

Handling City Data Deluge

Challenges and Applications

Veli Bicer

IBM Research, Ireland



Outline

- A Planet of Smarter Cities
- City Data and Information
- Challenges
- Applications
- Cloudy Cities
- Conclusion



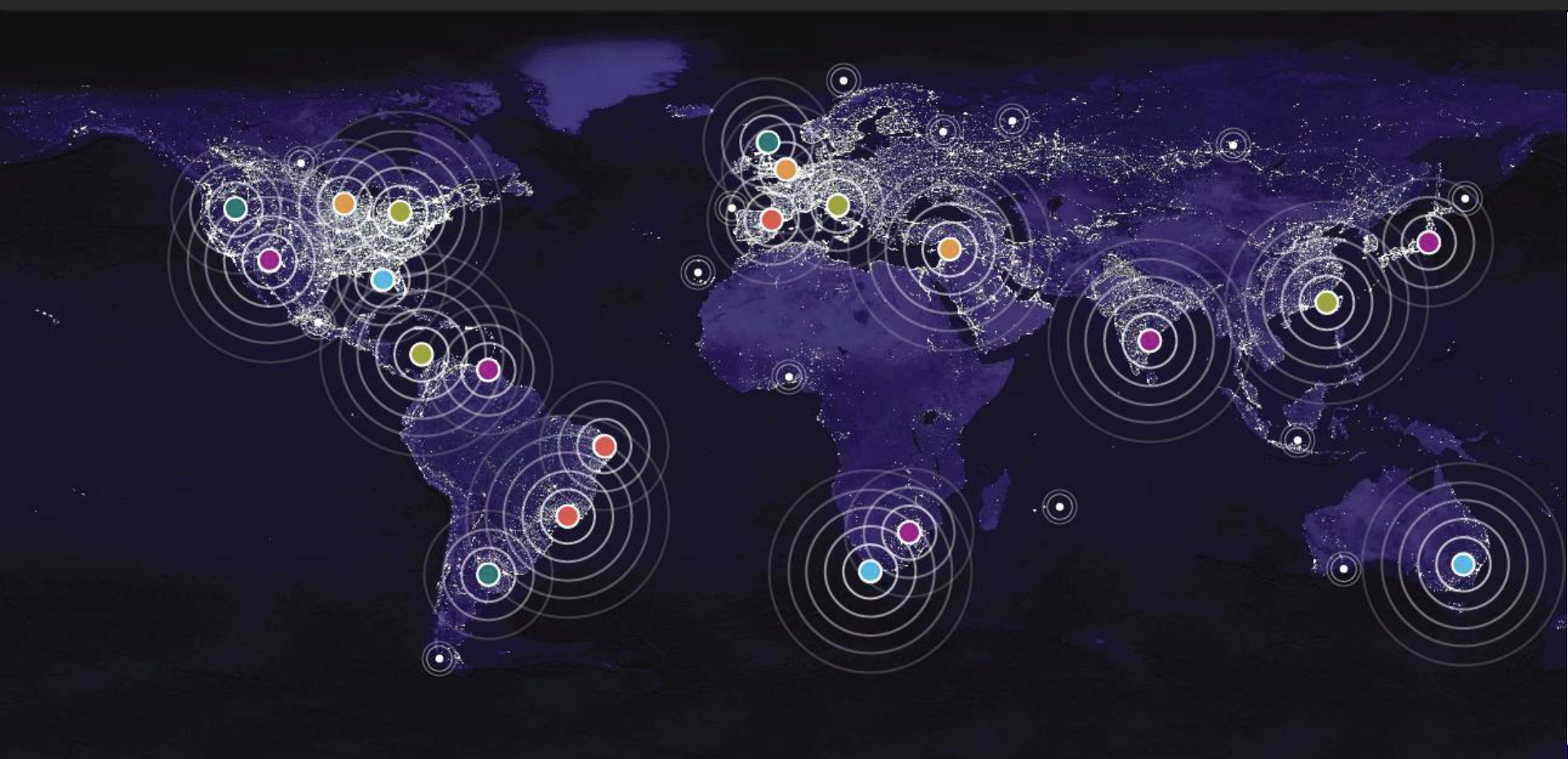
A Planet of Smarter Cities

“Cities have the capability of providing something for everybody, only because, and only when, they are created by everybody.”

