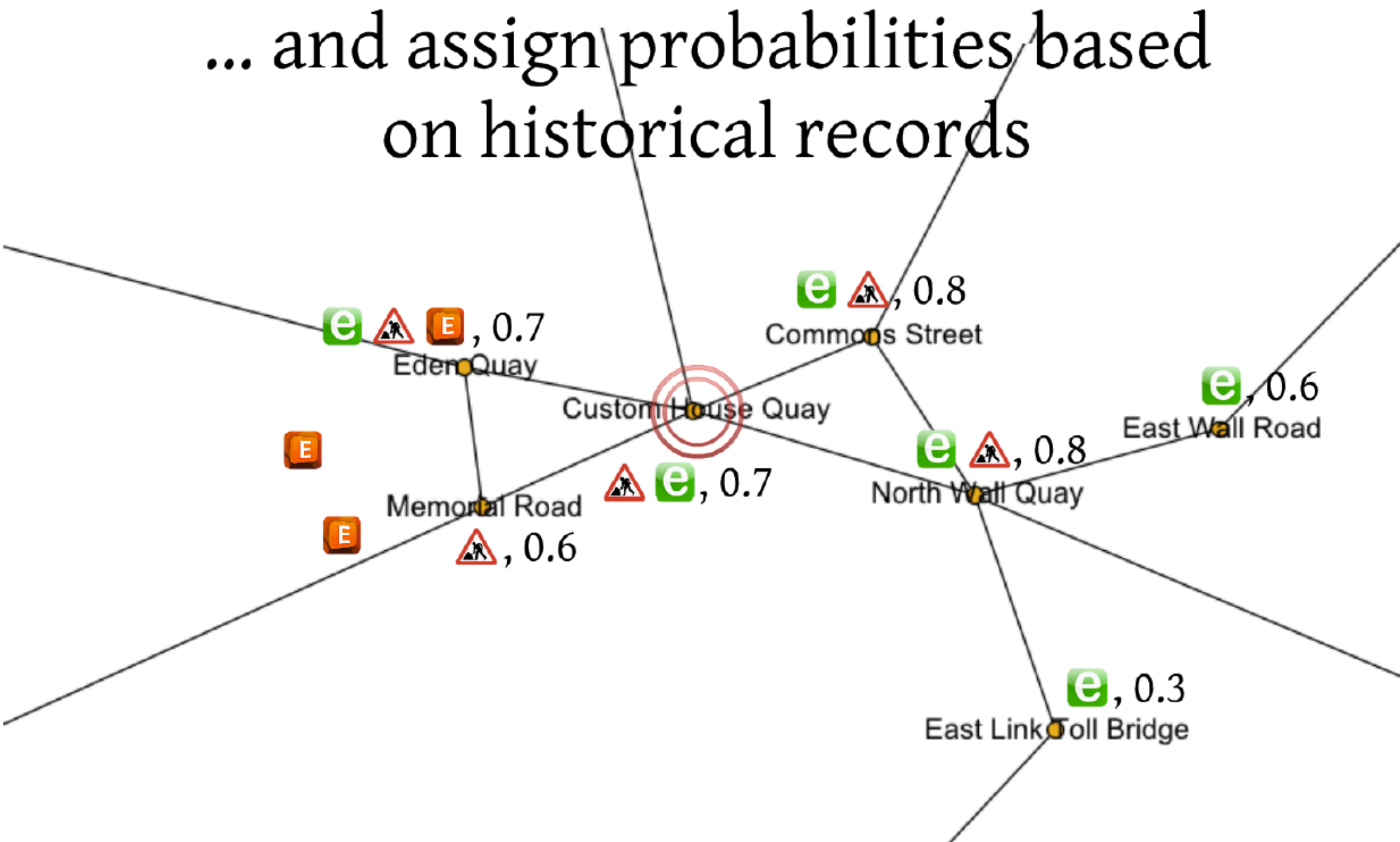
Diagnosis Reasoning (5)

Then we iterate to ALL "similar events" …



Explaining Traffic Conditions with Diagnosis Reasoning

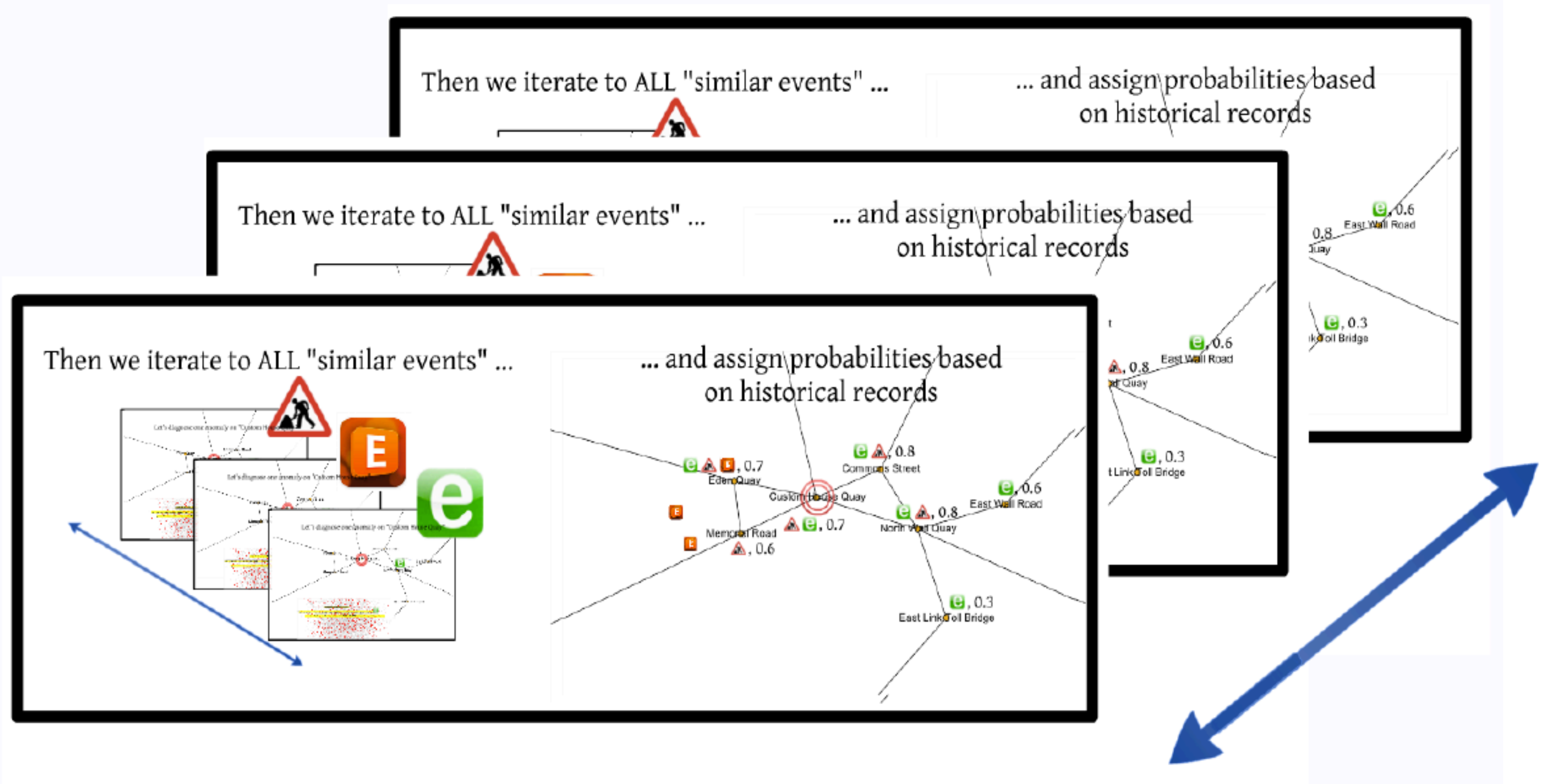
Diagnosis Reasoning (6)



Explaining Traffic Conditions with Diagnosis Reasoning

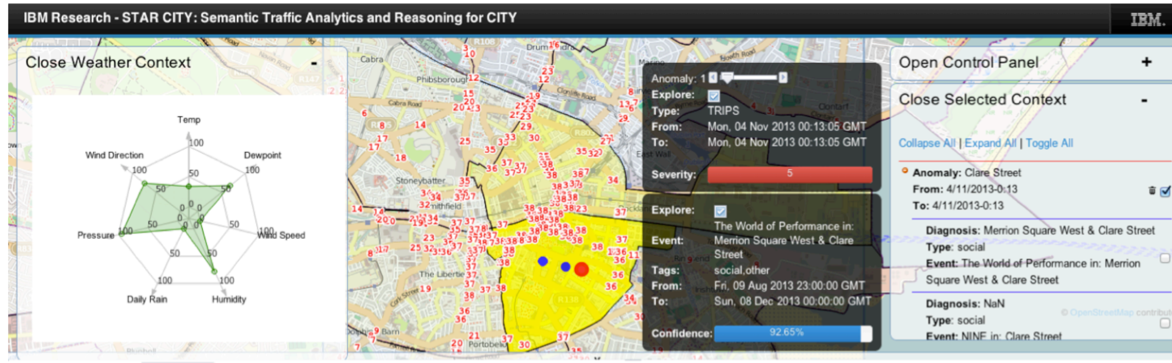
Diagnosis Reasoning (7)

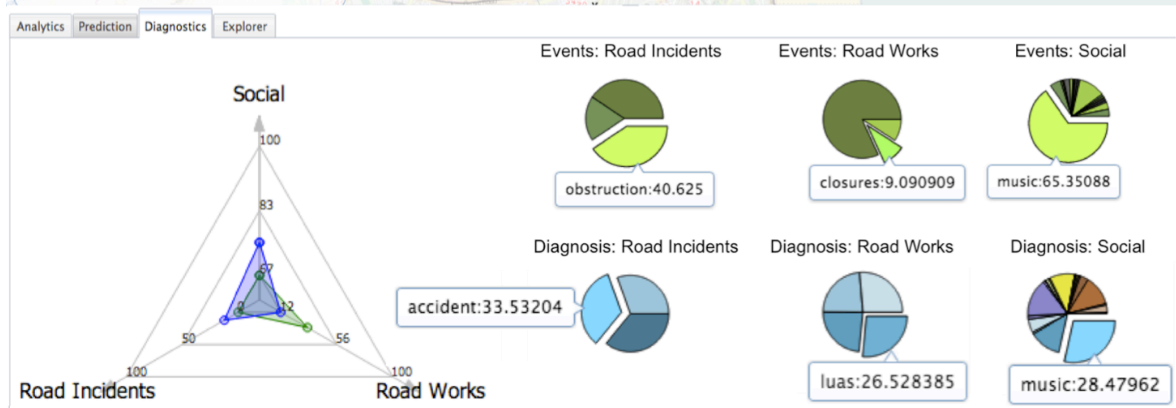
... where the time window needs also to be considered

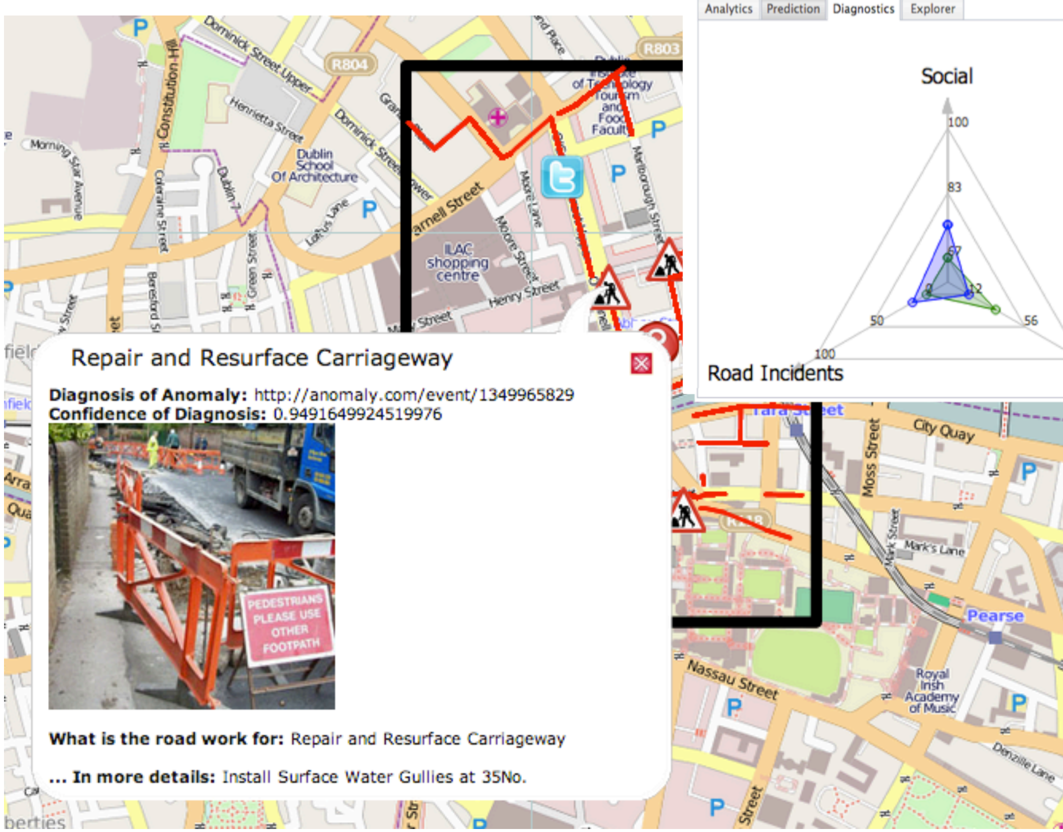


Explaining Traffic Conditions with Diagnosis Reasoning

Diagnosis Reasoning (8)







Explaining Traffic Conditions with Diagnosis Reasoning

Abductive Reasoning

Definition (Concept Abduction)

Let $\mathcal{L}$ be a DL, Out_s_i, In_s_j be two concepts in $\mathcal{L}$, and $\mathcal{T}$ be a set of axioms in $\mathcal{L}$. A Concept Abduction Problem (CAP), denoted as $\langle \mathcal{L}, Out\_s_i, In\_s_j, \mathcal{T} \rangle$ is finding a concept $H \in \mathcal{L}$ such that $\mathcal{T} \models Out\_s_i \sqcap H \sqsubseteq In\_s_j$.

Identification of what is underspecified in Out_si to completely satisfy In_sj in T

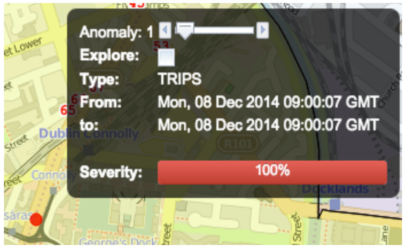
ABDUCTIVE REASONING
REPORTS HOW DIAGNOSIS IS APPROXIMATED
WITH SEMANTIC MATCHING

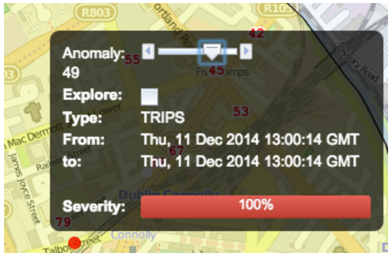
Different Cause / Event

Different Congestion

Different Time

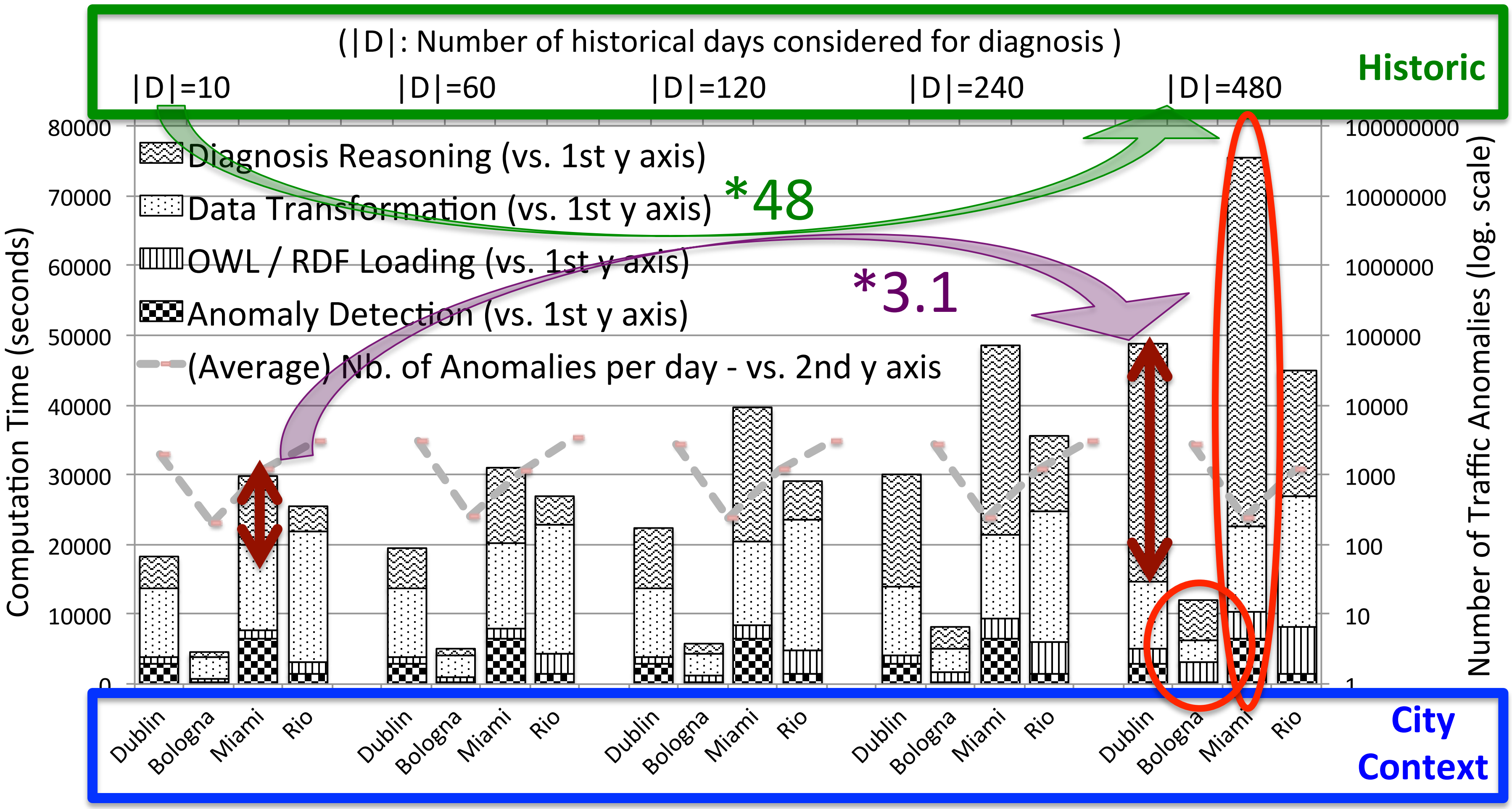
Different Location





Explaining Traffic Conditions with Diagnosis Reasoning

Experimentation: Evaluating the impact of data (size, historic, heterogeneity) on scalability

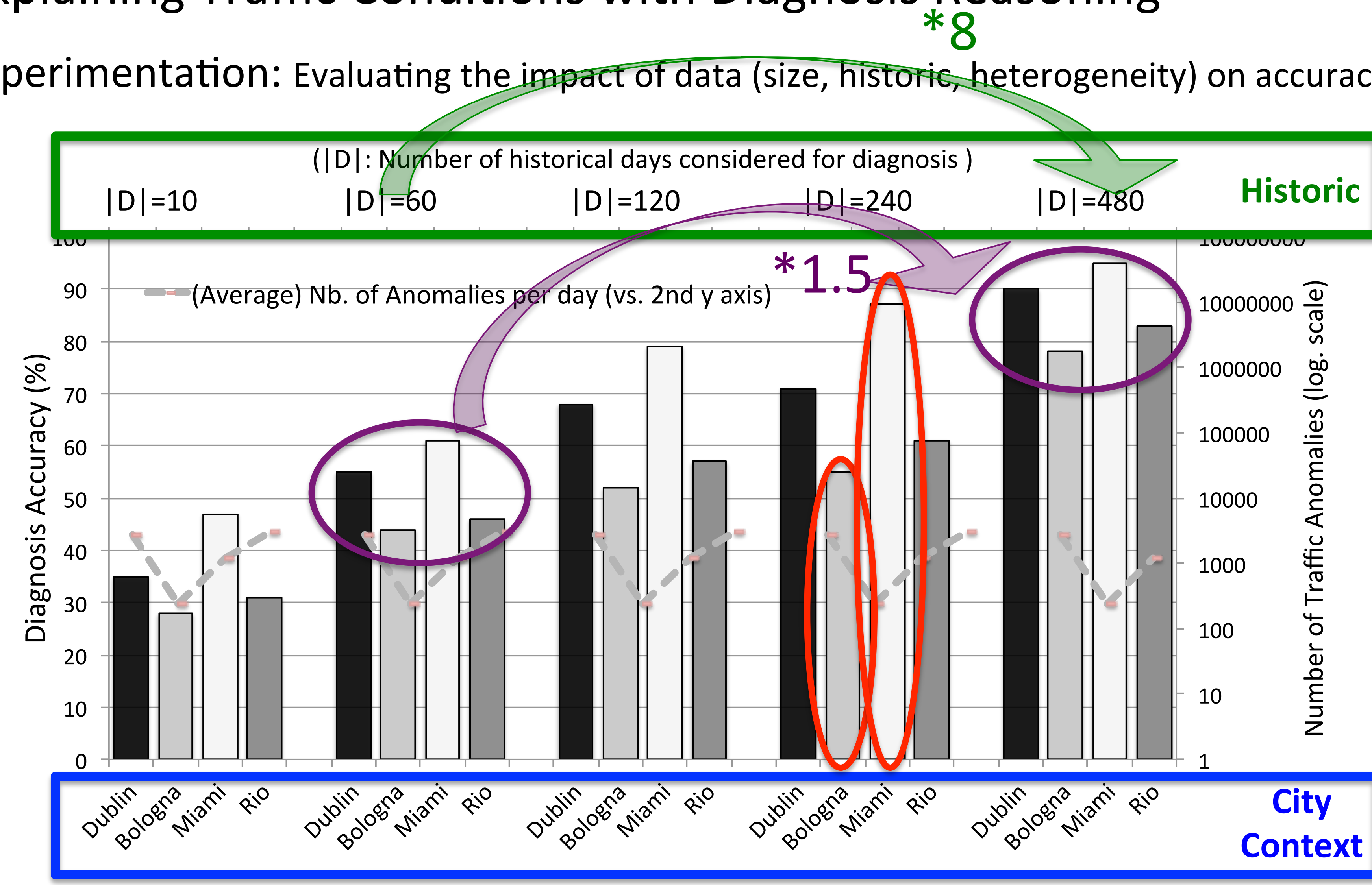


Lesson Learnt: + data sets + overhead on transformation, loading, and reasoning

+ historic data + overhead on reasoning

Explaining Traffic Conditions with Diagnosis Reasoning

Experimentation: Evaluating the impact of data (size, historic, heterogeneity) on accuracy