Jane Jacobs



A planet of smarter cities: In 2007, for the first time in history, the majority of the world's population—3.3 billion people—lived in cities. By 2050, city dwellers are expected to make up 70% of Earth's total population, or 6.4 billion people.



IBM Research Worldwide

12 Labs. 6 Continents.

Smarter Cities
Risk Analytics
HPC



IBM Research – Ireland: Mission

Smarter Cities



- Transportation
- Water
- Energy
- City Fabric
- Mobility
- Social Care

Analytics



- Risk Model Creation
- Efficient Decision Model Solvers
- Risk Communication
- City Analytics

HPC



- Exascale workload optimized systems
- Big+fast data and aggregate cloud workloads

Machine Learning

Data Mining

Geospatial Visualization

Robust Control

Transportation Science

Automated Reasoning

Social Semantic Web

Water Management

Power Systems

Expertise

Optimization

Systems Software

Parallel Algorithms

Workload Optimization

Real-time Stream Processing

Networking

Distributed Simulation



City Data and Information

“The country places and the trees don’t
teach me anything, but the people in the
city do”

Socrates





Claudio Munoz



City of Data and Information: Many Areas

- Large, open and continuous data environment from heterogeneous domains:



Energy Management



City Management



Transportation



Social Media



Supply Chain



Region



Food System



HealthCare



Water
Management

and even more...



Some Traffic-related Data Sets



Data Source	Description	Format Type	Temporal Frequency (s)	Historic (mm/yyyy)	Size Estimation (GBytes)
Dublin Bus	Vehicle activity (GPS location, line number, delay, stop flag)	SIRI (based on XML)	20	01/2011	4-6 per day
Dublin Traffic Flow Measurement along 24 Traffic Intersection IDs	Traffic Light Strategy	XML	30		0.055 per day
	Strategic Intersection Sensing		900		0.022 per day
Dublin Bike	Occupancy of bicycle on 44 stations	XML (through Web access)	600		0.02 per day
CCTV Monitoring	Real-Time monitoring of Dublin City	Stream Video	Real Time	No (privacy reasons)	under investigation
Pedestrians counter	Number of pedestrian (several sidewalk)	CSV, XML	Any	under investigation	Frequency-dependent
Wunderground for Dublin	Real-time weather information	CSV	[5, 600] (depending on stations)	11/2010	[0.050, 1.5] per day (depending on stations)
Road Weather Condition		under investigation			

- Big data
- Heterogeneous data
- Static, Continuous data
- Not all open yet,
- Not linked yet
- Noisy data (inconsistent, imprecise)



www.dublinked.ie

POWERED by
Open Innovation Portal

Dublinked™

Datastore

Apps

Activities

Members



 Traffic Volumes- City Centre Bridges Count Data 2007-2010	Transportation & Infrastructure
 Traffic Volumes- Canal Cordon Count Data 2008-2010	Transportation & Infrastructure
 Fats& Oils & Greases (FOG) Licences Register	Transportation & Infrastructure: Environment
 On Street Disabled Parking Bay in Dublin City Council area	Transportation & Infrastructure
 2011-2017 Dublin City Development Plan - Parking Zones	Zoning & Land Use
 Dublin City Council Ordnance Survey Ireland (OSI) Map Base 2011	Zoning & Land Use
 Air Quality Monitoring Traffic Projects -Dublin Port Tunnel(DPT)	Environment
 Paper and Plastic Recycling; Weights collected at Dublin City Bring Centres	Environment



Dublinked - outcomes

- Publish and put into context (100's datasets, 1000's of files)
- Create innovation ecosystem

Property management



Commercial valuations and rates



Business & Retail



Housing myhome.ie



Mapping



DECISIONMAP.IE

Demographics



Crime



Health



Transport & Access



Pool resources
Share results



Dublin Tourism



Events



Culturefox.ie

Heritage



Environment



Waste Collection



Fault Reporting



Challenges

“We cannot afford merely to sit down and deplore the evils of city life as inevitable, when cities are constantly growing, both absolutely and relatively. We must set ourselves vigorously about the task of improving them; and this task is now well begun.”