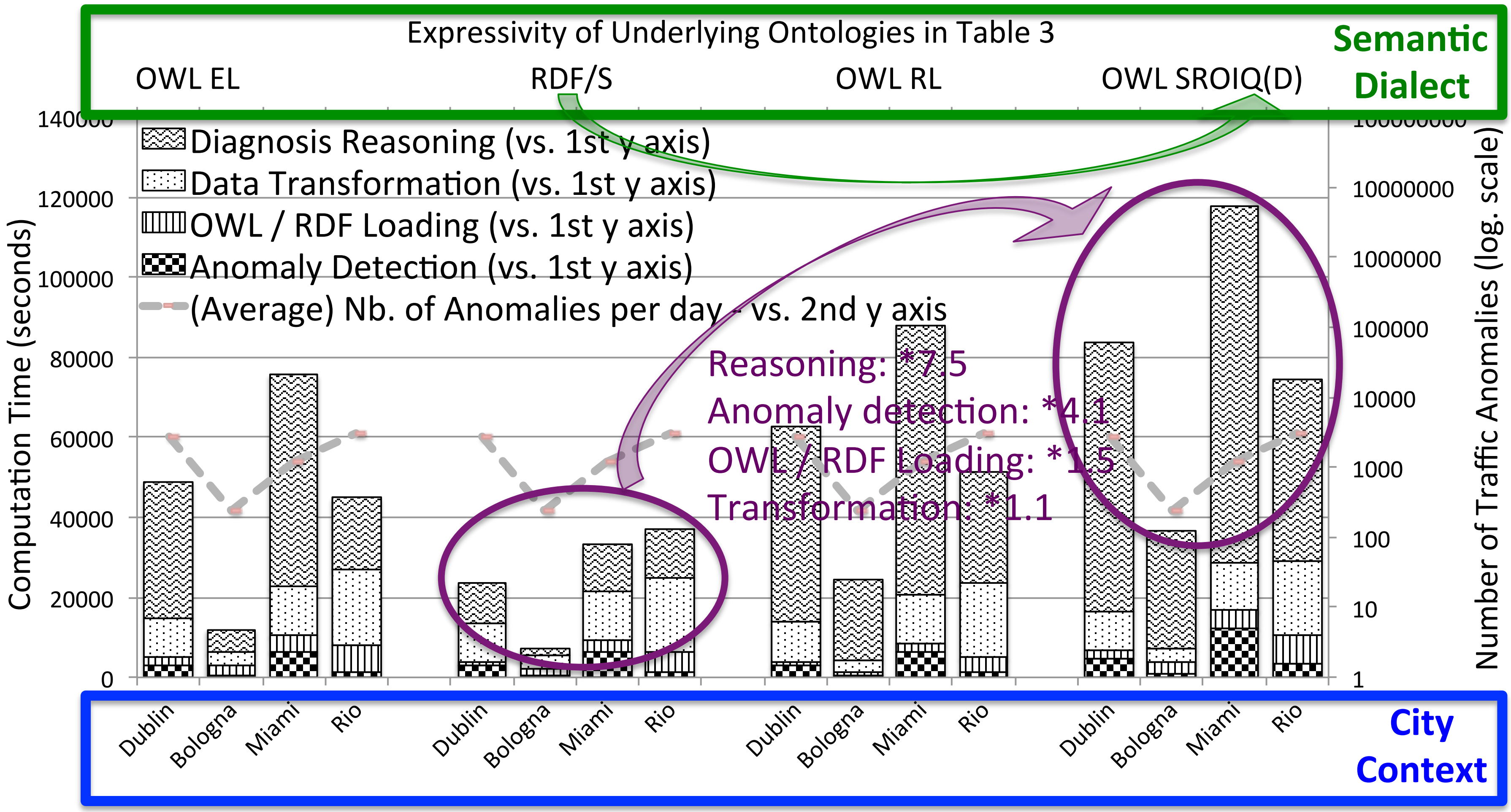


Lesson Learnt: + data sets + accurate diagnosis results

- historic data – accurate diagnosis results

Explaining Traffic Conditions with Diagnosis Reasoning

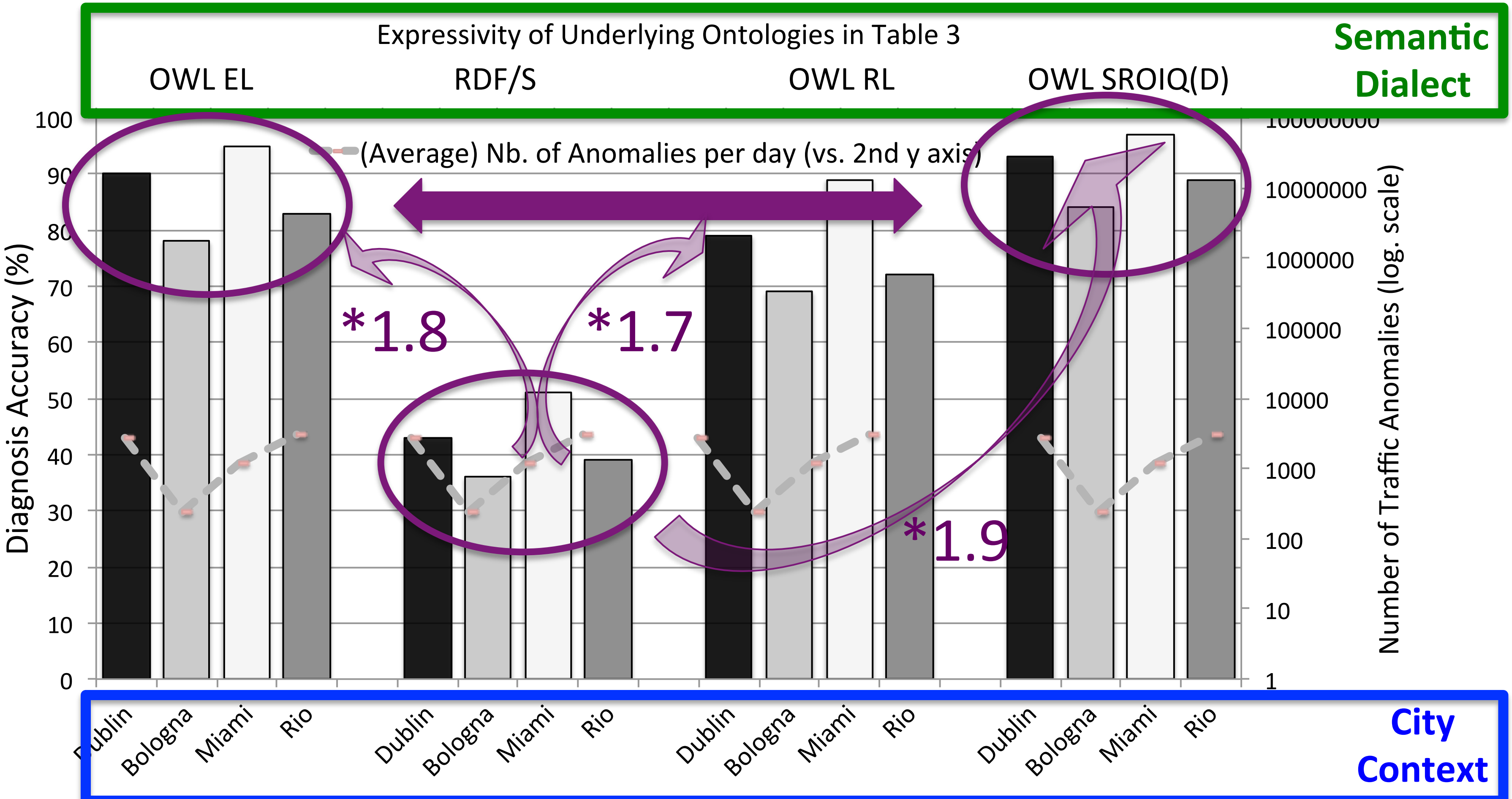
Experimentation: Evaluating the impact of expressivity on scalability



Lesson Learnt: + expressive - scalable accenture Technology Labs IJCAI-16

Explaining Traffic Conditions with Diagnosis Reasoning

Experimentation: Evaluating the impact of expressivity on accuracy

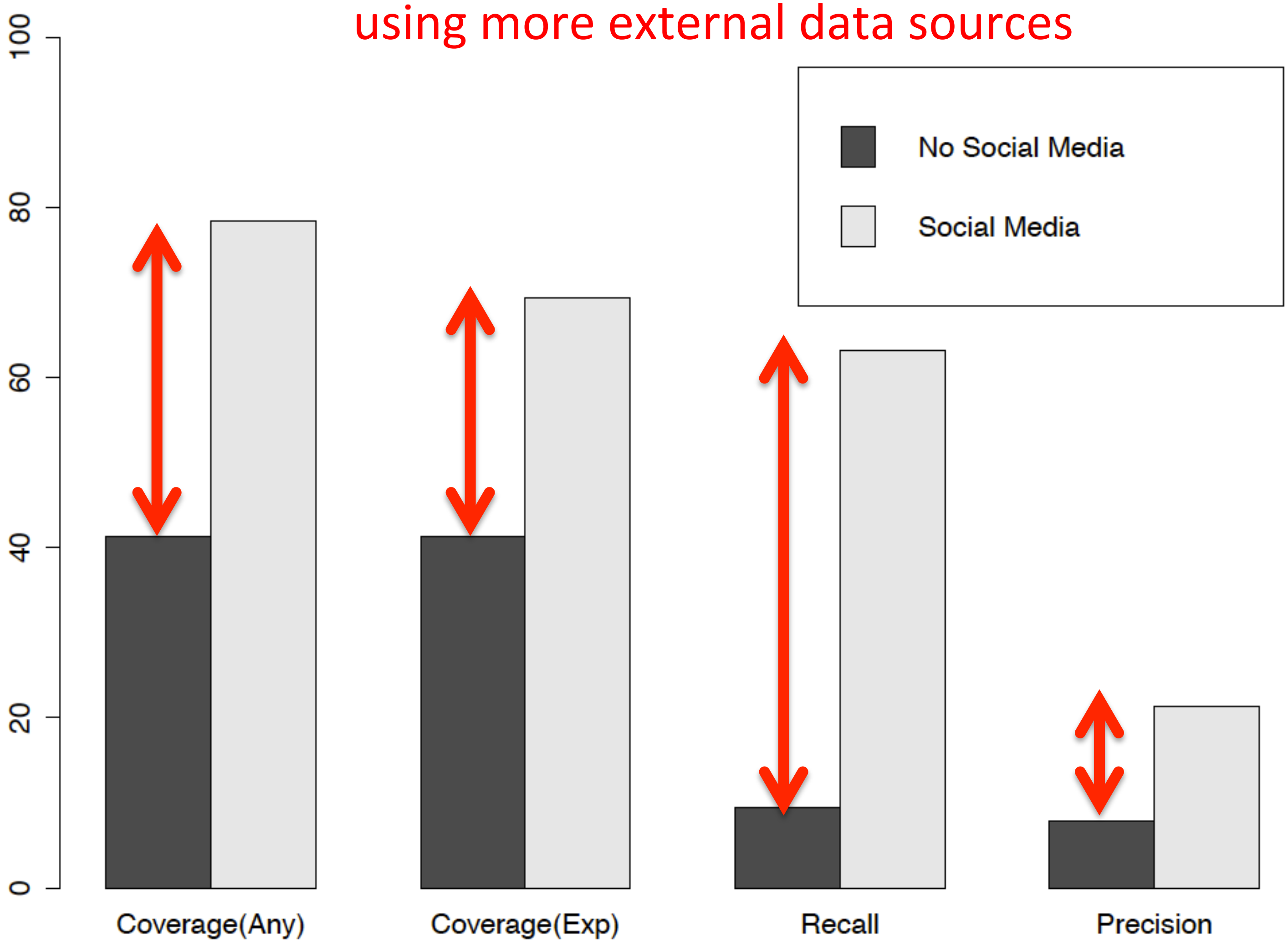


Lesson Learnt: + expressivity +/- accurate diagnosis results

Explaining Traffic Conditions with Diagnosis Reasoning

Experimentation: Evaluating the impact of data on accuracy

Improvement of coverage / precision / recall
using more external data sources



Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning (1)

Schema

Consistent?

Data

Update

Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning (2)

Heavyweight Reasoning

Schema

Static
Small

Consistent?

Data

Dynamic (stream)
Big

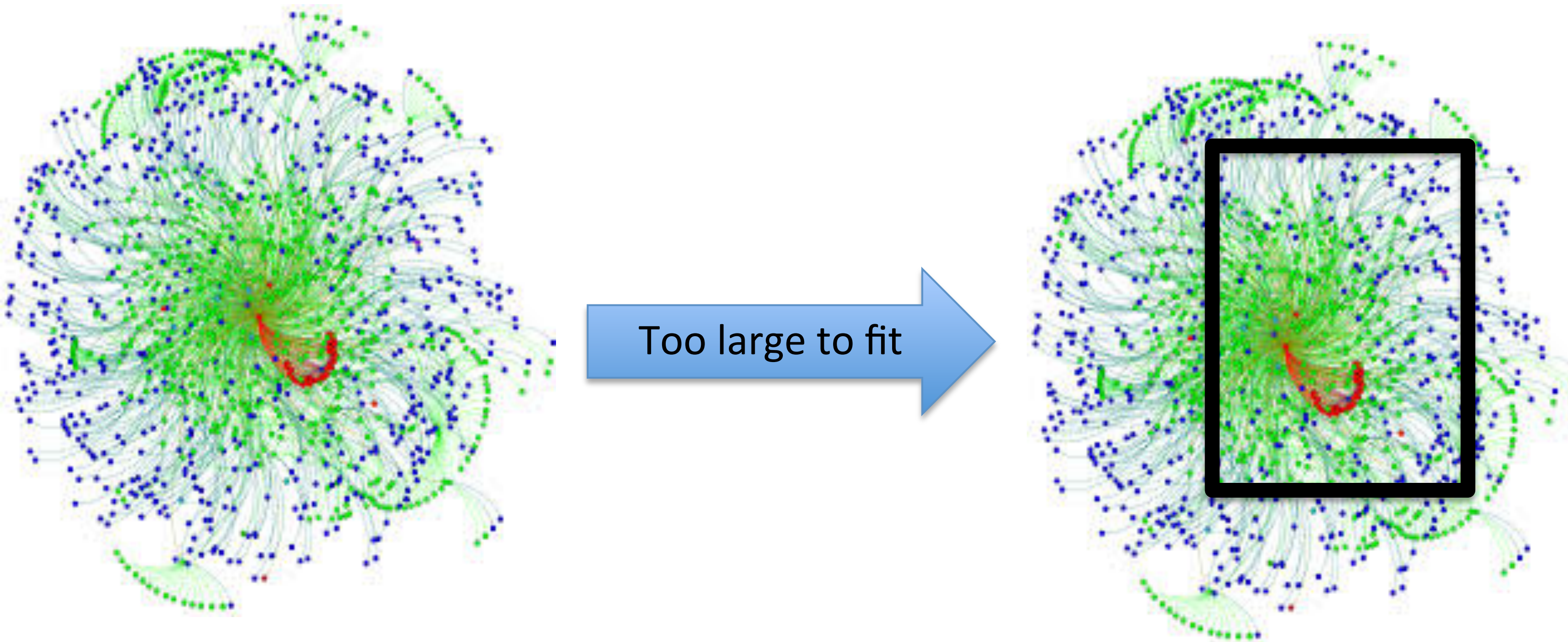
Update

Syntactic

Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning (3)

How to maintain ontology consistency more efficiently in this context?

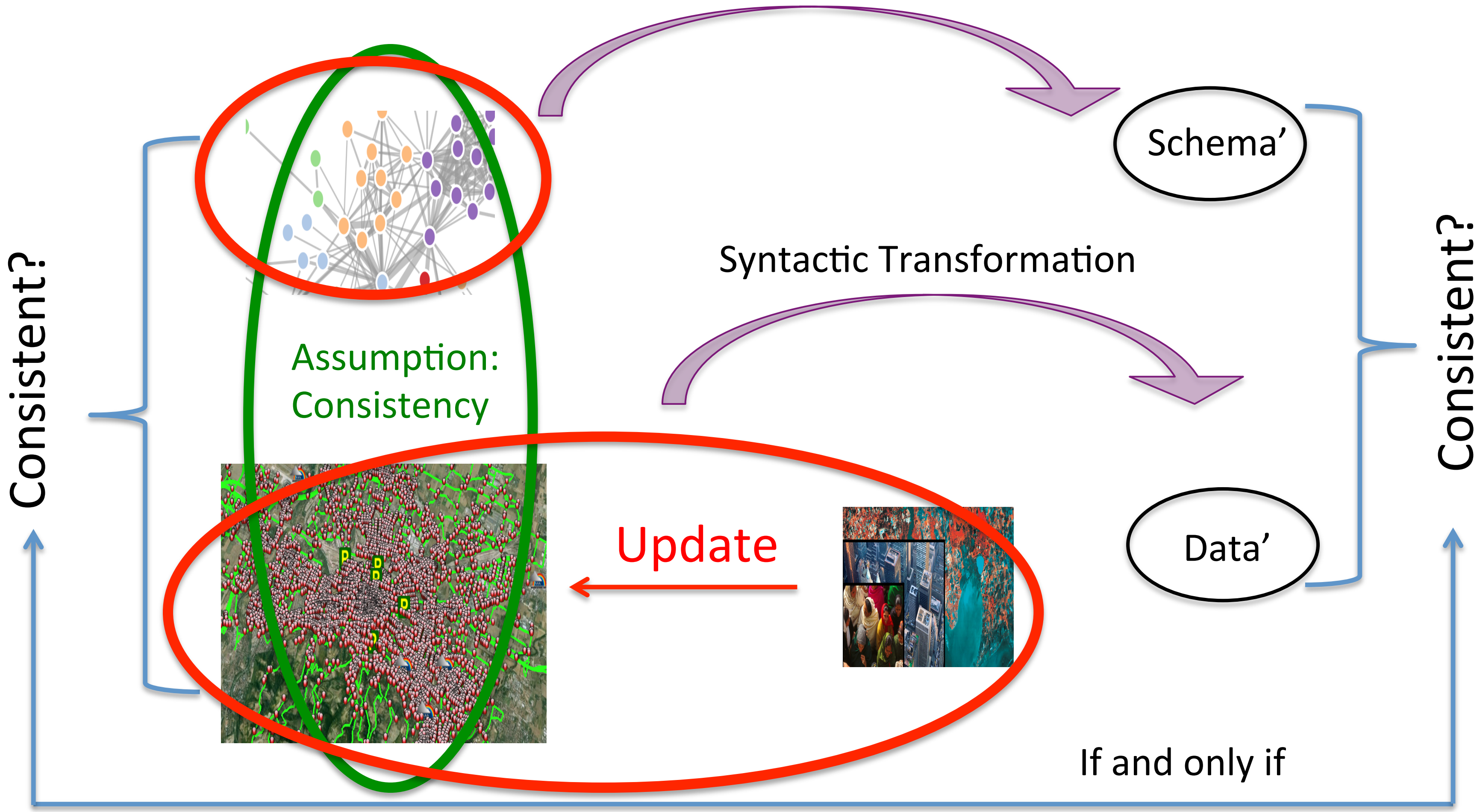


Existing incremental reasoning: maintain an in-memory structure

Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning (4)

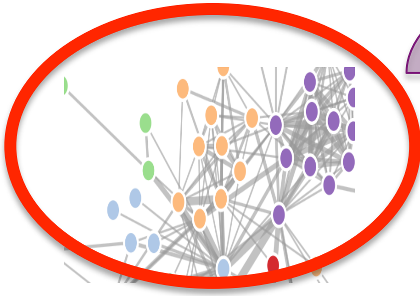
Our solution: extracting a subset of data + updates **syntactically**



Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning (5)

Our solution: extracting a subset of data + updates **syntactically**



Schema'

Schema simplification: to expose constraints that can possibly cause interaction between instances via role assertions.

$$L_1 \sqsubseteq \forall S.L_2,$$

$$L_1 \sqsubseteq \exists^{\leq n} S_p.L_2,$$

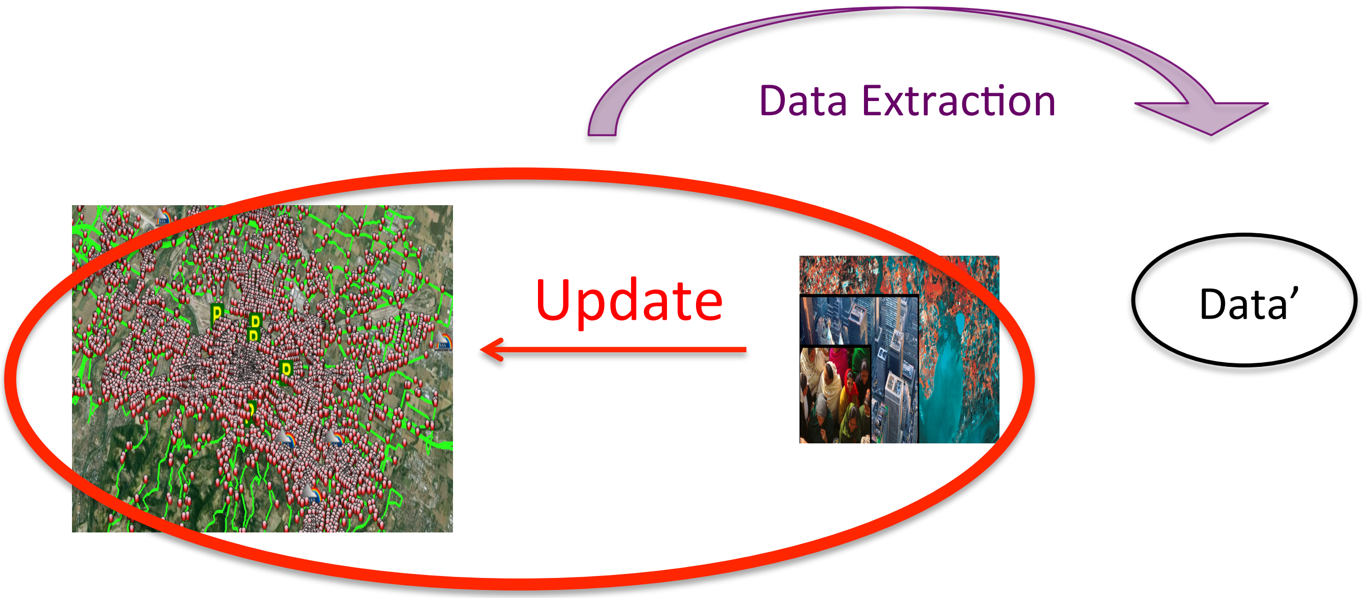
Where transitive roles are removed and all axioms with universal and maximum number restrictions are simplified on the RHS