Theodore Roosevelt



Smarter Cities share data ... Open Urban Data is at the center of a new wave of opportunity (*)



- More than 150 city agencies and authorities, worldwide, have already made over 1M datasets available through open data portals.
- Open data are generating new business: McKinsey & Associates estimate the economic value of big, open health data, at approximately \$350B annually.



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



City of Chicago Data Portal

Search & Browse Datasets and Views

Name	Description
1. Current Employee Names, Salaries, and Position Titles	Administration & Finance personnel This dataset is a listing of all current City of Chicago employees, complete with full names, departments, positions, and annual salaries.
2. CTA - Map of Fare Media Sales Outlets	cta, chicago transit authority
3. Crimes - Map	Public Safety crime, police This dataset reflects reported incidents of crime that have occurred in the City of Chicago over the past year, minus the most recent sev
4. Building Permits	Buildings permits Permits issued by the Department of Buildings in the City of Chicago from 2006 to the present. The dataset for each year contains more
5. Police Stations	Public Safety facilities, gis Chicago Police district station locations and contact information.
6. Food Inspections	Health & Human Services food, inspections This information is derived from inspections of restaurants and other food establishments in Chicago from January 1, 2010 to the presen



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



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

Community challenges	0	23
----------------------	---	----

4 V's of Big Data

Big city data

Volume

- Lots of relevant information
- Not linked to authoritative sources

Velocity

- Streams
- Frequent updates

Variety

- Different models and file formats
- Open domain - Unknown schema

Veracity

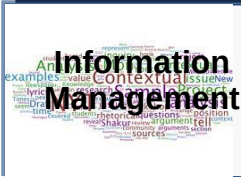
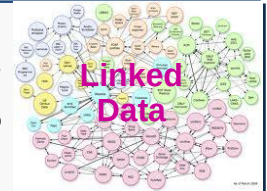
- Diverse sources
- Difficult to do assess quality



Research Streams

What would you do if you had access to all of the data in a City?

Could multiple sources of City data be linked together at scale to uncover new behaviours and provide new insights?



What technologies will enable contextual query across massive volumes of heterogeneous data, for applications and people?

How could we protect the City – and Citizens – from harm while still enabling insight?



How can we use computer reasoning to simplify City Operations through diagnosis and prediction?

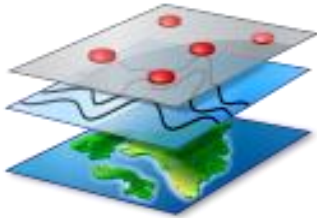


How can we incorporate human & social data sources to interpret and predict emergent behavior?



What do people search for?

Top 8 categories according to user scores [Kukka, PUC, 2013]



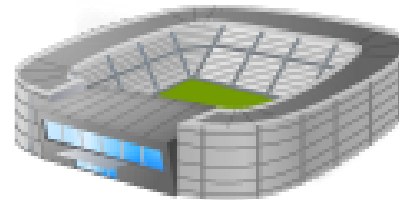
Maps

- Where places are and what's near me



Transport

- Public transportation schedules, location of transports etc.



Events

- What's happening today/tomorrow/next week



Food

- Restaurant menus, happy hours etc.



Info

- General information related to opening hours, local history, healthcare etc



Traffic

- Free parking spaces, construction sites, traffic jams etc.



Ads

- Offers from stores, where to buy etc.