- The simplified schema and the original schema are equi-satisfiable
- The simplification, a one-time process, can be done offline

Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning (6)

Our solution: extracting a subset of data + updates **syntactically**



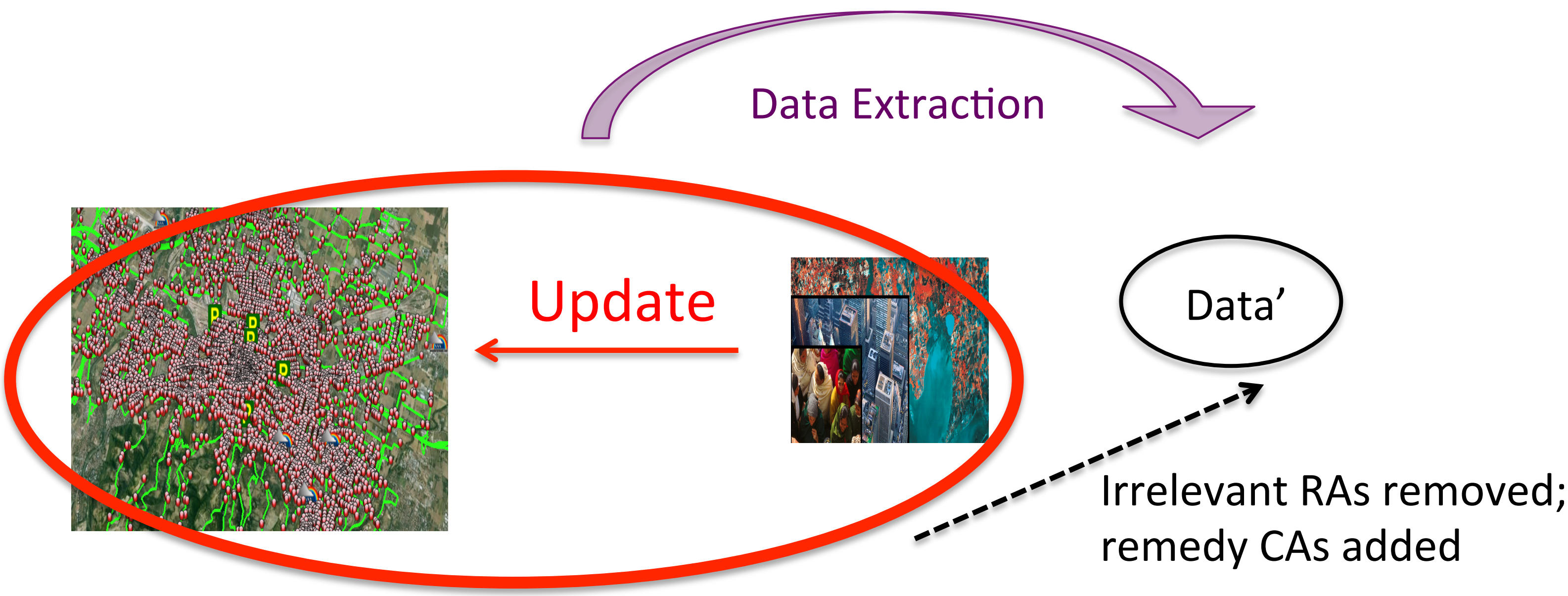
Key observation:

Role assertions of the form S(a,b) may cause interaction between instances, thus, they need be eliminated as much as possible for subset extraction

Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning (7)

Our solution: extracting a subset of data + updates **syntactically**



For any S(a,b), it is irrelevant for data extraction if, for every axiom $L_1 \sqsubseteq \forall S.L_2$ it is already consistent with this axiom, e.g., two sufficient conditions:

- b:L2 is in the data
- a:L1 is in the data (due to our consistency assumption)

Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning – Example

Schema: { }

Data: { IBM:COMP, *worksFor*(a, IBM),
a: age=38, a: dept="swg", b:EMP, c:COMP, … }

Inserts: { a:EMP }

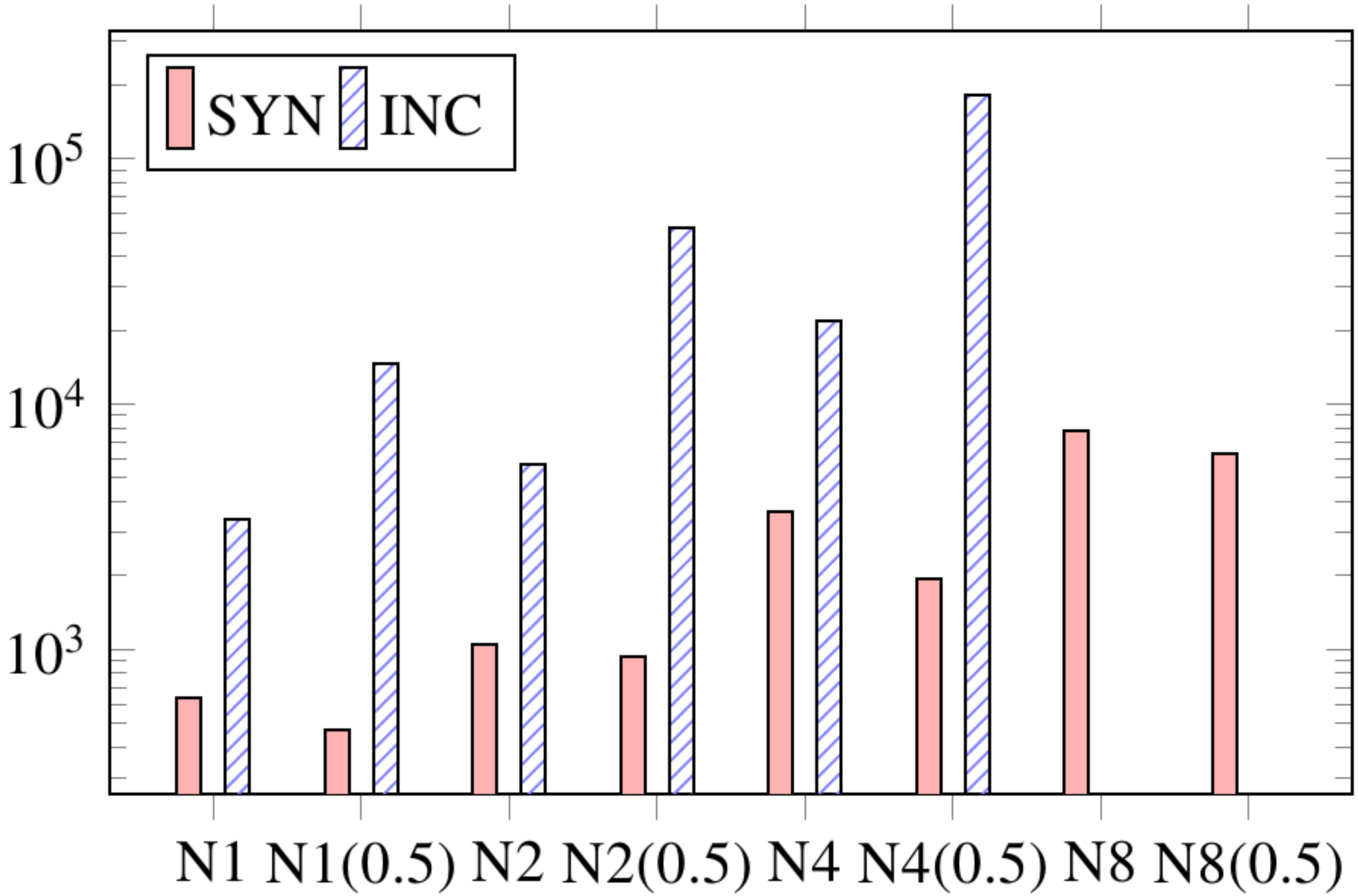
Steps to compute A_{add}:

- { a:EMP }
- { a:EMP, IBM:COMP } because *worksFor*(a, IBM) is irrelevant
- { a:EMP, IBM:COMP } is consistent with the schema

Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning – Experimentation

Figure 3: NPD ontologies with time (ms) in logarithmic scale.



Explaining Traffic Conditions with Diagnosis Reasoning

Optimization techniques to scale up reasoning – Take Away Notes

The Approach:

- Syntactic → quadratic in the size of the data
- Incremental → suits large datasets
- Reasoner independent → easy to implement

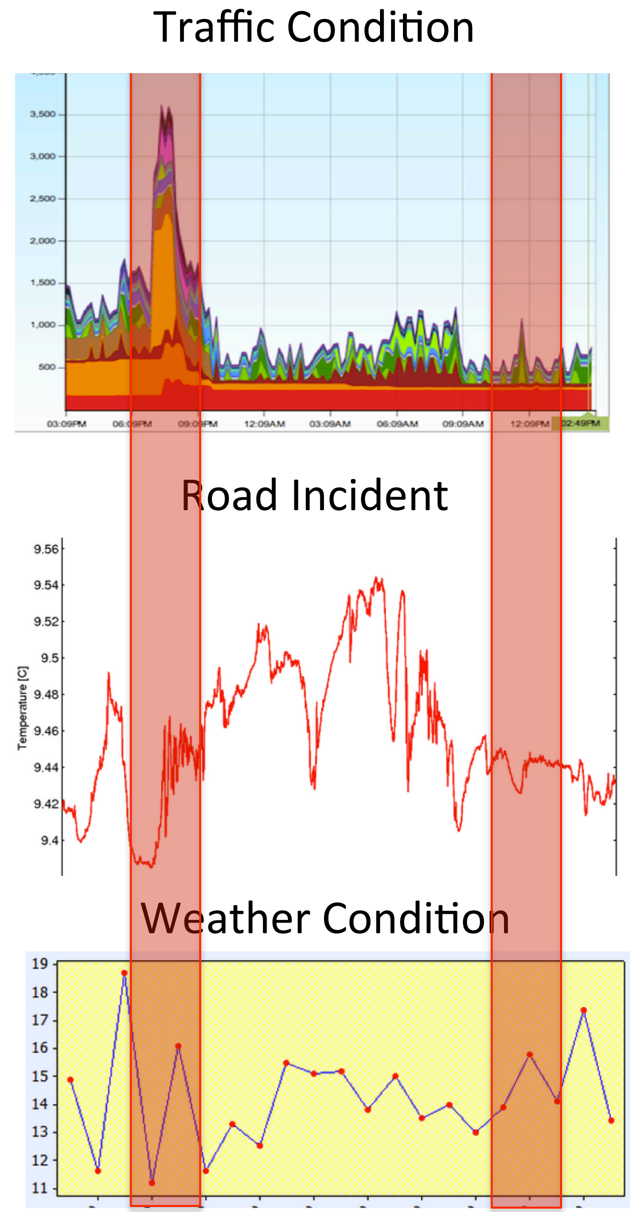
When is this approach preferred over Pellet (or similar native incremental reasoning support)?

- If the instance objects are not highly correlated via role assertions that have roles used by restrictions in the schema
- If the size of inserts is large, e.g., half of the original data
- If the data + inserts cannot be fully loaded into memory

Forecasting Traffic Congestion with Predictive Reasoning

Challenge:

Predictive **reasoning (as opposed to analytics)** in heterogeneous and dynamic settings

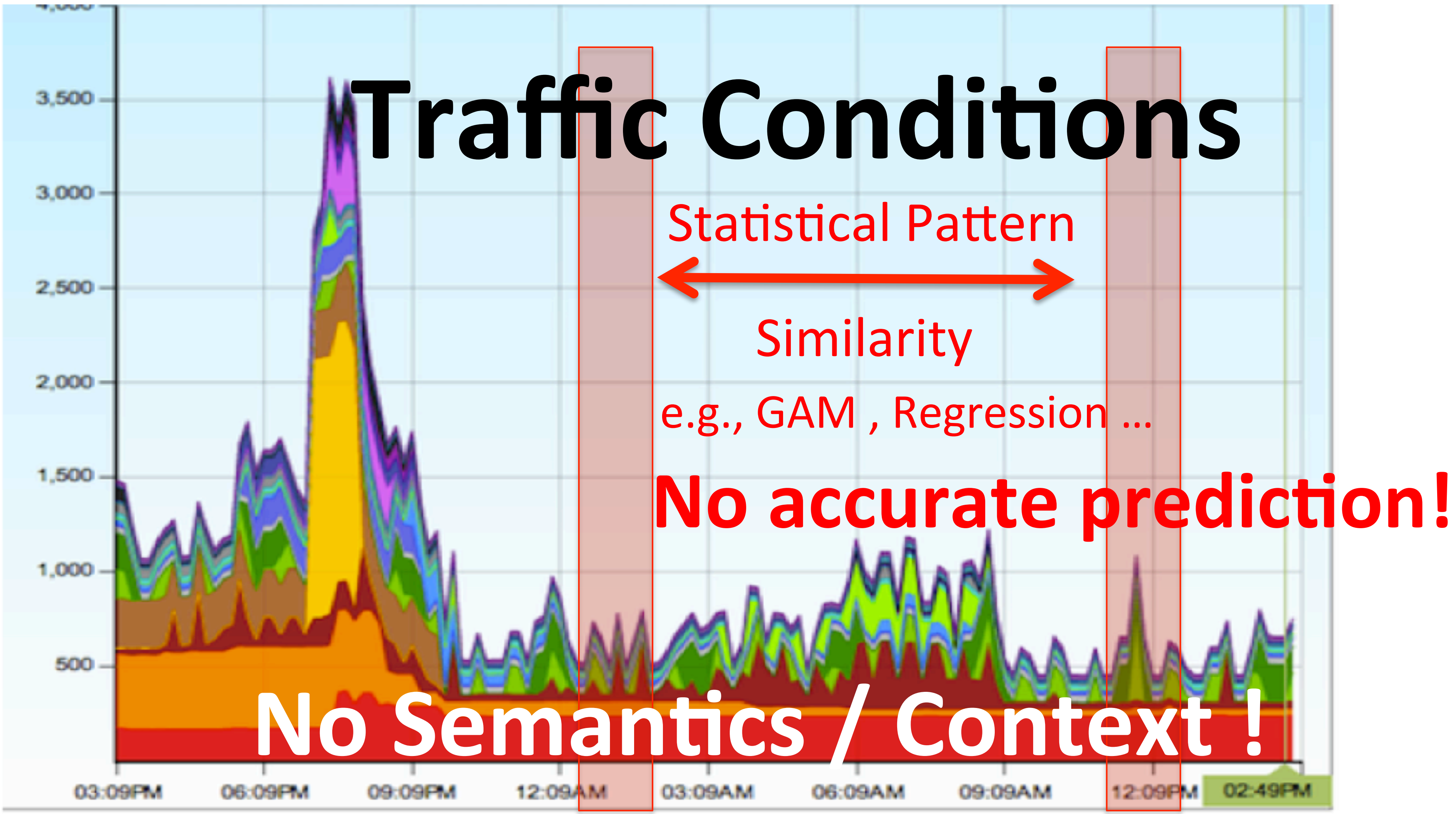


Core Areas / Problems:

- Knowledge Representation and Reasoning
- Machine Learning / Knowledge Discovery
- Database: Large scale data integration
- Signal Processing / Stream Reasoning

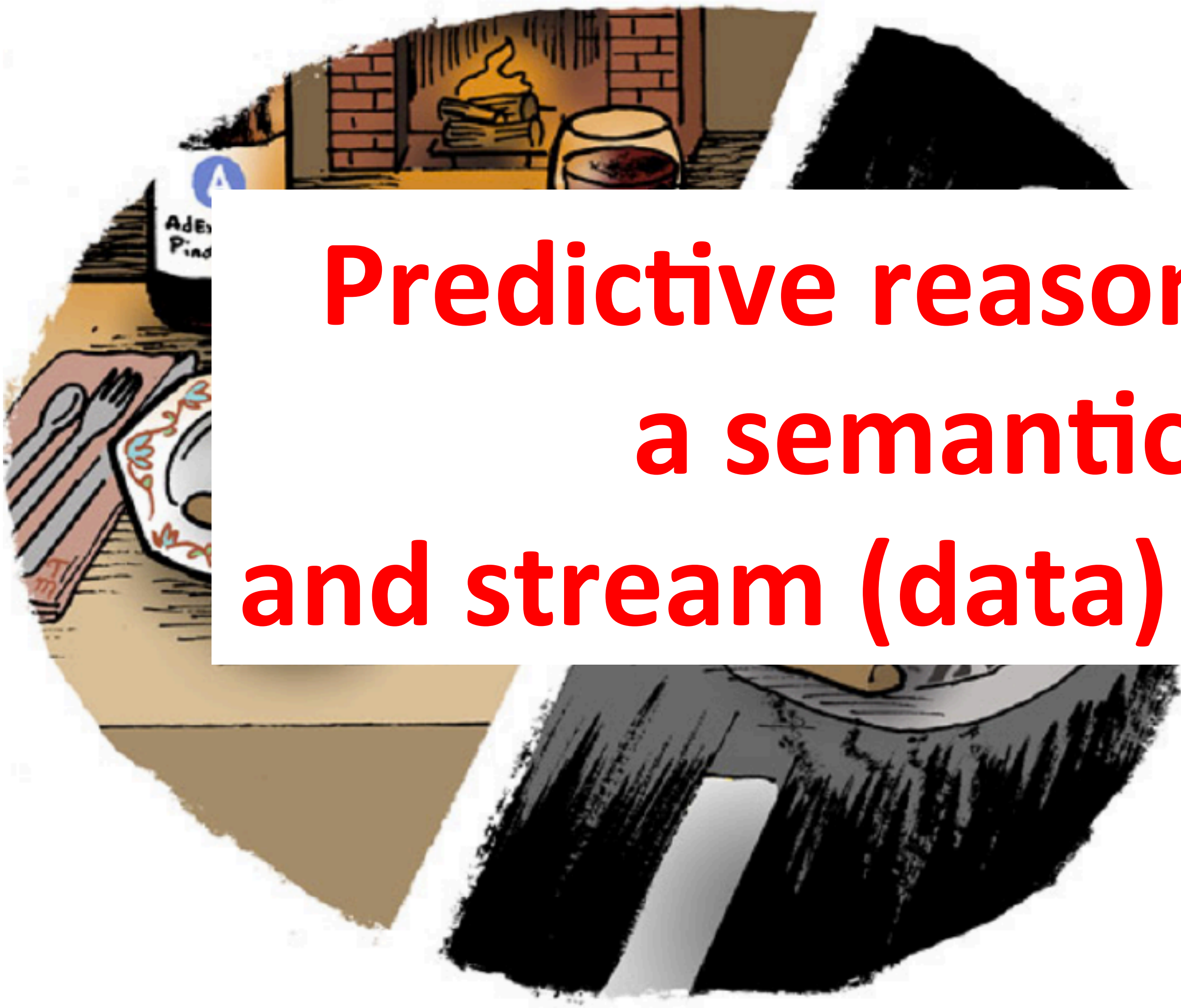
Forecasting Traffic Congestion with Predictive Reasoning

All existing prediction systems are based on **ONE signal / stream**



Forecasting Traffic Congestion with Predictive Reasoning

Why KR&R and Semantic Web Technologies ?