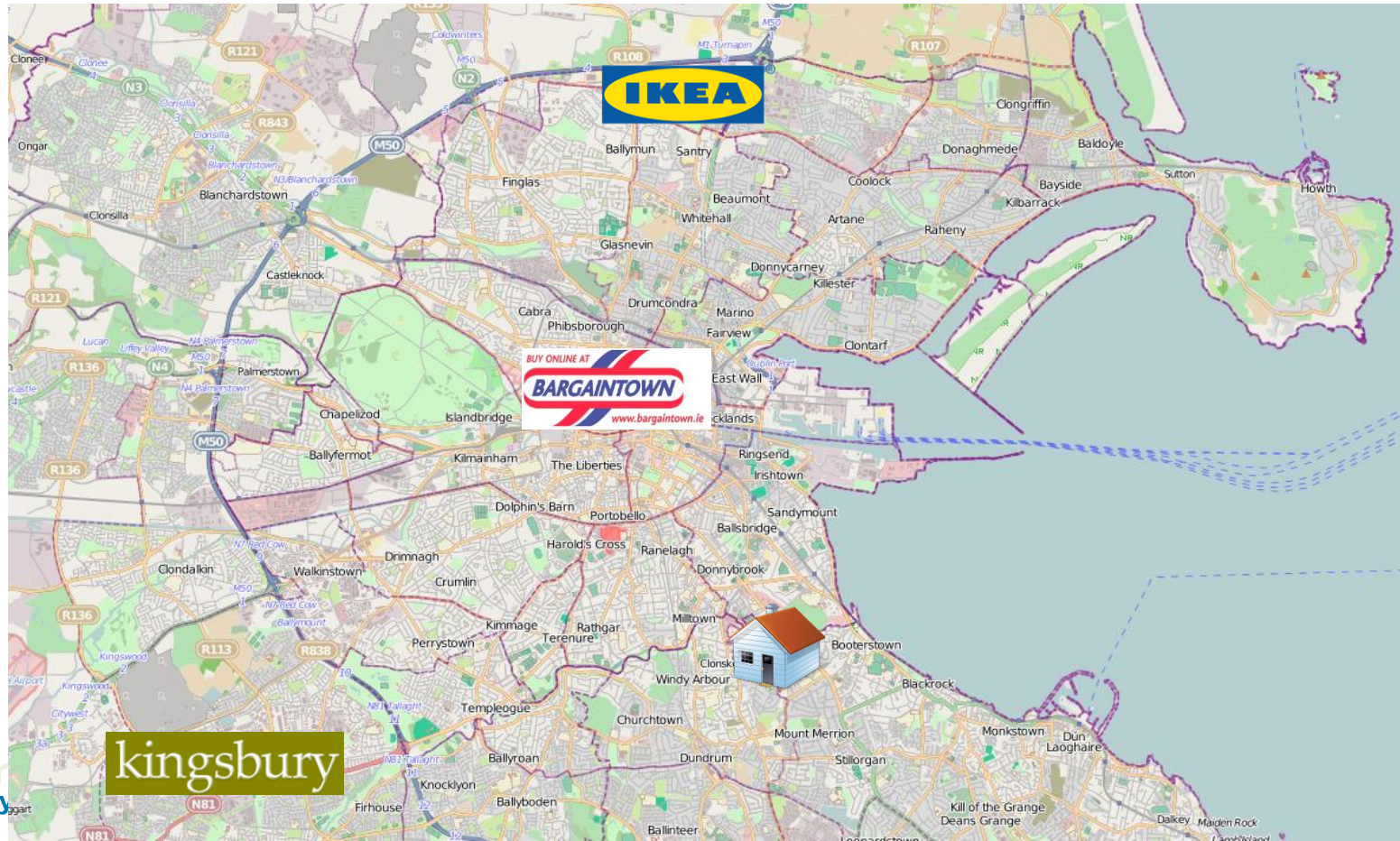
News

- News from national and international sources