Context Matters

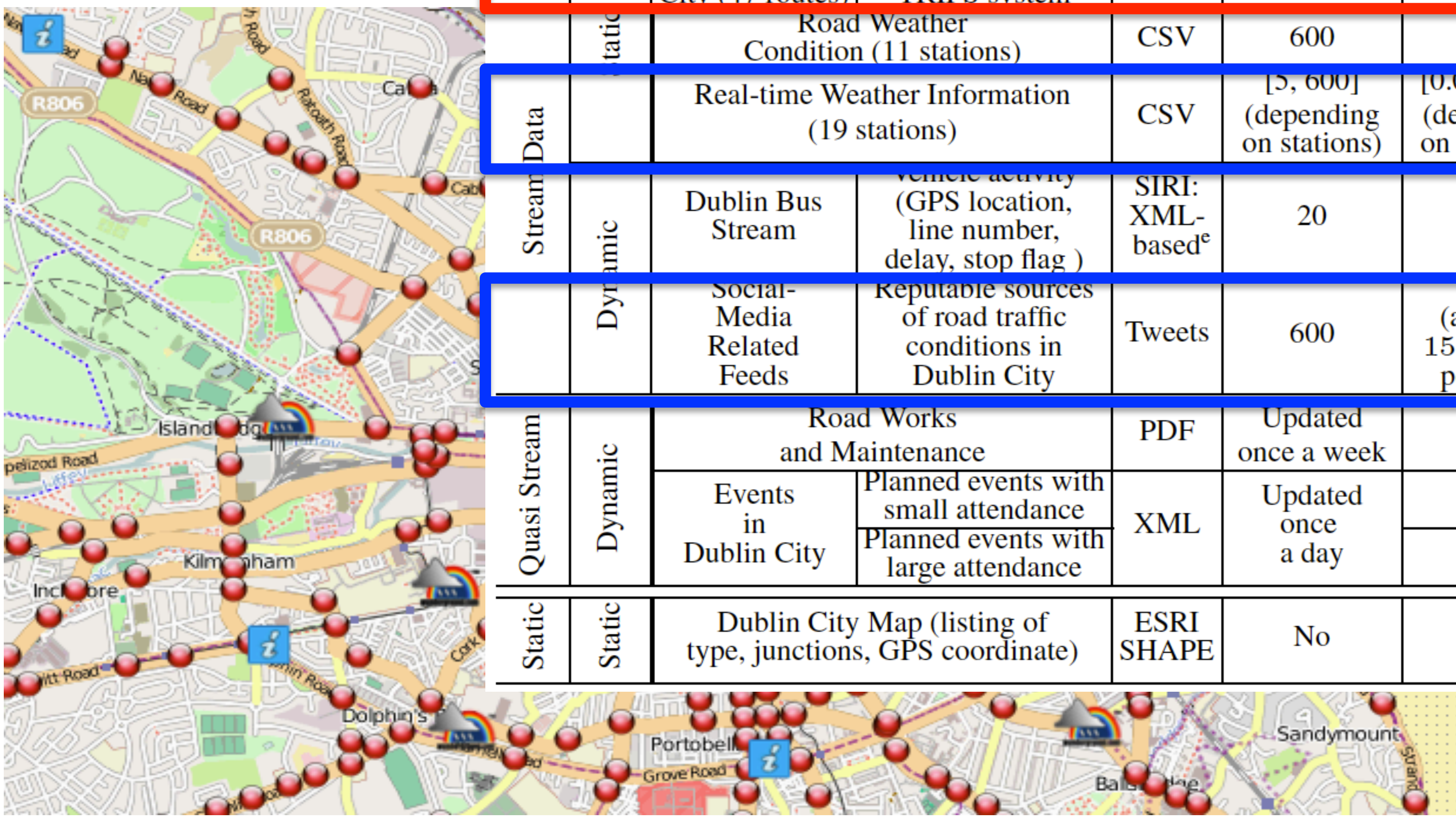
Predictive reasoning in a semantic and stream (data) context

Semantic Pattern Similarity

Forecasting Traffic Congestion with Predictive Reasoning

Context of Dublin City, Ireland

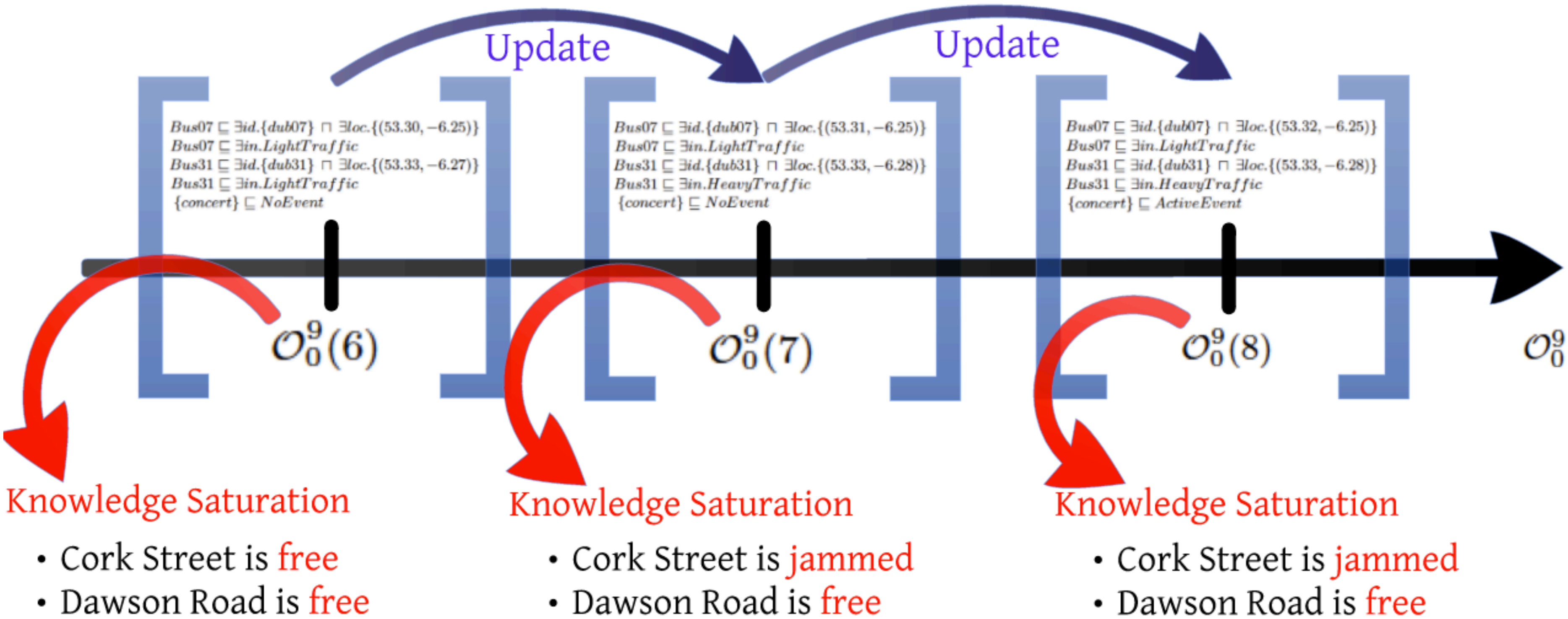
Type	Sensing	Data Source	Description	Format	Temporal Frequency (s)	Size per day (GBytes)	Data Provider (all open data)
Stream Data	Static	Journey times across Dublin City (47 routes)	Dublin Traffic Department's TRIPS system[a]	CSV	60	0.1	Dublin City Council via dublinked.ie[b]
		Road Weather Condition (11 stations)		CSV	600	0.1	NRA[c]
	Dynamic	Real-time Weather Information (19 stations)		CSV	[5, 600] (depending on stations)	[0.050, 1.5] (depending on stations)	Wunderground[d]
		Dublin Bus Stream	Vehicle activity (GPS location, line number, delay, stop flag)	SIRI: XML-based[e]	20	4-6	Dublin City Council via dublinked.ie[f]
		Social-Media Related Feeds	Reputable sources of road traffic conditions in Dublin City	Tweets	600	0.001 (approx. 150 tweets per day)	LiveDrive[g] Aaroadwatch[g] GardaTraffic[g]
Quasi Stream	Dynamic	Road Works and Maintenance		PDF	Updated once a week	0.001	Dublin City Council[h]
		Events in Dublin City	Planned events with small attendance	XML	Updated once a day	0.001	Eventbrite[i]
			Planned events with large attendance			0.05	Eventful[i]
Static	Static	Dublin City Map (listing of type, junctions, GPS coordinate)		ESRI SHAPE	No	0.1	Open StreetMap[j]



Weather Information station Journey Times station Road Weather Condition station

Forecasting Traffic Congestion with Predictive Reasoning

Deductive Reasoning: Ontology Stream as any sequence of ontologies



Application:

Data exposed by The Internet of Thing

accenture
Technology Labs

IJCAI-16

Forecasting Traffic Congestion with Predictive Reasoning

Deductive Reasoning in Ontology Streams

Background Knowledge

$SocialEvent \sqcap \exists type.Music \sqsubseteq Event \sqcap \exists disruption.Steady$
$SocialEvent \sqsubseteq Event$
$Incident \sqcap \exists impact.Serious \sqsubseteq Event \sqcap \exists disruption.High$
$Road \sqcap \exists adj.(\exists occur.(\exists disruption.High)) \sqsubseteq DisruptedRoad$
$BusRoad \sqcap \exists travel.Long \sqsubseteq Road \sqcap \exists with.CongestedBus$
$Road \sqcap \exists with.Bus \sqsubseteq BusRoad$ (6) $\quad Road(r_0)$
$Steady \sqsubseteq High$ (8) $\quad Stop \sqsubseteq Long \sqsubseteq Abnormal$
$Bus(b_1)$ (10) $\quad Bus(b_2)$ (11) $\quad Bus(b_3)$
$Road(r_1)$ (13) $\quad Road(r_2)$ (14) $\quad Road(r_3)$
$adj(r_0, r_1)$ (16) $\quad adj(r_0, r_2)$ (17) $\quad adj(r_0, r_3)$

Always Valid

Rule Engine

R_1	If $X \sqsubseteq A,\ A \sqsubseteq B$ then $X \sqsubseteq B$
R_2	If $X \sqsubseteq A_1, \cdots A_n,\ A_1 \sqcap \cdots \sqcap A_n \sqsubseteq B$ then $X \sqsubseteq B$
R_3	If $X \sqsubseteq A,\ A \sqsubseteq \exists r.B$ then $X \sqsubseteq \exists r.B$
R_4	If $X \sqsubseteq \exists r.A,\ A \sqsubseteq A',\ \exists r.A' \sqsubseteq B$ then $X \sqsubseteq B$
R_5	If $X \sqsubseteq \exists r.A,\ A \sqsubseteq \bot$ then $X \sqsubseteq \bot$
R_6	If $X \sqsubseteq \exists r.A,\ r \sqsubseteq s$ then $X \sqsubseteq \exists s.A$
R_7	If $X \sqsubseteq \exists r_1.A,\ A \sqsubseteq \exists r_2.B,\ r_1 \circ r_2 \sqsubseteq r_3$ then $X \sqsubseteq \exists r_3.B$

Ontology Streams

$\mathcal{P}_0^9(5) : (Event \sqcap \exists disruption.High)(e_1),\quad occur(r_1, e_1)$
$\mathcal{Q}_0^9(5) : (Road \sqcap \exists travel.Long)(r_1)$
$\mathcal{R}_0^9(5) : with(r_1, b_1)$
$\mathcal{P}_0^9(6) : (SocialEvent \sqcap \exists type.Music)(e_2),\quad occur(r_2, e_2)$
$\mathcal{Q}_0^9(6) : (Road \sqcap \exists travel.Abnormal)(r_2)$
$\mathcal{R}_0^9(6) : with(r_2, b_2)$
$\mathcal{P}_0^9(7) : (Incident \sqcap \exists impact.Serious)(e_3),\quad occur(r_3, e_3)$
$\mathcal{Q}_0^9(7) : (Road \sqcap \exists travel.Stop)(r_3)$
$\mathcal{R}_0^9(7) : with(r_3, b_3)$

Valid at some point of time

Derived Knowledge

$$\mathcal{O}, \mathcal{P}_0^9(7) \models^{(3\text{-}4),(7),(18),(30)}_{R_1,R_2,R_3,R_7} DisruptedRoad(r_0)$$

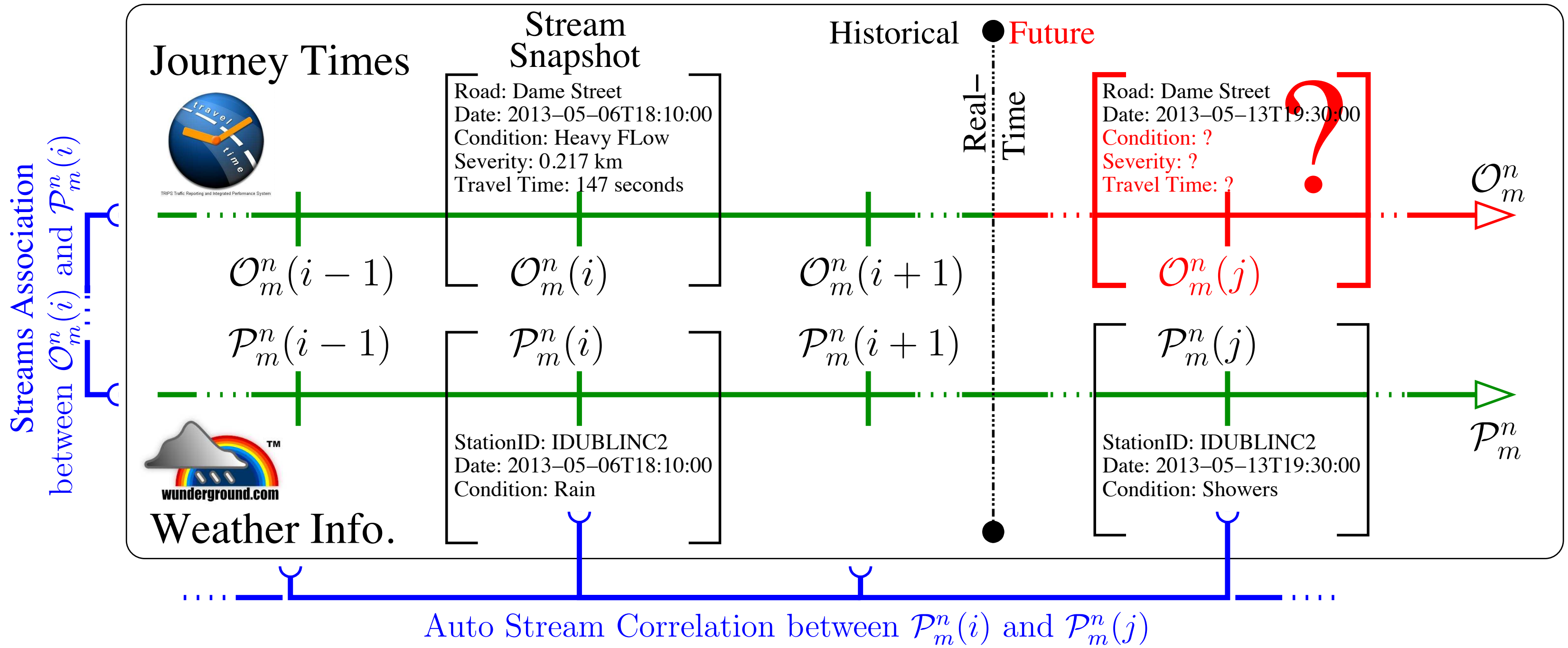
$$\mathcal{O}, \mathcal{Q}_0^9 \cup \mathcal{R}_0^9(7) \models^{(5\text{-}6),(9),(12),(31\text{-}32)}_{R_1,R_2,R_3,R_4} \exists with.CongestedBus(r_3)$$

accenture Technology Labs

IJCAI-16

Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Reasoning on Multiple Streams (1)



How to discover knowledge association across ontology streams?

Representation: **Association Rules**

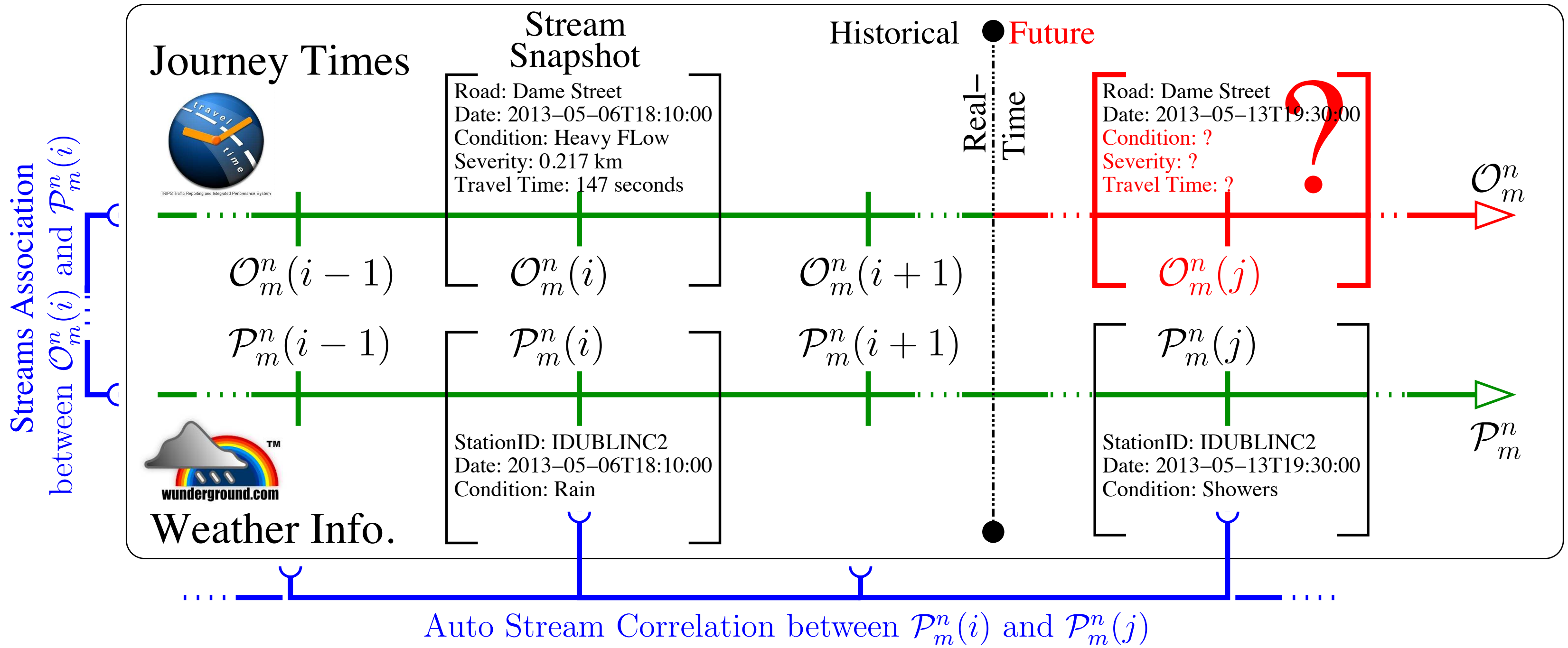
Knowledge Discovery: **Rules Mining**

~~with Semantic data~~

Deductive and Inductive Reasoning

Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Reasoning on Multiple Streams (2)



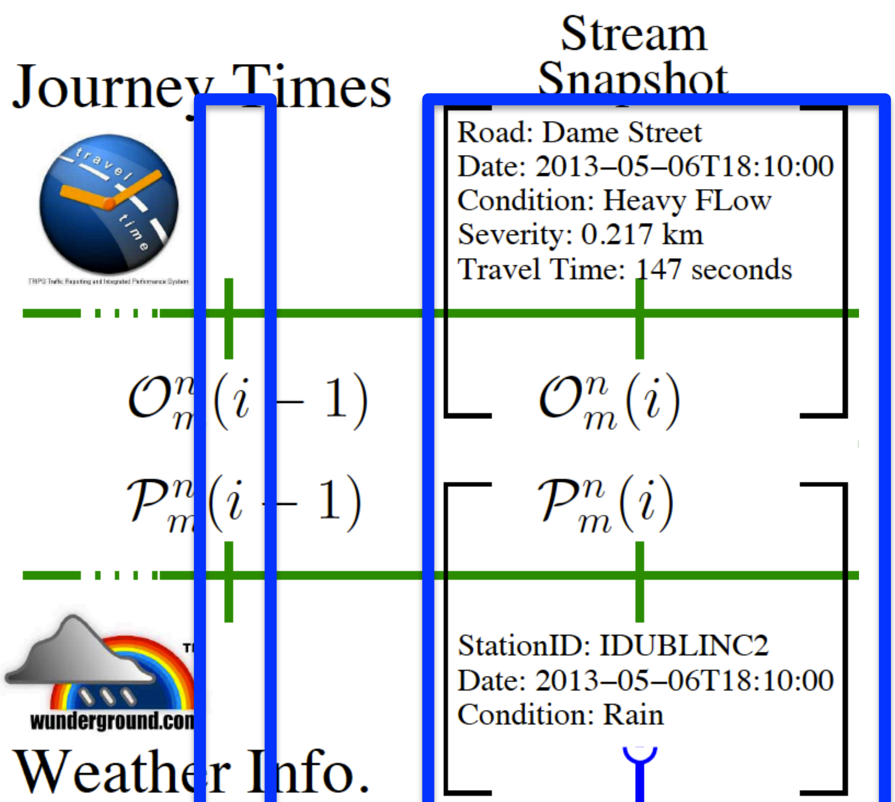
Association Rules as EL++ Rules:

$$(Event \sqcap \exists disruption.High)(x_1) \wedge$$
$$occur(x_2, x_1) \;\wedge\; adj(x_3, x_2)$$
$$\rightarrow (Road \sqcap \exists with.CongestedBus)(x_3)$$

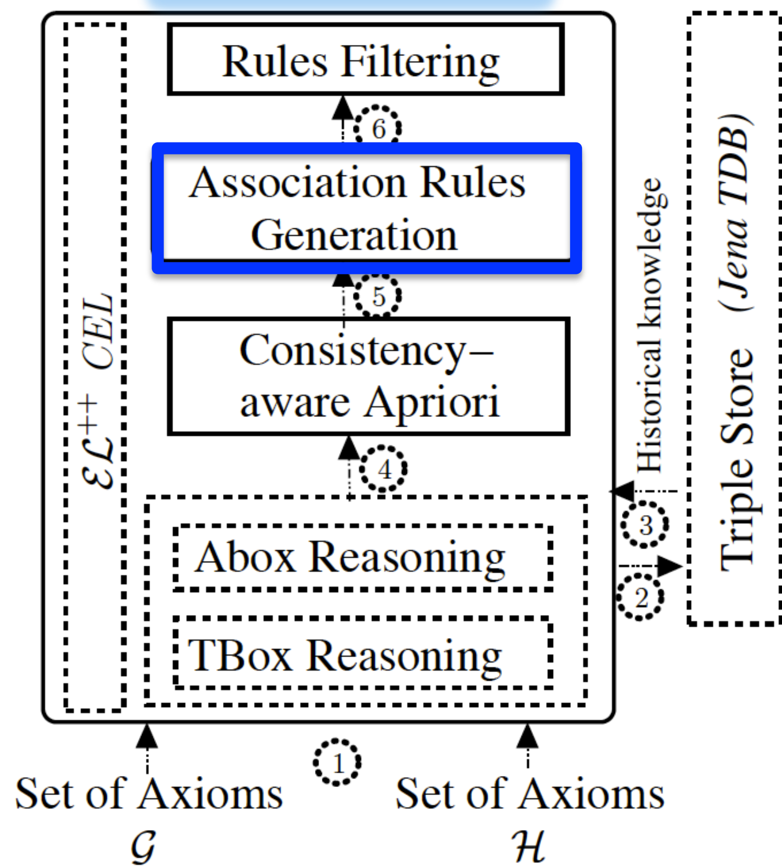
Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Semantic Rules Mining (1)

In a Nutshell



In-use System



Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Semantic Rules Mining (2)

Data Mining

Knowledge Mining

By revisiting concepts of:

- Support
- Confidence
- Weight

With notion of consistent rules

Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Semantic Rules Mining (3)

Data Mining

Knowledge Mining

By revisiting concepts of:

- Support
- Confidence
- Weight

With notion of consistent rules

Ontology Streams

$$\mathcal{P}_0^9(5) : (Event \sqcap \exists disruption.High)(e_1), \quad occur(r_1, e_1) \tag{24}$$
$$\mathcal{Q}_0^9(5) : (Road \sqcap \exists travel.Long)(r_1) \tag{25}$$
$$\mathcal{R}_0^9(5) : with(r_1, b_1) \tag{26}$$
$$\mathcal{P}_0^9(6) : (SocialEvent \sqcap \exists type.Music)(e_2), \quad occur(r_2, e_2) \tag{27}$$
$$\mathcal{Q}_0^9(6) : (Road \sqcap \exists travel.Abnormal)(r_2) \tag{28}$$
$$\mathcal{R}_0^9(6) : with(r_2, b_2) \tag{29}$$
$$\mathcal{P}_0^9(7) : (Incident \sqcap \exists impact.Serious)(e_3), \quad occur(r_3, e_3) \tag{30}$$
$$\mathcal{Q}_0^9(7) : (Road \sqcap \exists travel.Stop)(r_3) \tag{31}$$
$$\mathcal{R}_0^9(7) : with(r_3, b_3) \tag{32}$$

Ontology	$\mathcal{P}_5^7$		$\mathcal{Q}_5^7$		$\mathcal{R}_5^7$	$\mathcal{Q}_5^7 \cup \mathcal{R}_5^7$
Atomset $\mathbb{B}$	$\{(21)\}$	$\{(22)\}$	$\{(28)\}^*$	$\{(31)\}^*$	$\{(32)\}^*$	$\{(23)\}$
Variable $\mathcal{V}$	(x_1)	(x_1, x_2, x_3)	(x_3)	(x_3)	(x_4, x_3)	(x_3)
Binding at Time 5	$\{(e_1)\}$	$\{(e_1, r_1, r_0)\}$	$\{(r_1)\}$		$\{(b_1, r_1)\}$	$\{(r_1)\}$
6	$\{(e_2)\}$	$\{(e_2, r_2, r_0)\}$	$\{(r_2)\}$		$\{(b_2, r_2)\}$	
7	$\{(e_3), (e_4)\}$	$\{(e_3, r_3, r_0), (e_4, r_3, r_0)\}$	$\{(r_3)\}$	$\{(r_3)\}$	$\{(b_3, r_3)\}$	$\{(r_3)\}$
$\sigma(\mathbb{B})$	1	1	1	$1/3$	1	$2/3$
$\omega(\mathbb{B})$	4	4	3	1	3	2

(Confidence, Weight)

$$\left(\frac{\sigma(\{(21),(22),(23)\})}{\sigma(\{(21),(22)\})}, \frac{\omega(\{(21),(22),(23)\})}{\omega(\{(21),(22)\})} \right)$$

$$i.e., \left(\frac{2/3}{3/3}, \frac{3}{4} \right)$$

Rule:

$$(Event \sqcap \exists disruption.High)(x_1) \wedge \tag{21}$$
$$occur(x_2, x_1) \wedge adj(x_3, x_2) \tag{22}$$
$$\twoheadrightarrow (Road \sqcap \exists with.CongestedBus)(x_3) \tag{23}$$

Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Semantic Rules Mining (4)

Data Mining

By revisiting concepts of:
- Support
- Confidence
- Weight

With notion of consistent rules

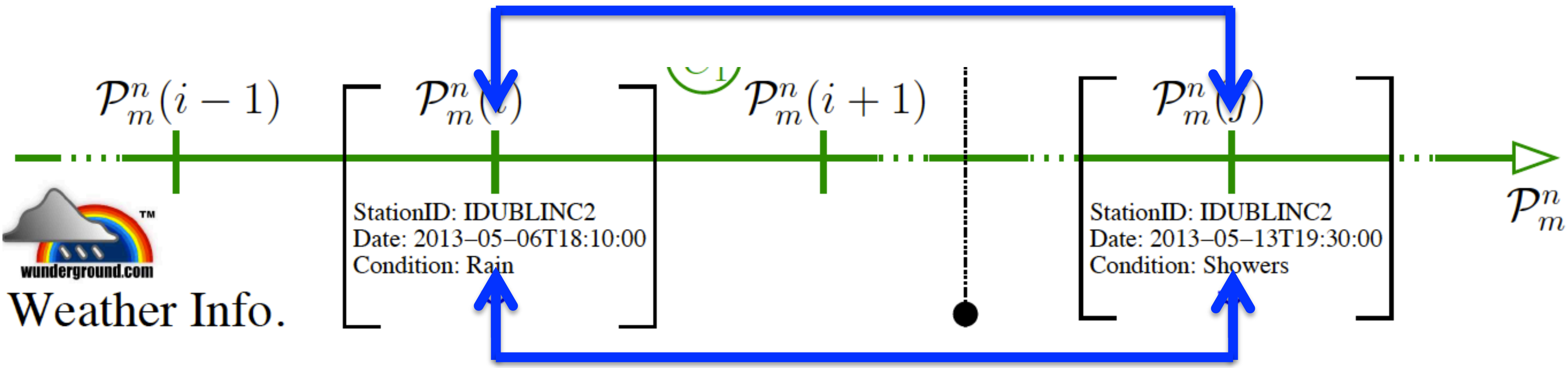
Knowledge Mining

- **Knowledge Mining in Ontology Streams:**
  - Atomset mining: identification of all potential consistent combinations
  - Rule generation: identification of significative combination of atomsets

- **Limitations**:
  - Exponential number of rules (although semantics helped)
  - Rules filtering is the bottleneck
  - No cross temporal association (e.g., t+1 and t)

Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Stream Auto-Correlation (1)

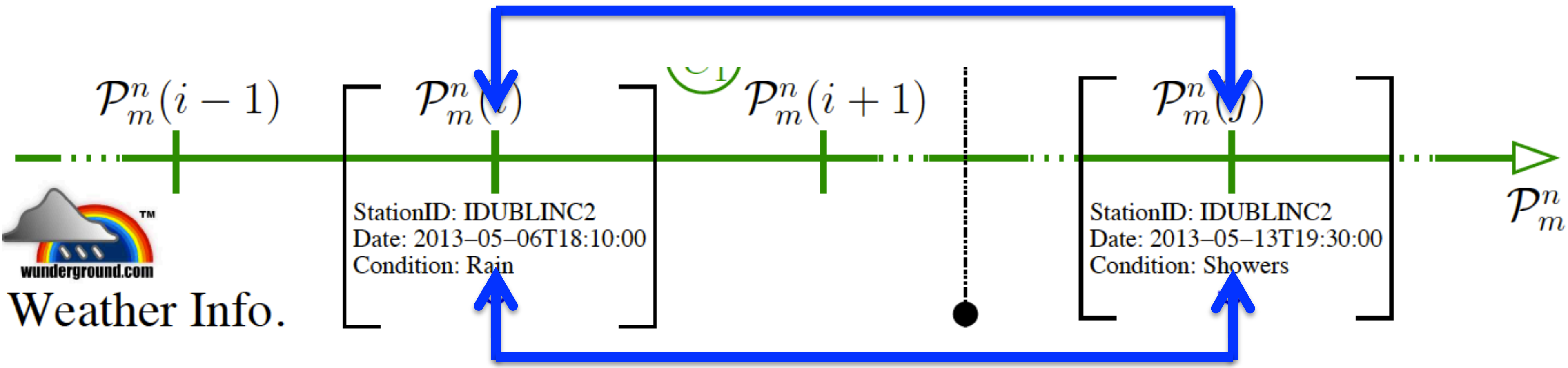


Identification of similarity among ontology snapshots:

- New, obsolete, invariant Abox assertions;
- Any other knowledge similarity metric could apply.

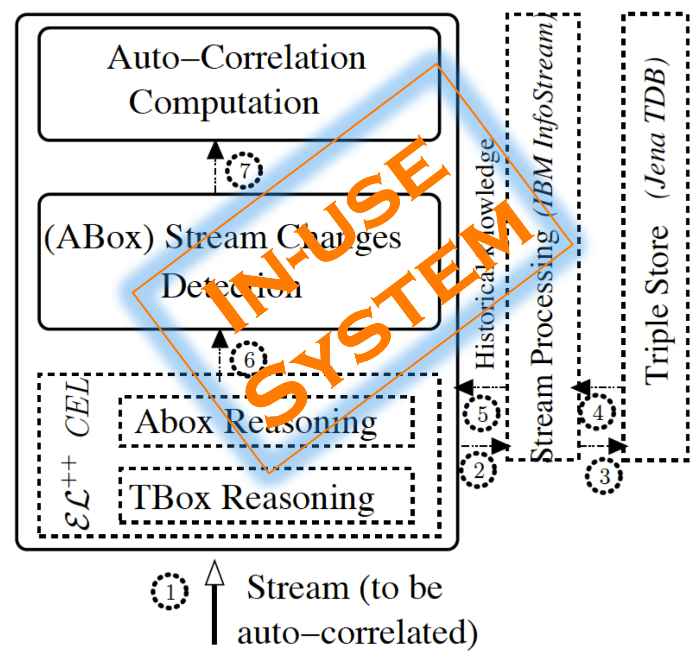
Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Stream Auto-Correlation (2)



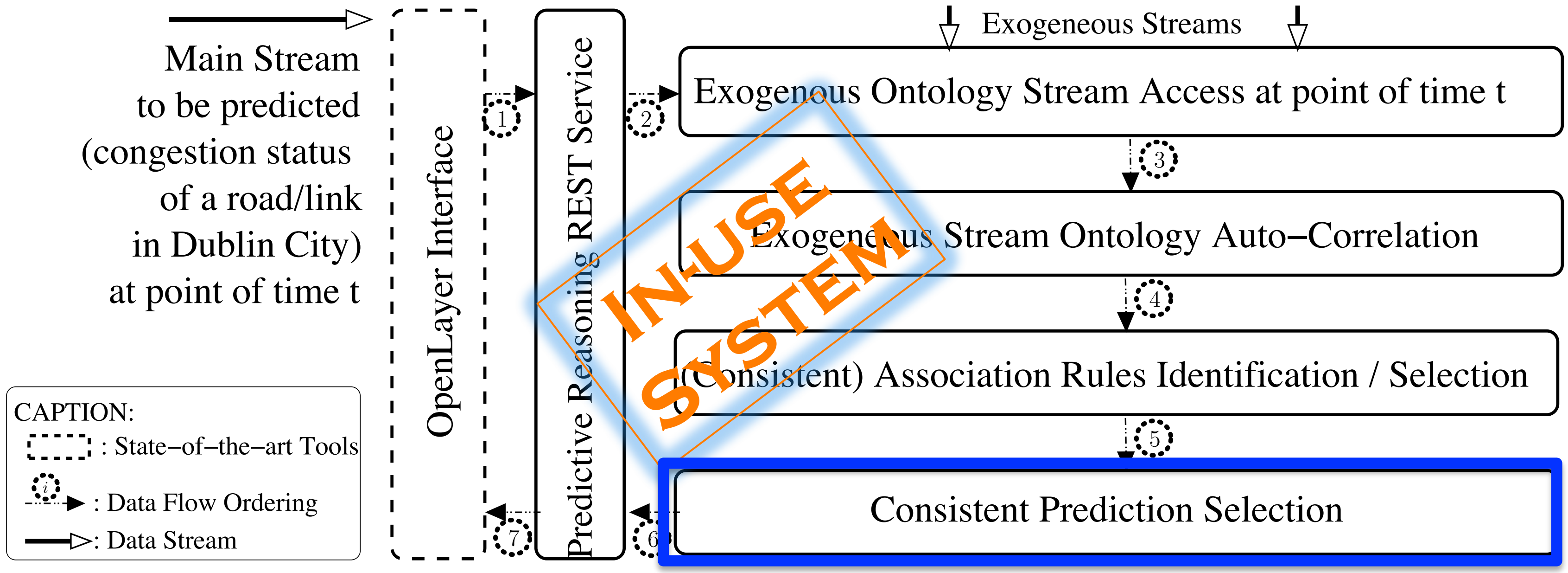
- **Limitations**:
  1. The **more** historical **data / streams** (the more auto-correlation evaluation)
  2. The **more expressive** data **streams**

  ➢ The **less scalable** the approach



Forecasting Traffic Congestion with Predictive Reasoning

Knowledge Discovery in Ontology Streams – Scalable and Consistent Prediction



Example of consistent rule

$$
\begin{aligned}
HeavyTrafficFlow(s) \leftarrow\, & Road(r_1) \wedge Road(r_2) \wedge isAdjacentTo(r_1, r_2) \wedge \\
& hasTravelTimeStatus(r_1, s) \wedge hasWeatherPhenomenon(r_1, w) \wedge \\
& OptimunHumidity(w) \wedge hasTrafficPhenomenon(r_2, a) \wedge \\
& RoadTrafficAccident(a) \qquad (12)
\end{aligned}
$$

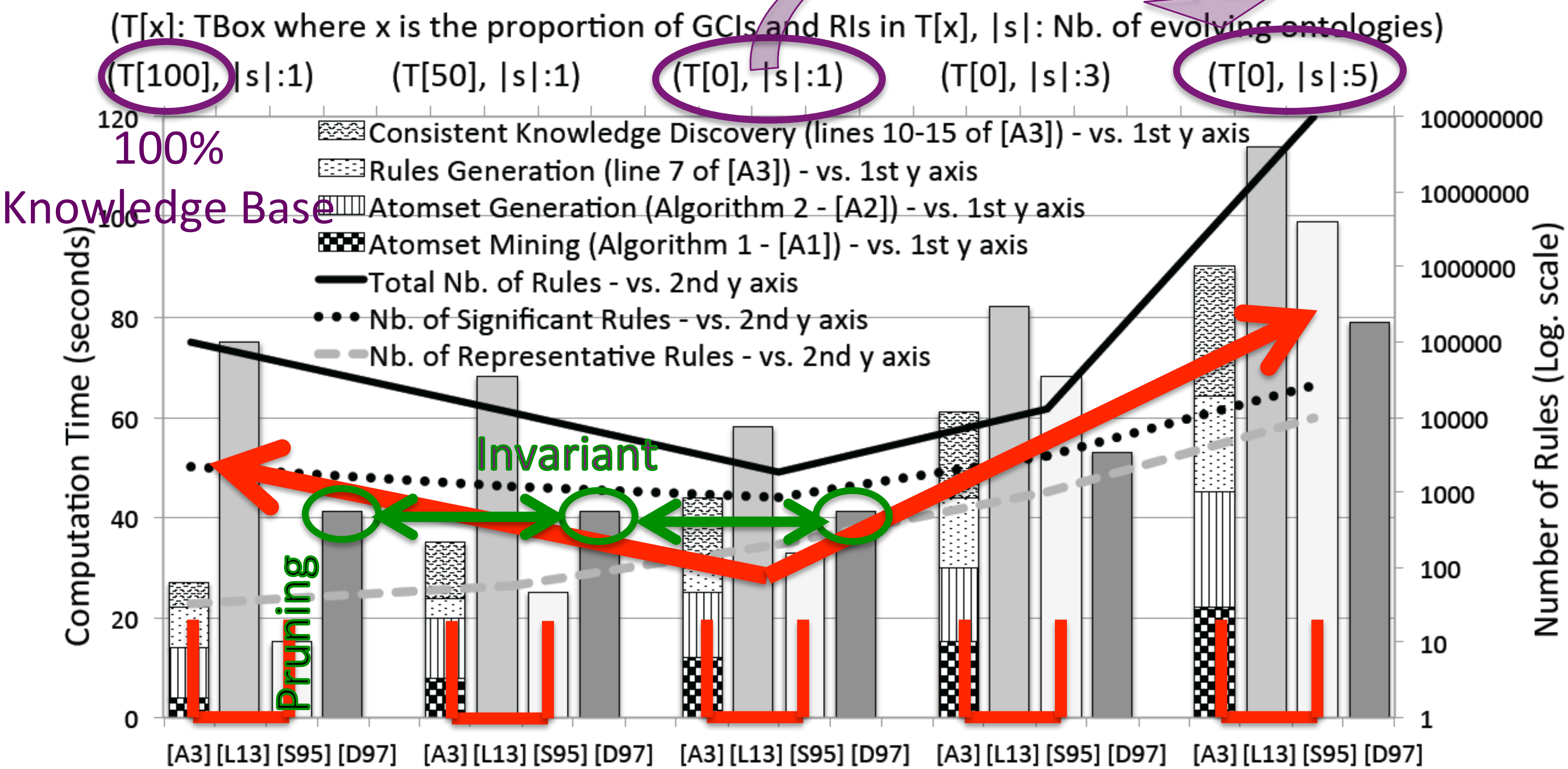
Limitations:

- **Scalability of consistency checking** on rules consequent

Forecasting Traffic Congestion with Predictive Reasoning

Experimentation: Scalability

DataSet	Size (Mb / Day)	#Axioms / Update	#RDF Triples / Update
[a] Weather	3	53	318
[b] Travel Time	43	270	810
[c] Incident	0.1	81	324
[d] Event	9.5	480	1, 150
[e] Bus	120	3, 000	12, 000

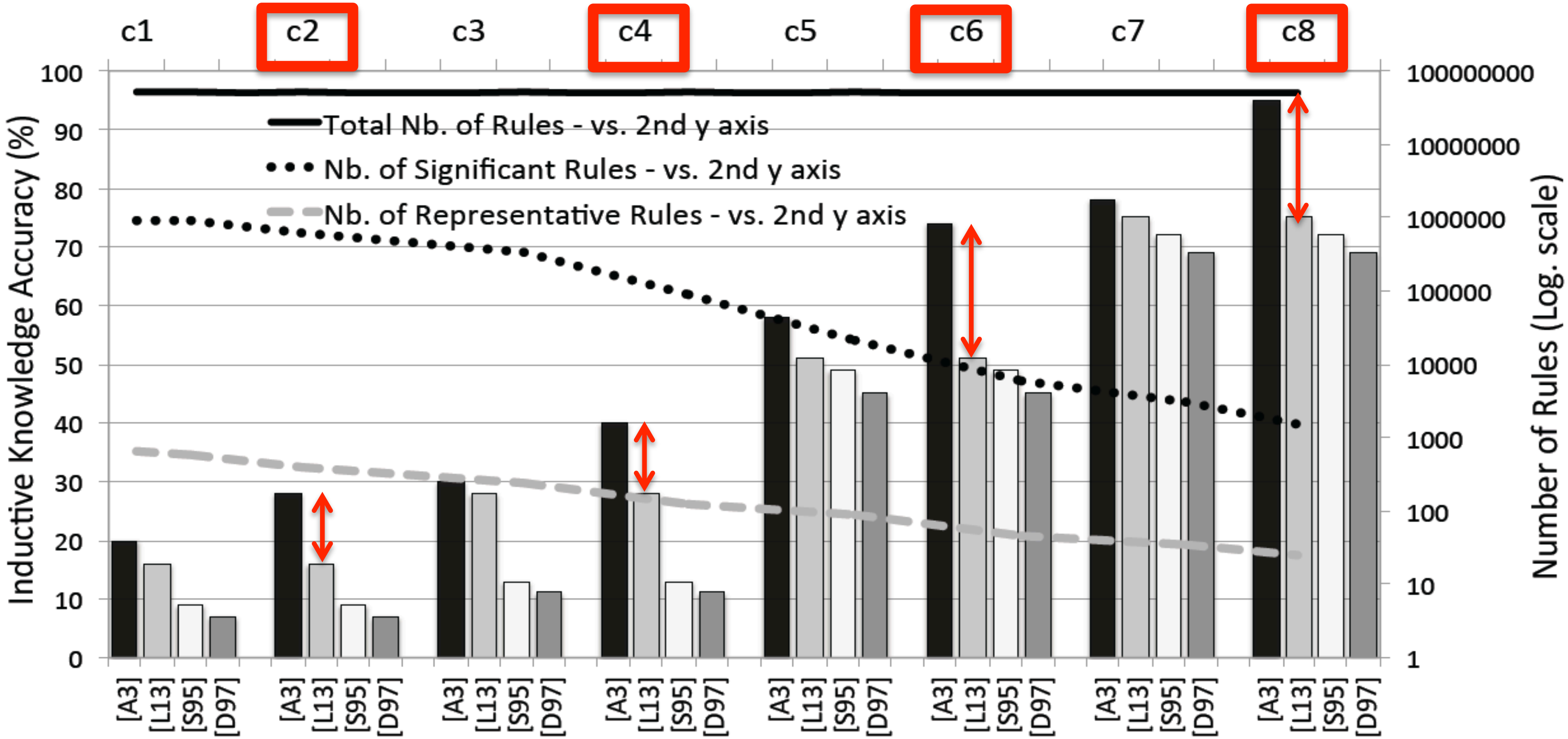


[A3] Lécué F., Pan J.: Consistent Knowledge Discovery from Evolving Ontologies. AAAI 2015
[L13] Lécué F., Pan J.: Predicting Knowledge in an Ontology Stream. IJCAI 2013
[S95] Srikant R., Agrawal R. Mining generalized association rules. VLDB 1995
[D97] Dehaspe L., Raedt L. D. Mining association rules in multiple relations. ILP 1997

Forecasting Traffic Congestion with Predictive Reasoning

Experimentation: Accuracy of Prediction

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
$\sigma_{\min}$	.4	.4	.4	.4	.8	.8	.8	.8
$\gamma_{\min}$	(.4, .4)	(.4, .8)	(.8, .4)	(.8, .8)	(.4, .4)	(.4, .8)	(.8, .4)	(.8, .8)



[A3] Lécué F., Pan J.: Consistent Knowledge Discovery from Evolving Ontologies. AAAI 2015

[L13] Lécué F., Pan J.: Predicting Knowledge in an Ontology Stream. IJCAI 2013

[S95] Srikant R., Agrawal R. Mining generalized association rules. VLDB 1995

[D97] Dehaspe L., Raedt L. D. Mining association rules in multiple relations. ILP 1997

Distributed reasoning to scale up deductive reasoning

Motivation

- Ontologies generated from sensors can grow very large in a short span.
- Streaming traffic data produces several million axioms in a day.
- Output of deductive reasoning process is larger than input.