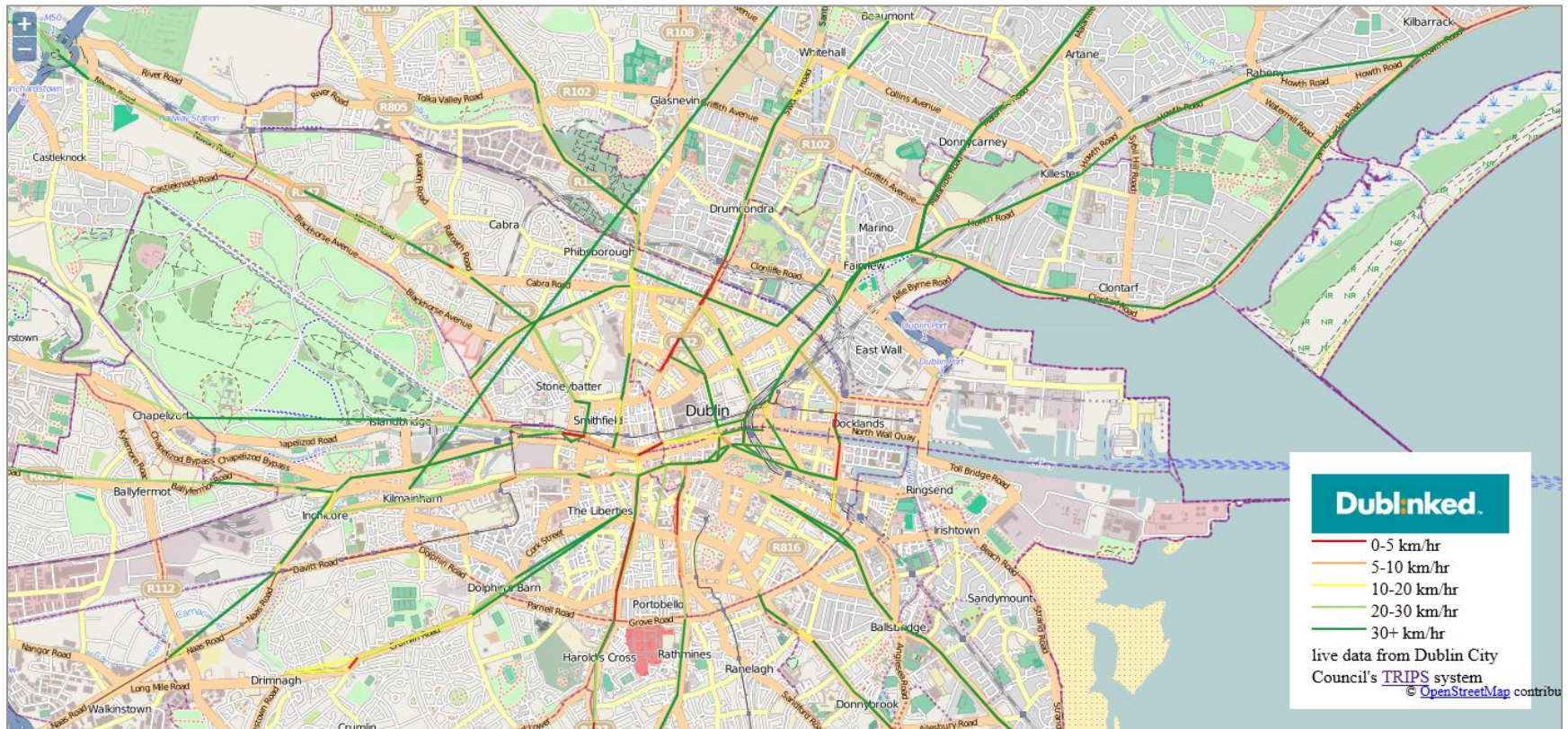
Relevance

- Need to buy new “furniture”?