Distributed reasoning to scale up deductive reasoning

Motivation

- Current in-memory single machine reasoners cannot scale.
- A distributed approach to ontology reasoning is required.
- Easy to add more memory and processing power in a distributed setup.

Distributed reasoning to scale up deductive reasoning

Preliminaries

- Classification: For each class in the ontology, find all its superclasses.

- Completion Rules
  - To classify an ontology a set of completion rules are applied iteratively on the axioms.
  - They are applied until no new output is produced.
  - $S(X)$ is the set containing all the superclasses of X.
  - $R(r) = \{(A, B)\}$ implies that $A \sqsubseteq \exists r.B$

Distributed reasoning to scale up deductive reasoning

Preliminaries

- **R1** If $C' \in S(C), C' \sqsubseteq D \in \mathcal{C}$, and $D \notin S(C)$
  then $S(C) := S(C) \cup \{D\}$
- **R2** If $C_1, C_2 \in S(C), C_1 \sqcap C_2 \sqsubseteq D \in \mathcal{C}$, and $D \notin S(C)$
  then $S(C) := S(C) \cup \{D\}$
- **R3** If $C' \in S(C), C' \sqsubseteq \exists r.D \in \mathcal{C}$, and $(C, D) \notin R(r)$
  then $R(r) := R(r) \cup \{(C, D)\}$
- **R4** If $(C, D) \in R(r), D' \in S(D), \exists r.D' \sqsubseteq E \in \mathcal{C}$,
  and $E \notin S(C)$
  then $S(C) := S(C) \cup \{E\}$

Distributed reasoning to scale up deductive reasoning

Preliminaries

- R5 If $(C, D) \in R(r)$, $\bot \in S(D)$, and $\bot \notin S(C)$, then $S(C) := S(C) \cup \{\bot\}$

- R6 If $(C, D) \in R(r)$, $r \sqsubseteq s \in \mathcal{C}$, and $(C, D) \notin R(s)$ then $R(s) := R(s) \cup \{(C, D)\}$

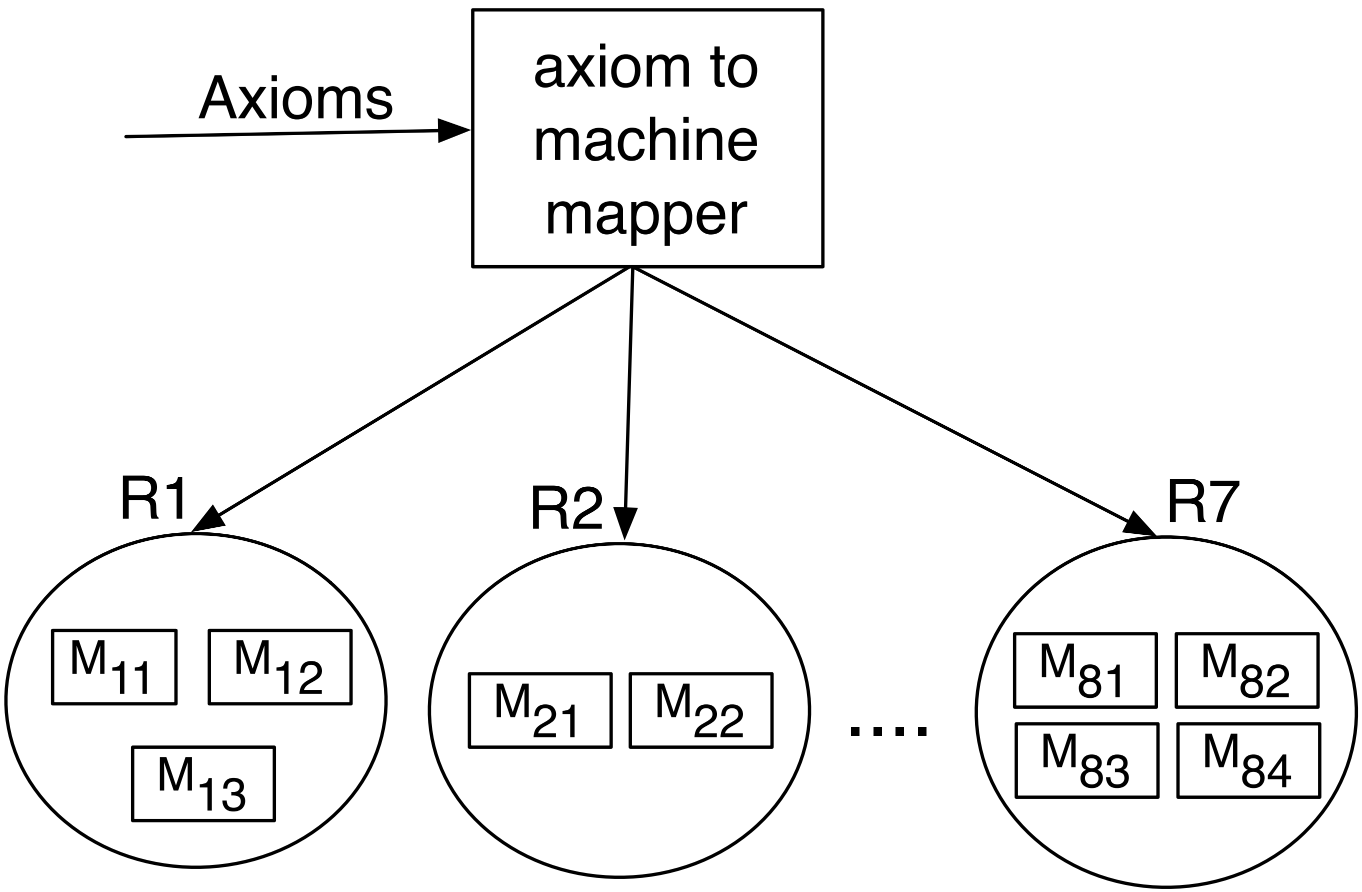
- R7 If $(C, D) \in R(r_1)$, $(D, E) \in R(r_2)$, $r_1 \circ r_2 \sqsubseteq r_3 \in \mathcal{C}$, and $(C, E) \notin R(r_3)$ then $R(r_3) := R_i(r_3) \cup \{(C, E)\}$

Distributed reasoning to scale up deductive reasoning

- Data distribution and communication management play a crucial role.
- Axiom distribution
  - Data locality
  - $O = O_1 \cup \ldots \cup O_7$
  - Each O_i is assigned to a group of nodes.

Distributed reasoning to scale up deductive reasoning

Data distribution



Distributed reasoning to scale up deductive reasoning

Targeted communication

- Rule R_i is applied on the corresponding O_i.
- The output of one rule is relevant to only a specific set of rules.
- Rule processes can directly send messages to the relevant one.

Distributed reasoning to scale up deductive reasoning

Dynamic load balancing

- Improper load balancing due to number and type of axioms.
- Idle nodes help the busy nodes by stealing some work from them.
- Performance improvement outweighs communication cost.

Distributed reasoning to scale up deductive reasoning

Implementation

- DistEL, implemented in Java. https://github.com/raghavam/DistEL
- A key-value store named Redis is used to store axioms.
- Redis provides set operations, database sharding, transactions, server-side scripting.

Distributed reasoning to scale up deductive reasoning

Experiments

- Amazon EC2, m3.xlarge instances are used.
- Biomedical ontologies such as GO, SNOMED CT and traffic data are used. 1441 bursts of traffic data was considered.

	GO	SNOMED	2-SNO	3-SNO	5-SNO	Traffic
Before classification	87,137	1,038,481	2,076,962	3,115,443	5,192,405	7,151,328
After classification	868,996	14,796,555	29,593,106	44,389,657	73,982,759	21,840,440

Distributed reasoning to scale up deductive reasoning

Experiments

- 5GB is available to JVM.
- All reasoners invoked through OWL API.

Ontology	ELK	jCEL	Srocket	Pellet	HermiT	FaCT++
GO	23.5	57.4	40.3	231.4	91.7	367.89
SNOMED	31.8	126.6	52.34	620.46	1273.7	1350.5
2-SNOMED	77.3	OOM[a]	OOM[a]	OOM[a]	OOM[a]	OOM[a]
3-SNOMED	OOM[a]	OOM[a]	OOM[a]	OOM[a]	OOM[a]	OOM[a]
5-SNOMED	OOM[a]	OOM[a]	OOM[a]	OOM[a]	OOM[a]	OOM[a]
Traffic	OOM[b]	OOM[c]	OOM[c]	OOM[b]	OOM[b]	OOM[c]

Distributed reasoning to scale up deductive reasoning

Experiments

Ontology	8 nodes	16 nodes	24 nodes	32 nodes	64 nodes
GO	134.49	114.66	109.46	156.04	137.31
SNOMED	544.38	435.79	407.38	386	444.19
2-SNOMED	954.17	750.81	717.41	673.08	799.07
3-SNOMED	1362.88	1007.16	960.46	928.41	1051.80
5-SNOMED	2182.16	1537.63	1489.34	1445.30	1799.13
Traffic	60004.54	41729.54	39719.84	38696.48	34200.17

- Runtime decreases as nodes increase.
- With 64 nodes, for traffic data, each burst takes 23.73 seconds.

Distributed reasoning to scale up deductive reasoning

Experiments

- Speedup on SNOMED

Nodes	Runtime	Speedup
8	544.38	1
16	435.79	1.24
24	407.38	1.33
32	386	1.41
64	444.19	1.22

- After a point, advantages of distributed approach is overshadowed by distribution costs.

Distributed reasoning to scale up deductive reasoning

Experiments

- Memory taken by Redis on each node for Traffic data.

Node	Memory (MB)
R1	186.72
R2	0.81
R3	257.47
R4	0.79
R5	1970
R6	380.61
R7	0.79
R8	1470
Result	654.53
Total	4921.72

Distributed reasoning to scale up deductive reasoning

Future work

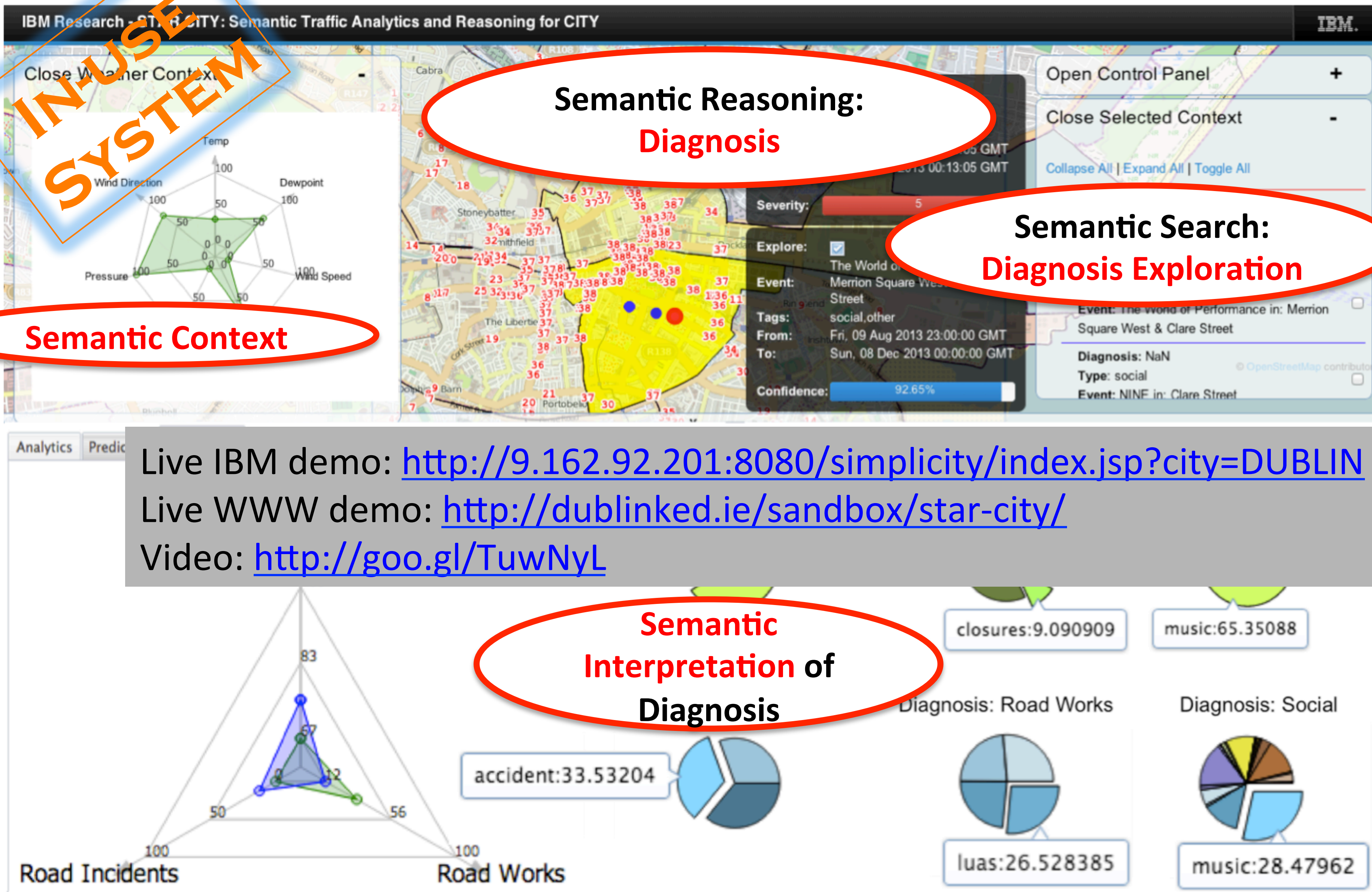
- Estimate number of nodes required for optimum performance for a given ontology.
- Alternate rule sets and classification procedure can be tried (eg., ELK).
- Distributed approaches to more expressive ontologies.

PART III Applications in Bologna, Dublin, Miami, Rio

accenture Technology Labs

IJCAI-16

STAR-CITY (Semantic Traffic Analytics and Reasoning for CITY)



Live IBM demo: http://9.162.92.201:8080/simplicity/index.jsp?city=DUBLIN

Live WWW demo: http://dublinked.ie/sandbox/star-city/

Video: http://goo.gl/TuwNyL

STAR-CITY (Semantic Traffic Analytics and Reasoning for CITY)

Context

- **Dublin**: (Diagnosis of) Traffic congestion
- **Bologna**: (Diagnosis of) Bus congestion
- **Miami**: (Diagnosis of) Bus bunching
- **Rio**: (Diagnosis of) Low on-time performance of buses

Source of Anomaly

Source of Diagnosis

Source Type	Data Source	Description	City			
			Dublin (Ireland)	Bologna (Italy)	Miami (USA)	Rio (Brazil)
Traffic Anomaly	Journey travel times across the city	Traffic Department's TRIPS system[a]	CSV format (47 routes, 732 sensors) 0.1 GB per day[b]	✗ (not available)		
Traffic Anomaly	Dublin Bus Dynamics	Vehicle activity (GPS location, line number, delay, stop flag)	✗ (not used)	SIRI: XML format[c] (596 buses, 80KB per update 11GB per day[d])	CSV format (893 buses, 225 KB per update 43 GB per day[e])	CSV format (1, 349 buses, 181 KB per update 14 GB per day[f])
Traffic Diagnosis	Social-Media Related Feeds	Reputable sources of road traffic conditions in Dublin City	"Tweet" format - Accessed through Twitter streaming API[g]			
			Approx. 150 tweets per day[h] (approx. 0.001 GB)	✗ (not available)	Approx. 500 tweets per day[i] (approx. 0.003 GB)	✗ (not available)
Traffic Diagnosis	Road Works and Maintenance		PDF format (approx. 0.003 GB per day[j])	XML format (approx. 0.001 GB per day[k])	HTML format (approx. 0.001 GB per day[l])	✗ (not available)
Traffic Diagnosis	Social events e.g., music event, political event	Planned events with small attendance	XML format - Accessed once a day through Eventbrite[m] APIs			
			Approx. 85 events per day (0.001 GB)	Approx. 35 events per day (0.001 GB)	Approx. 285 events per day (0.005 GB)	Approx. 232 events per day (0.01 GB)
Traffic Diagnosis	Social events e.g., music event, political event	Planned events with large attendance	XML format - Accessed once a day through Eventful[n] APIs			
			Approx. 180 events per day (0.05 GB)	Approx. 110 events per day (0.04 GB)	Approx. 425 events per day (0.1 GB)	Approx. 310 events per day (0.08 GB)
Traffic Diagnosis	Bus Passenger Loading / Unloading (information related to number of passenger getting in / out)		✗ (not available)	✗ (not available)	CSV format (approx. 0.8 GB per day[e])	CSV format (approx. 0.1 GB per day[e])

STAR-CITY (Semantic Traffic Analytics and Reasoning for CITY)

Data

Real Time, Live Data	City	Frequency of Update (s)	Raw Update Size (KB)	Semantic Update Size (KB)	Semantic Update #RDF Triples	Semantic Conversion Computation Time (s)
[a] Journey Times	Dublin	60	20.2	6, 102	63, 000	0.61
[b] Bus	Bologna	120	31.8	1, 166	4, 550	0.295
	Miami	40	66.8	1, 766	11, 000	0.415
	Rio	60	96.8	2, 366	16, 145	0.595
[c] Incident	Dublin	600	0.2	1.0	7	0.002
	Miami	180	0.2	1.0	9	0.002
[d] Road Works	Dublin	once a week	146.6	77.9	820	3.988
	Bologna	once a day	78.9	133.2	1, 100	0.988
	Miami	3600	102.6	103.6	912	1.388
[e] City Events	Dublin	once a day	240.7	297	612	1.018
	Bologna		111.2	149	450	0.434
	Miami		637.2	789	1, 190	1.876
	Rio		585.3	650	950	1.633
[f] Bus Loading	Miami	40	833	2, 500	4, 500	0.390
	Rio	60	69.7	650	1, 230	0.147

Vocabulary

Ontology	Size (KB)	#Concepts	#Object Properties	#Data Properties	#Individuals	Imported Ontologies	Data Sets Covered
IBM Travel Time	4, 194	41	49	22	1, 429	Time Geo	[a]
IBM SIRI-BUS [4]	41.9	21	17	18	-		[b]
LODE[a] (initial)	12	14	16	-	-		[e]
(extended)	56	87	68	31	-	Time, Geo	[c-f]
W3C Time[b]	25.2	12	24	17	14	-	[a-f]
W3C Geo[c]	7.8	2	4	-	-	-	[a-f]
DBpedia	Only a subset is used for annotation i.e., 28 concepts, 9 data properties						[c-e]

accenture Technology Labs

IJCAI-16

STAR-CITY (Semantic Traffic Analytics and Reasoning for CITY)

Objective: Real-Time and Historical Traffic Diagnosis (1)

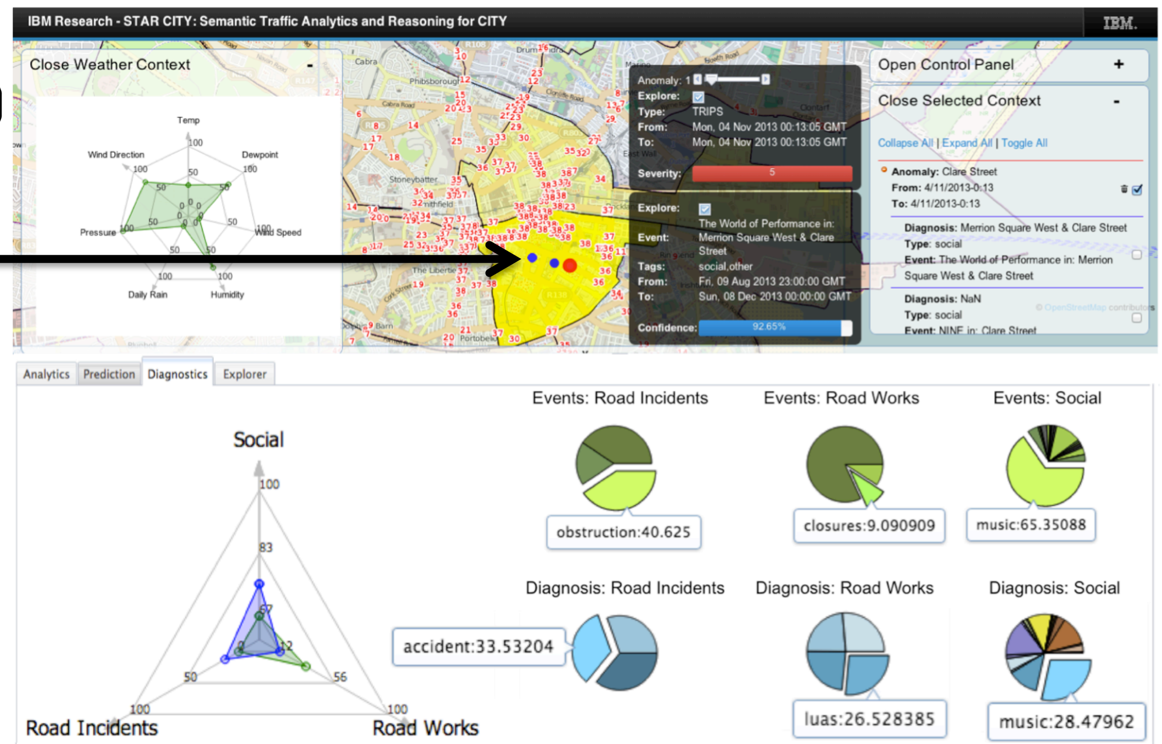
Live demo: http://9.162.92.201:8080/simplicity/index.jsp?city=DUBLIN
Video: http://goo.gl/TuwNyL

Input

- Type of anomaly (bus speed, ambulance delay…)
- Type of explanation (city events, unplanned events, road works, …)

Output

- Real-time/Historical diagnosis results
- Evolution of diagnosis over time/space
- Comparison vs. historical



STAR-CITY (Semantic Traffic Analytics and Reasoning for CITY)

Objective: Real-Time and Historical Traffic Diagnosis (2)

Live demo: http://9.162.92.201:8080/simplicity/index.jsp?city=DUBLIN
Video: http://goo.gl/TuwNyL

Input

- Type of anomaly (bus speed, ambulance delay…)
- Type of explanation (city events, unplanned events, road works, …)