Relevance

- Dublin TRIPS data:



Relevance

- Dublin Trips Data:
 - Journey times throughout the city
 - Real-time data with updates in every minute
 - Historical data is available for every day since 9/7/2012
 - **Mined from SCATS-based (Sydney Coordinated Adaptive Traffic System)** intelligent transportation system for 500+ sites around Dublin
- Accessible from:
 - <http://dublinked.ie/datastore/datasets/dataset-215.php>
- Visualization
 - <http://www.dublinked.ie/traffic/>



Relevance

- More transportation data
 - Public Transport Route Networks
 - <http://dublinked.ie/datastore/datasets/dataset-258.php>
 - Dublin Bus GPS Data
 - <http://dublinked.com/datastore/datasets/dataset-304.php>
 - Dublin Bus GTFS data
 - <http://dublinked.ie/datastore/datasets/dataset-254.php>
 - Accessible Parking Places
 - <http://dublinked.com/datastore/datasets/dataset-049.php>
 - Roads and Streets in Dublin City
 - <http://dublinked.com/datastore/datasets/dataset-123.php>



Relevance

Buying your dream house



Finding the houses?

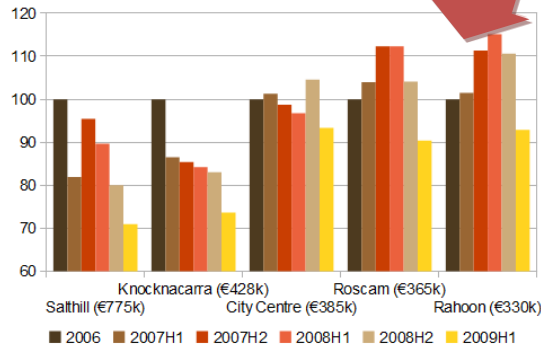


€235,000
Rathfarnham
Co. Dublin



€249,000
Dundrum
Co. Dublin

Perfect match!!



Is the price reasonable?