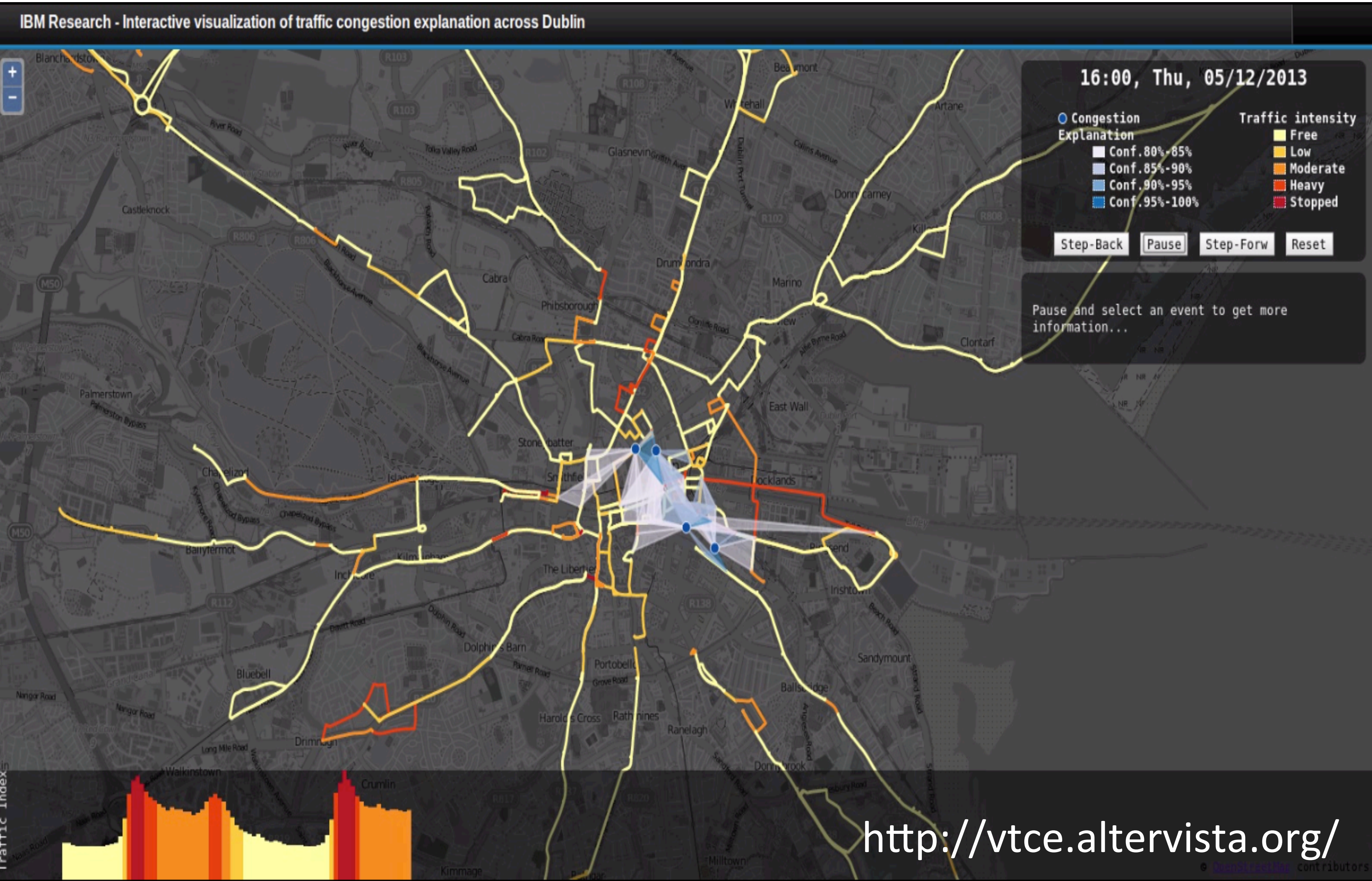
Output

- Categorization of diagnosis

accenture Technology Labs

IJCAI-16

Reverse STAR-CITY system for city managers



http://vtce.altervista.org/

Reverse STAR-CITY system for city managers

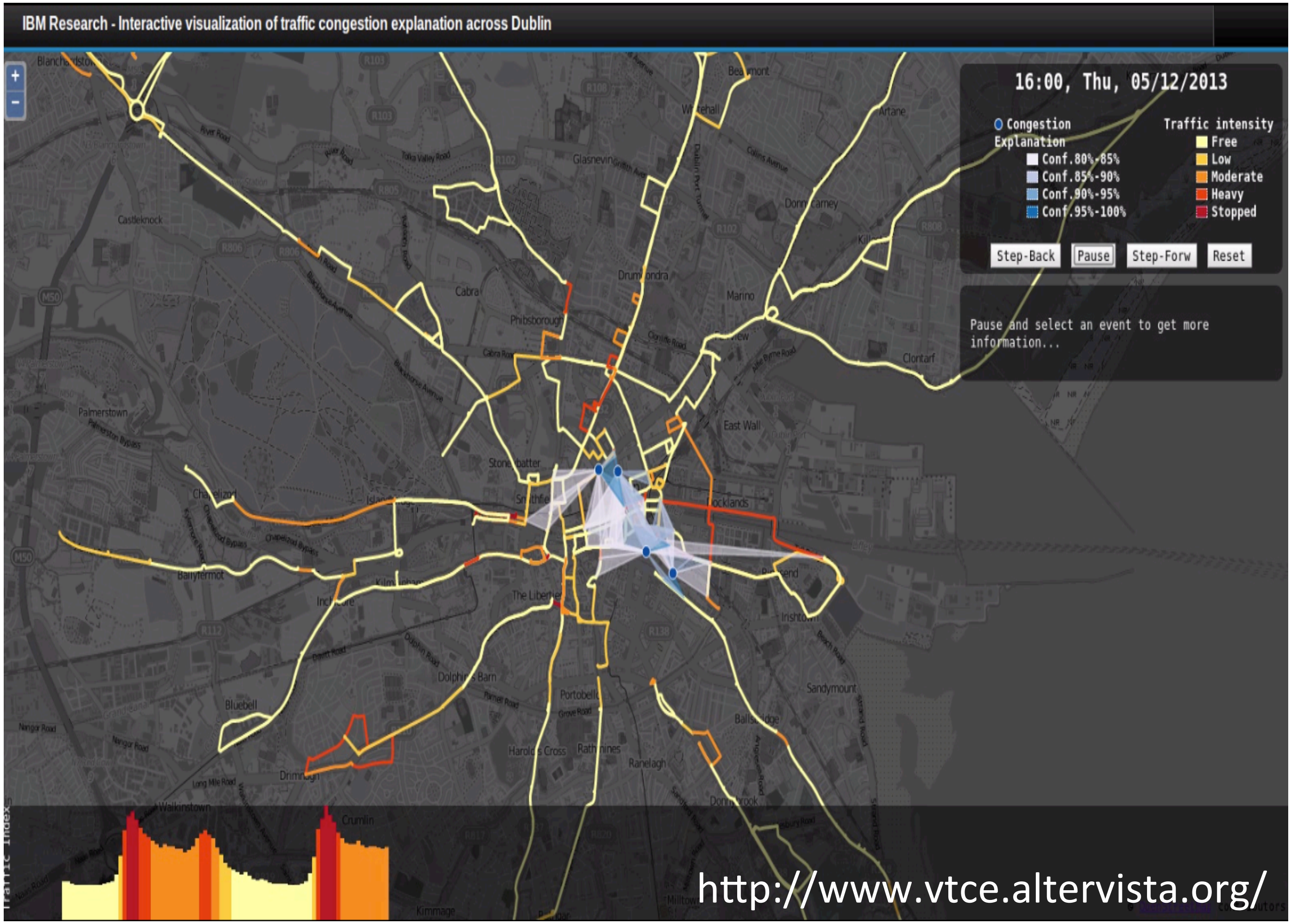
<u>Objective: City Planning</u>

Input

- Type of anomaly (bus speed, ambulance delay…)
- Type of explanation (city events, unplanned events, road works, …)

Output

- Impact of events and their characteristics on anomalies



Reverse STAR-CITY system for city managers

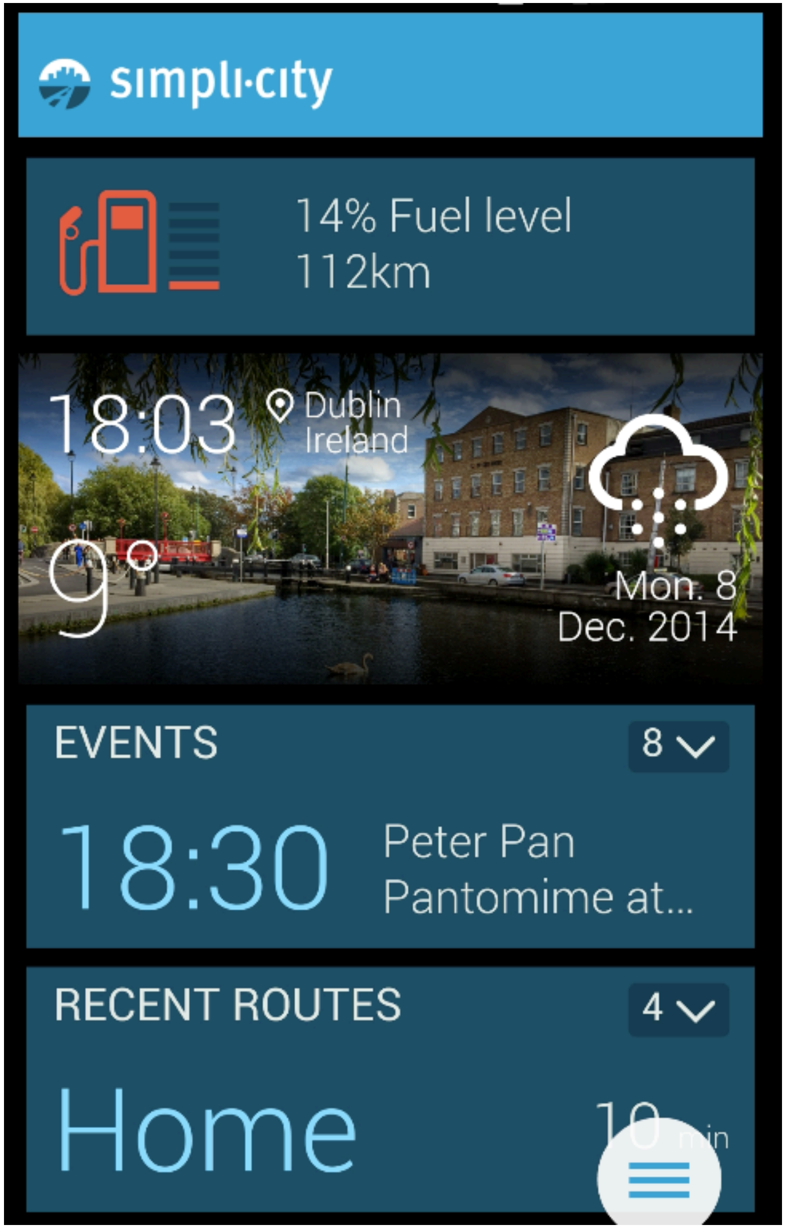
Data

	Data Source	Description	Format Type	Temporal Frequency (s)	Historic (mm/yyyy)	Size Estimation per day (GBytes)	Data Provider
Source of Effects	Dublin Bus	Vehicle activity (GPS location, line number, delay, stop flag)	SIRI[a] XML-based	20	11/2010	4-6	(Private) DCC
Source of Causes	Wunderground for Dublin	Real-time weather information	CSV	[5, 600] (depending on stations)	01/1996	[0.050, 1.5] (depending on stations)	(Public) Wunderground[b]
	Road Weather Condition (54 stations)		CSV	600	11/2010	0.1	(Public) NRA[c]
	Road Works and Maintenance		CSV	3600	11/2010	0.01	(Public) Dublinked[d]
	Events in Dublin	Events with small attendance	XML	Not considered	11/2011	0.001	(Public) Eventbrite[e]
		Events with large attendance			11/2011	0.05	(Public) Eventful[f]
Semantic Data	DBPedia	Structured facts extracted from wikipedia	RDF	No	No	3.5×10^6 concepts	(Public) DBPedia[g]
	Dublin City Roads (listing of type, junctions, GPS coordinate)		RDF	No	No	0.1	(Public) Linked-geodata[h]

Vocabulary

Ontology	Size (KB)	#Concepts	#Object Properties	#Data Properties	#Individuals	Imported Ontologies
NASA SWEET[12] (IBM adaptation)	158.8	90	40	34	63	W3C Time, Geo
IBM Travel Time	4, 194	41	49	22	1, 429	
IBM SIRI-BUS	41.9	21	17	18	-	
W3C Time[9]	25.2	12	24	17	14	-
W3C Geo[10]	7.8	2	4	-	-	-
DBpedia	Only a subset is used for annotation i.e., 28 concepts, 9 data properties					

Mobile STAR-CITY app for citizen in Dublin and Bologna



Mobile STAR-CITY app for citizen in Dublin and Bologna

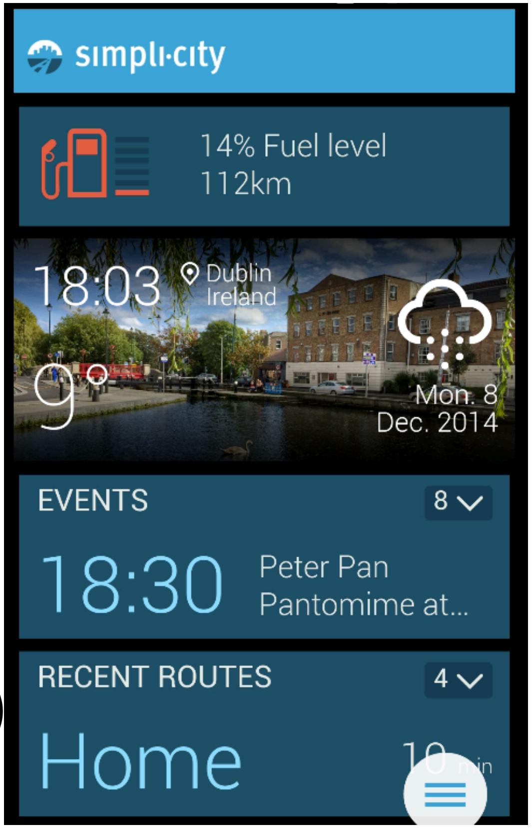
Context

Use case scenario: Meeting the Increased Mobility Demand:

- Scenario 1 “Road Traffic Diagnosis”
- Scenario 2 “Road Traffic Prediction”
- Scenario 3 “Personalized Traffic Restrictions”

Outcome:

- One mobile app
- Two pilot cities (Dublin and Bologna)
- Live and real-time environment
- Real data (user calendar, open data: traffic congestion, weather, events, road works, accident …)
- but with simulated car sensor data – Austin ;-)

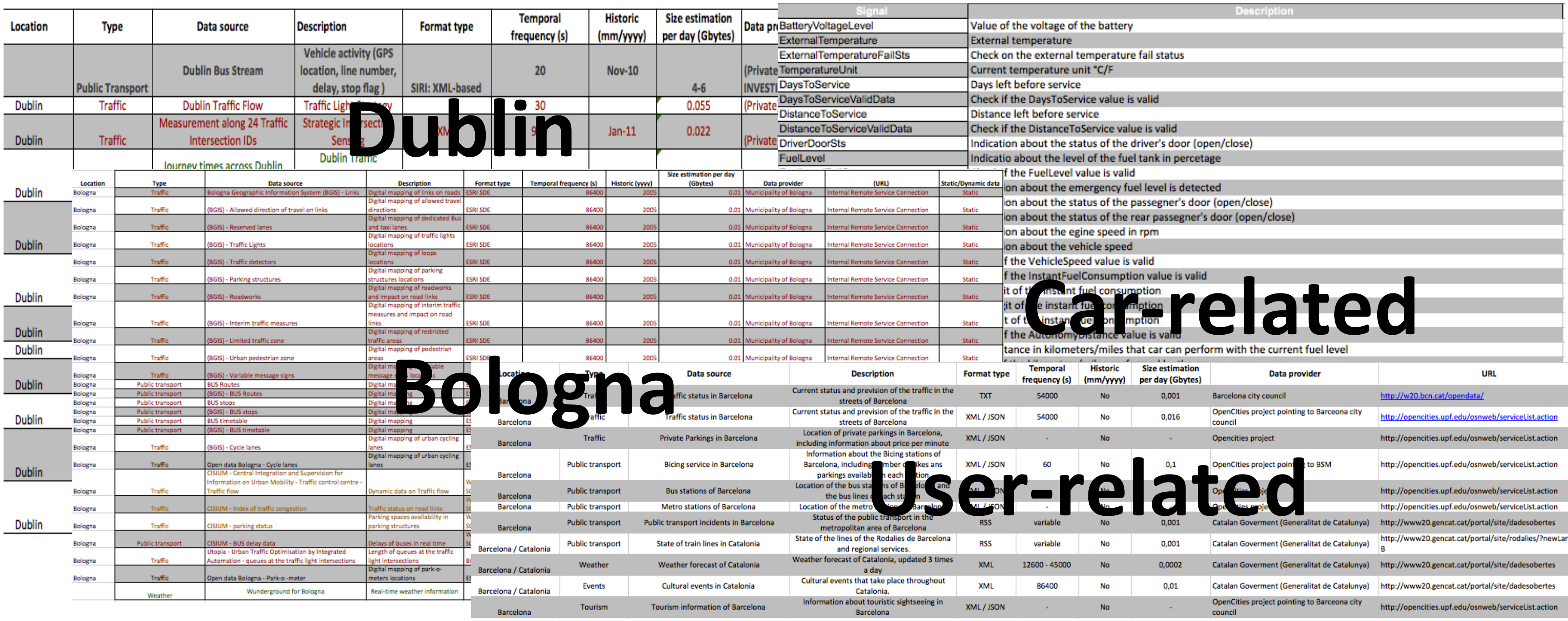


accenture Technology Labs

IJCAI-16

Mobile STAR-CITY app for citizen in Dublin and Bologna

Data



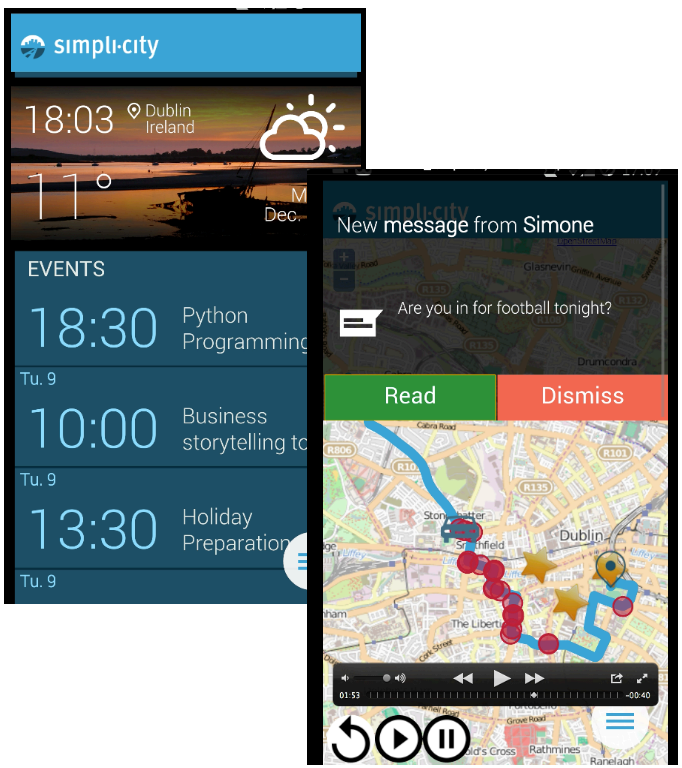
Vocabulary

Ontology	Size (KB)	#Concepts	#Object Properties	#Data Properties	#Individuals	Imported Ontologies
NASA SWEET[12] (IBM adaptation)	158.8	90	40	34	63	W3C Time, Geo
IBM Travel Time	4, 194	41	49	22	1, 429	
IBM SIRI-BUS	41.9	21	17	18	-	
W3C Time[9]	25.2	12	24	17	14	-
W3C Geo[10]	7.8	2	4	-	-	-
DBpedia	Only a subset is used for annotation i.e., 28 concepts, 9 data properties					

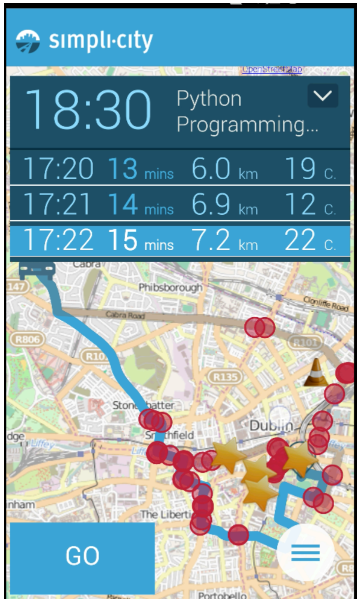
Mobile STAR-CITY app for citizen in Dublin and Bologna

Objective: Context-aware driving experience (1)

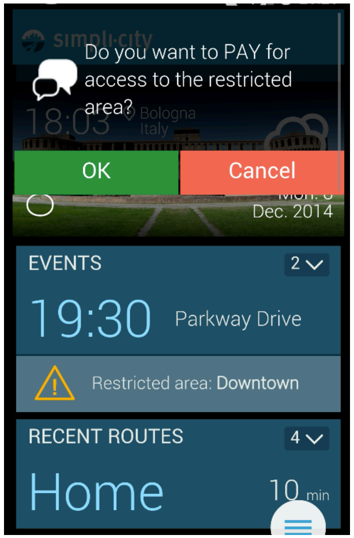
User Context-aware Driving (User data)



Open Context-aware Driving (Open data)



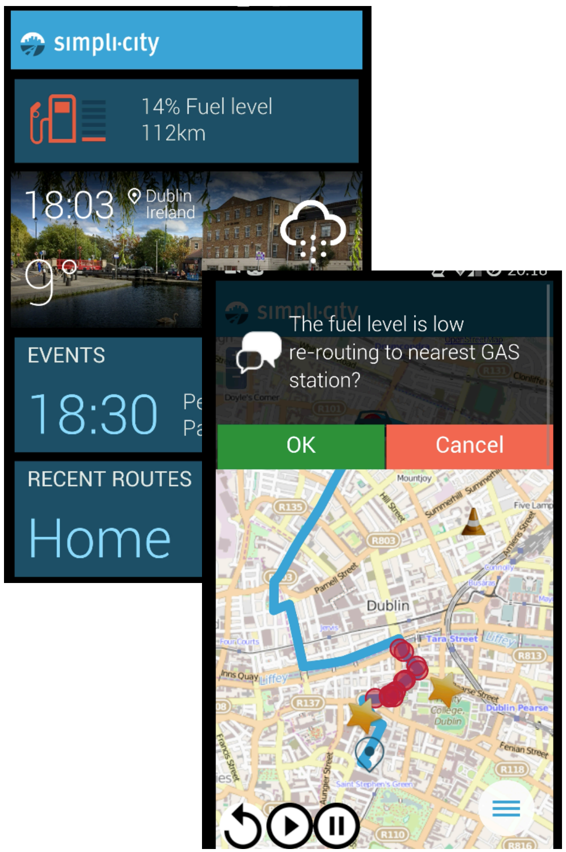
Private Context-aware Driving (City data)



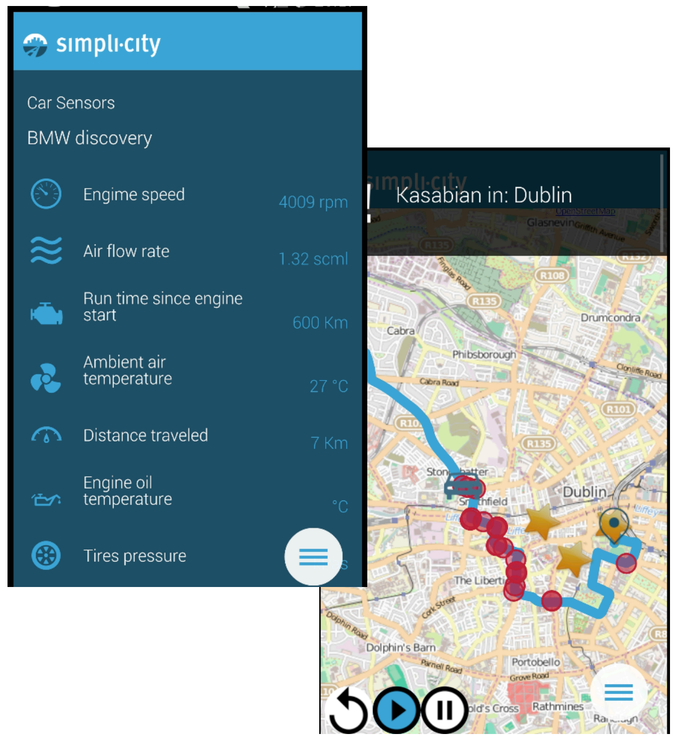
Mobile STAR-CITY app for citizen in Dublin and Bologna

Objective: Context-aware driving experience (2)

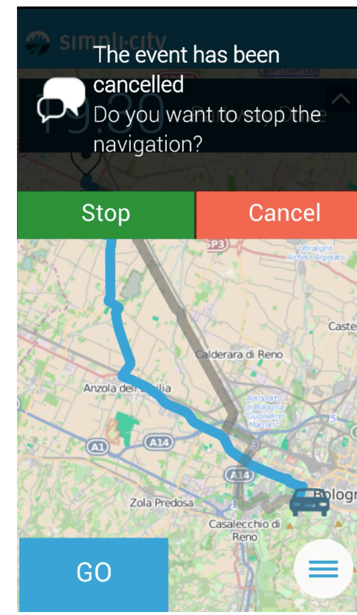
Car Sensor-aware Driving



Car Sensor Simulation



Personal Events-aware Driving

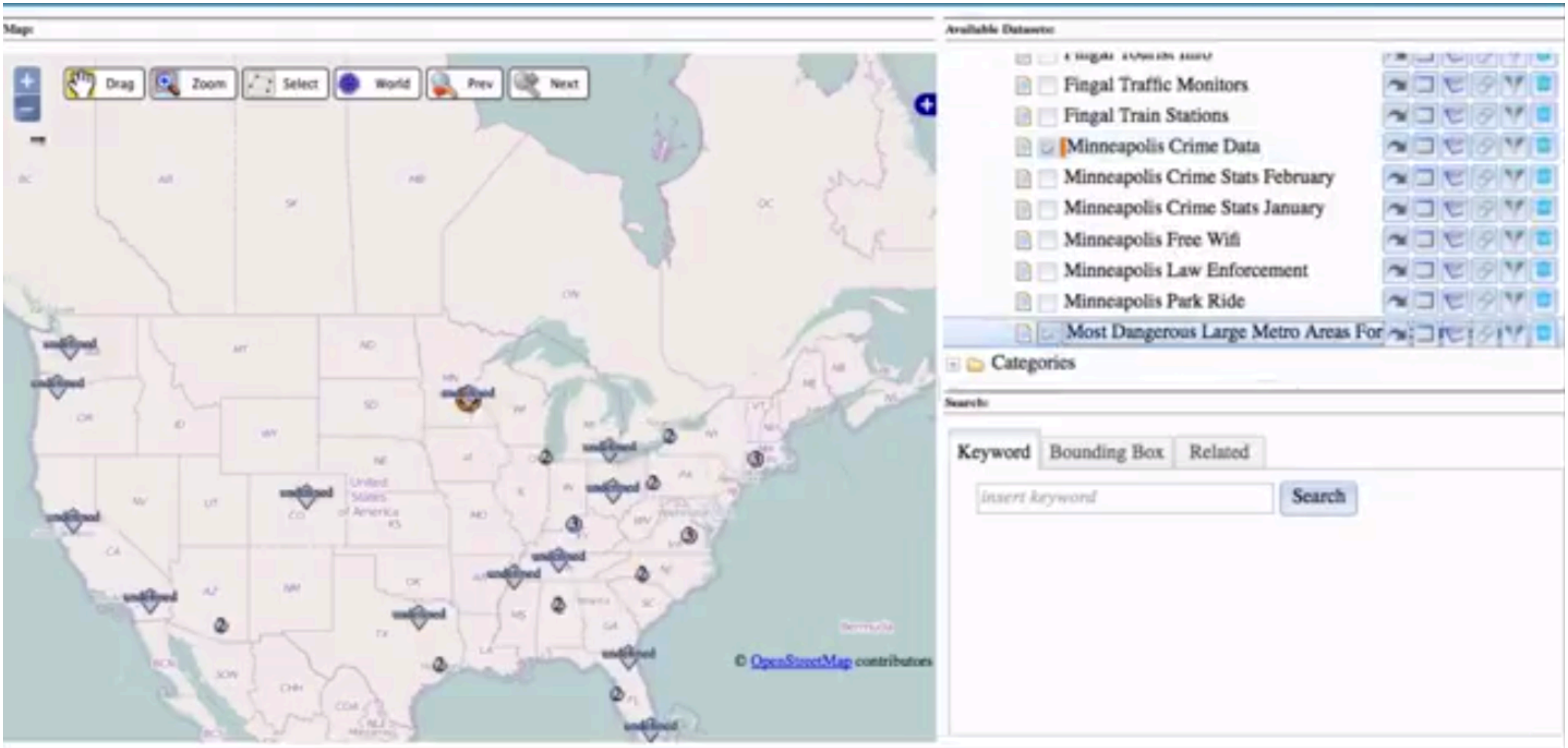


Real-Time Urban Monitoring in Dublin

Green: Dublin Bike availability
Purple dot: Bus in congestion
Blue: Noise
Purple bar: Pollution
Red: Amenities
Yellow: Cameras

https://www.youtube.com/watch?v=ImTI0jm3OEw

Semantic Processing of Urban Data



https://www.youtube.com/watch?v=lrUHet5awzw&feature=youtu.be

http://50.97.192.242:8080/Dali/

PART IV

"Finally, in conclusion, let me say just this."

Peter Sellers

Conclusion

Cities are characterized by:

- Big Data
- Complex Systems
- Integrated Problems
- Scalability Challenges

Cities want to be Smarter:

- More efficient
- More reliable
- More secure
- More open
- Cheaper
- Faster
- More integrated
- More citizen-centric
- More attractive
- More Intelligent
- Sustainable
- Better city planning

Cities can benefits from

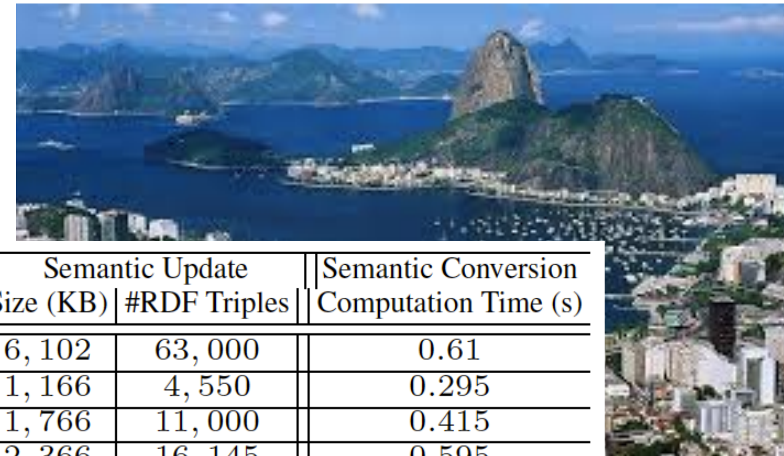
- REAL World data: Open Data to get Smarter
- Advances in AI

AI already helped a lot!! ... and should even contribute further

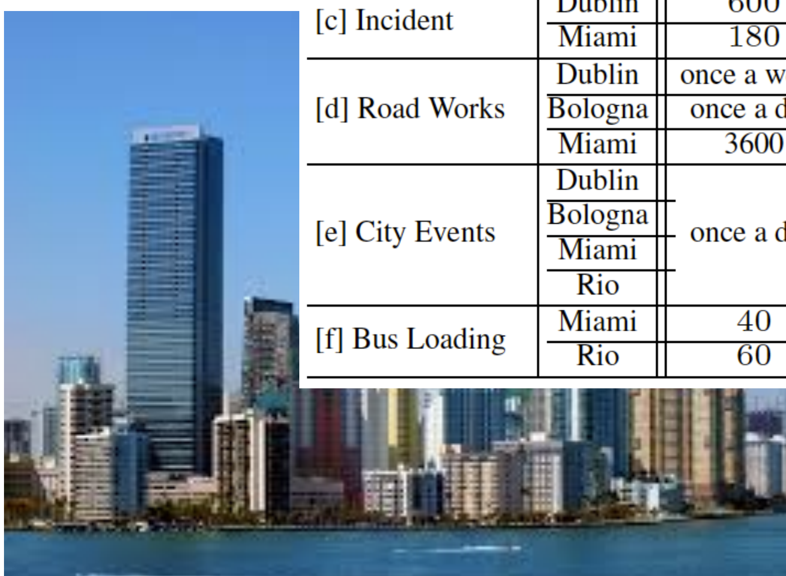
- Optimization, coordination ...

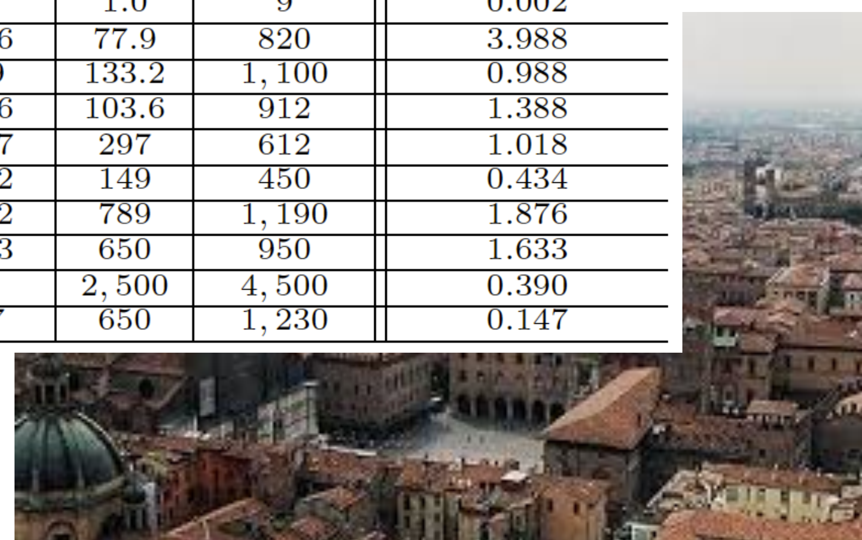
Future Work

Analytics and Reasoning: Scalability from One City to Another One



Real Time, Live Data	City	Frequency of Update (s)	Raw Update Size (KB)	Semantic Update Size (KB)	Semantic Update #RDF Triples	Semantic Conversion Computation Time (s)
[a] Journey Times	Dublin	60	20.2	6,102	63,000	0.61
[b] Bus	Bologna	120	31.8	1,166	4,550	0.295
	Miami	40	66.8	1,766	11,000	0.415
	Rio	60	96.8	2,366	16,145	0.595
[c] Incident	Dublin	600	0.2	1.0	7	0.002
	Miami	180	0.2	1.0	9	0.002
[d] Road Works	Dublin	once a week	146.6	77.9	820	3.988
	Bologna	once a day	78.9	133.2	1,100	0.988
	Miami	3600	102.6	103.6	912	1.388
[e] City Events	Dublin	once a day	240.7	297	612	1.018
	Bologna		111.2	149	450	0.434
	Miami		637.2	789	1,190	1.876
	Rio		585.3	650	950	1.633
[f] Bus Loading	Miami	40	833	2,500	4,500	0.390
	Rio	60	69.7	650	1,230	0.147

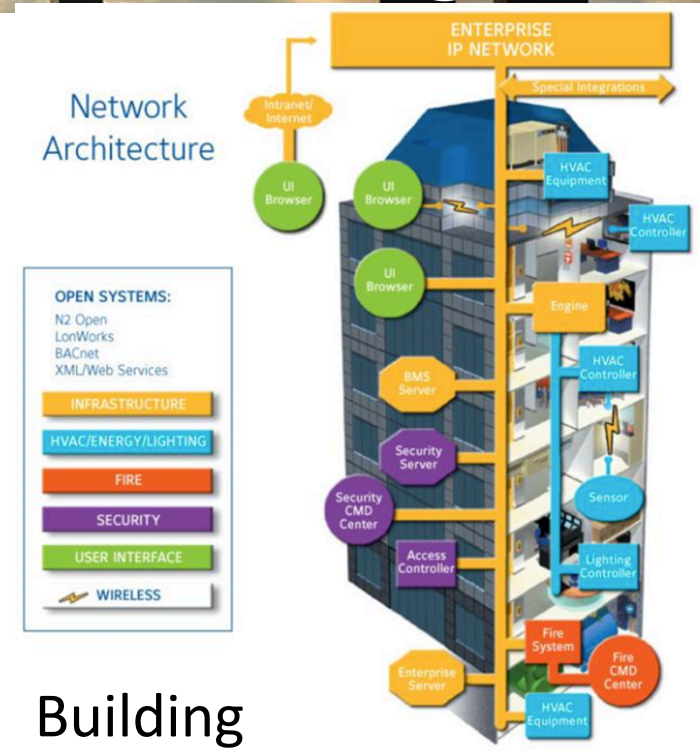




Future Work

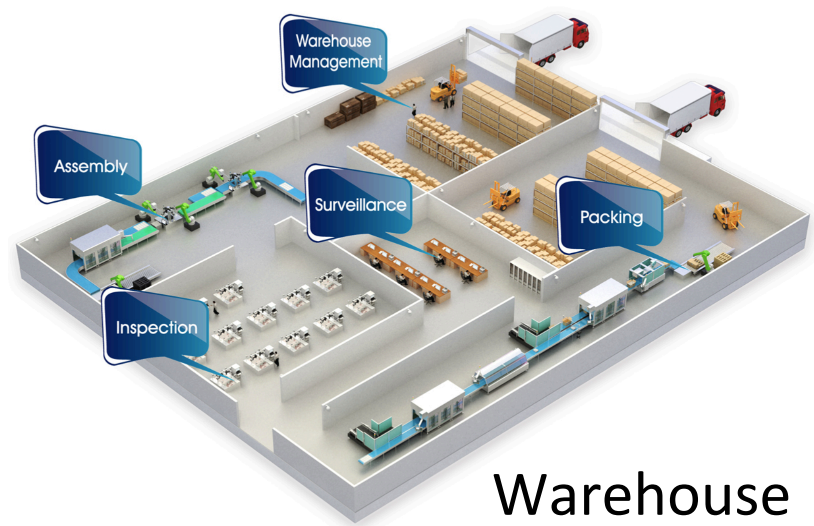
Application: From Cities to mini-Cities

Airport



Building

Supply-Chain



Warehouse

Future Work

More Multi-disciplinary: AI (Planning, KRR, ML, …), Database, Mathematics …

More Science Integration

References (1)

- **Freddy Lecue, Jeff Z.Pan.** Consistent Knowledge Discovery from Evolving Ontologies. AAAI 2015. Austin, Texas, USA, January, 2015.
- **Anika Schumann, Freddy Lecue.** Minimizing User Involvement for Accurate Ontology Matching Problems. AAAI 2015. Austin, Texas, USA, January, 2015.
- **Freddy Lecue, Simone Tallevi-Diotallevi, Jer Hayes, Robert Tucker, Veli Bicer, Marco Sbodio, Pierpaolo Tommasi**. Smart traffic analytics in the semantic web with STAR-CITY: Scenarios, system and lessons learned in Dublin City. In Jour- nal of Web Semantics. Web Semantics: Science, Services and Agents on the World Wide Web (JWS). Vol XX (2014).
- **Spyros Kotoulas, Vanessa Lopez, Raymond Lloyd, Marco Luca Sbodio, Freddy Lecue, Martin Stephenson, Elizabeth M. Daly, Veli Bicer, Aris Gkoulalas-Divanis, Giusy Di Lorenzo, Anika Schumann, Pol Mac Aonghusa**. SPUD - Semantic Pro- cessing of Urban Data. In Journal of Web Semantics. Web Semantics: Science, Services and Agents on the World Wide Web (JWS), Vol 24 (2014).
- **Freddy Lecue, Robert Tucker, Simone Tallevi-Diotallevi, Rahul Nair, Yiannis Gkoufas, Giuseppe Liguori, Mauro Borioni, Alexandre Rademaker and Luciano Barbosa.** Semantic Traffic Diagnosis with STAR-CITY: Architecture and Lessons Learned from Deployment in Dublin, Bologna, Miami and Rio, ISWC 2014, Riva, Italy, October 2014 (Best In Use Paper Award)
- **Joern Ploennigs, Anika Schumann, Freddy Lecue. Adapting Semantic Sensor Networks for Smart Building Diagnosis**. In Proceedings of the 13th International Semantic Web Conference (ISWC 2014), pages 308-323, October 19-23, 2014, Riva del Garda, Italy. Springer 2014 ISBN 978-3-319-11914-4. (Best In Use Paper Award Nominee)
- **Jiewen Wu, Freddy Lecue.** Towards Consistency Checking over Evolving Ontologies. Proceedings of the 23rd ACM Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014.
- **Anika Schumann, Freddy Lecue, Joern Ploennigs**. Exploiting the Semantic Web for Systems Diagnosis. In Proceedings of the Twenty-First European Conference on Arti cial Intelligence (ECAI 2014), pages ??-??, August 18-22, 2014, Prague, Czech Republic. IOS Press 2014.

References (2)

- **Joern Ploennigs, Anika Schumann, Freddy Lecue**. Extending Semantic Sensor Networks for Automatically Tackling Smart Building Problems. In Proceedings of the Twenty-First European Conference on Arti cial Intelligence (ECAI 2014), pages ??-??, August 18-22, 2014, Prague, Czech Republic. IOS Press 2014.
- **Freddy Lécué**: Towards Scalable Exploration of Diagnoses in an Ontology Stream. AAAI 2014: 87-93
- **Freddy Lécué, Robert Tucker, Veli Bicer, Pierpaolo Tommasi, Simone Tallevi-Diotallevi, Marco Luca Sbodio**: Predicting Severity of Road Traffic Congestion Using Semantic Web Technologies. ESWC 2014: 611-627 (Best In Use Paper Award)
- **Freddy Lécué, Simone Tallevi-Diotallevi, Jer Hayes, Robert Tucker, Veli Bicer, Marco Luca Sbodio, Pierpaolo Tommasi**: STAR-CITY: semantic traffic analytics and reasoning for CITY. IUI 2014: 179-188
- **Simone Tallevi-Diotallevi, Spyros Kotoulas, Luca Foschini, Freddy Lecue, Antonio Corradi**. Real-Time Urban Monitoring in Dublin Using Semantic and Stream Technologies. In Proceedings of the 12th International Semantic Web Confer- ence (ISWC 2013), pages 178-194, October 21-25, 2013, Sydney, NSW, Australia. Springer 2013 Lecture Notes in Computer Science ISBN 978-3-642-41337-7.
- **Freddy Lecue, Jeff Z, Pan.** Predicting Knowledge in an Ontology Stream. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2013)
- **Elizabeth M. Daly, Freddy Lecue, Veli Bicer.** Westland Row Why So Slow? Fusing Social Media and Linked Data Sources for Understanding Real-Time Traffic Conditions. In Proceedings of the ACM 2013 International Conference on Intelligent User Interfaces (IUI 2013)
- **Freddy Lecue**, **Anika Schumann**, **Marco Luca Sbodio**. Applying Semantic Web Technologies for Diagnosing Road Traffic Congestions. In Proceedings of the 11th International Semantic Web Conference (ISWC 2012), Springer (Best In Use Paper Award Nominee)
- **Freddy Lecue**: Diagnosing Changes in An Ontology Stream: A DL Reasoning Approach. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI 2012), July 22-26, 2012, Toronto, Ontario, Canada. AAAI Press 2012.

Questions

Thank you!

Credits: Pol Mac Aonghusa, Luciano Barbosa, Veli Bicer, Antonio Corradi, Elizabeth Daly, Luca Foschini, Yiannis Gkoufas, Jer Hayes, Vanessa Lopez, Rahul Nair, Joern Ploennigs, Alexandre Rademaker, Marco Luca Sbodio, Anika Schumann, Martin Stephenson, Simone Tallevi-Diotallevi, Pierpaolo Tommasi, Robert Tucker, Yuan Ren