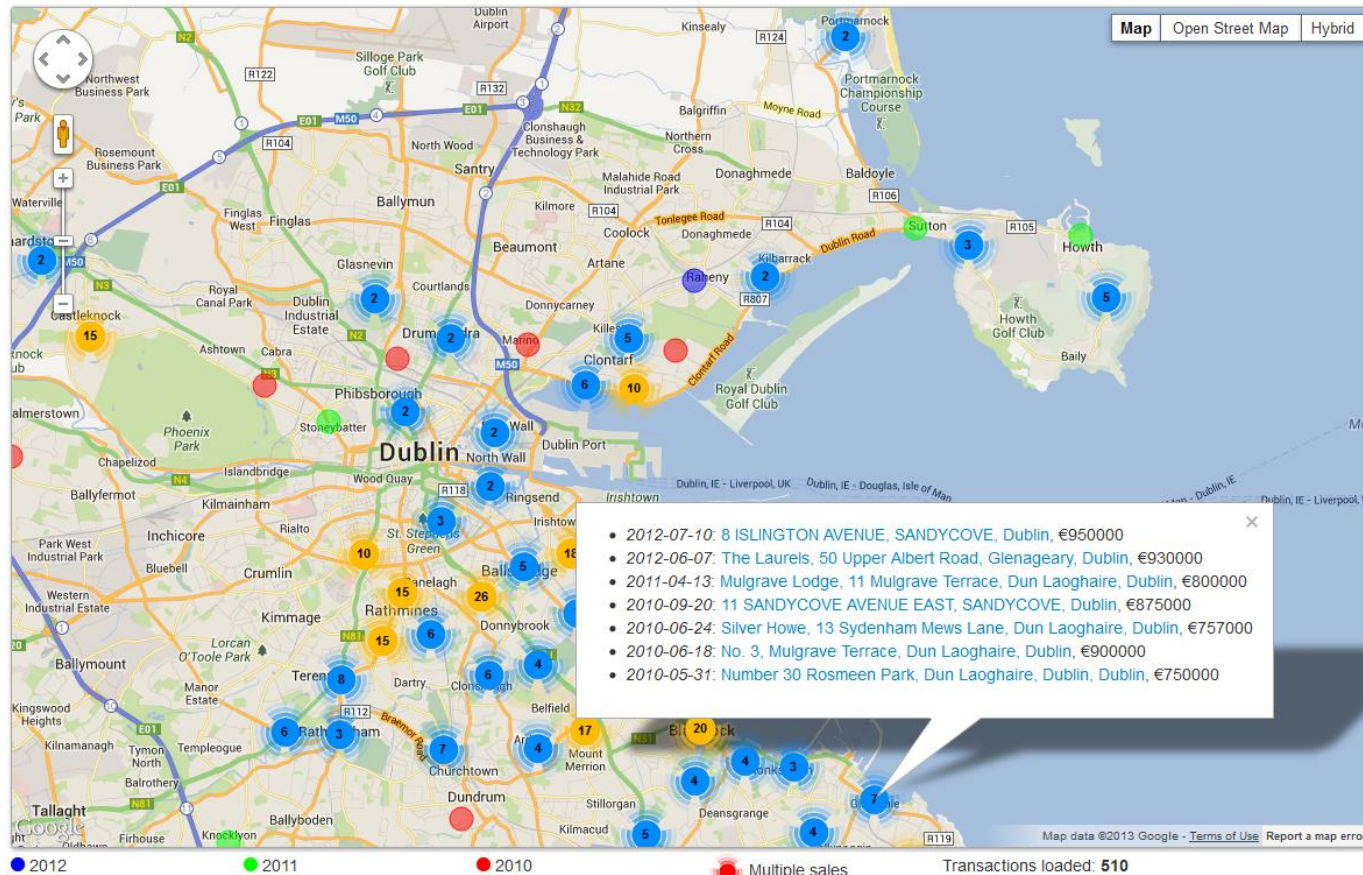


How is the neighborhood?

Relevance



Property Register Index : ~52000 property sales



Relevance

- More city data:
 - Amenities & Recreation
 - <http://dublinked.ie/datastore/by-category/amenities-recreation.php>
 - Schools
 - <http://dublinked.com/datastore/datasets/dataset-099.php>
 - Key developing areas
 - <http://dublinked.ie/datastore/datasets/dataset-134.php>
 - Air pollution monitoring data
 - <http://dublinked.ie/datastore/datasets/dataset-185.php>



Business case

- Why are ambulances late?

Sources of information

- 100's of datasets from four municipal authorities in Dublin
 - Most static, some dynamic
- Social Media: twitter, LiveDrive, eventful, eventBright, ...
- Linked Data: DBpedia, ..
- Vocabularies: IPSV, FOAF, VOID, PROV, DCAT, WSG



Domain of information

- Locations of Health Services
- Ambulance call outs and response times
- Tweets about traffic congestion
- Geo-located tweets about people movement
- Road network
- Event Web Services
- ...

We actually have this information. But we require novel search and exploration methods

Business case: traffic diagnosis

Problem: diagnosis and reasoning

How can we provide City decision makers with **explanations and diagnoses** for events by applying machine reasoning techniques to a fusion of massive, rich, complex and dynamic data? How can we move from explanation to prediction?

Challenges

- **Identifying** relevant data and information
- Capturing and **representing anomalies**
- **Correlating time-evolving knowledge** on heterogeneous data sources
- **Advanced fusion of 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: XML-based	20	11/2010	4-6	(Private) DCC
	Wunderground for Dublin	Real-time weather information	CSV	[5, 600] (depending on stations)	01/1996	[0.050, 1.5] (depending on stations)	(Public) Wunderground ^a
Source of Causes	Road Weather Condition (54 stations)		CSV	600	11/2010	0.1	(Public) NRA ^b
	Road Works and Maintenance		CSV	NA	2011	a few KBytes (need more!)	(Public) Dublinked ^c
	Events in Dublin	Events with small attendance Events with large attendance	XML	Not considered	02/2012 11/2011	0.001 0.05	(Public) Eventbrite ^d (Public) Eventful ^e
	DBpedia	Structured facts extracted from wikipedia	RDF	No	No	3.5 × 10 ⁶ concepts	(Public) DBpedia ^f

^a <http://www.wunderground.com/weather/api/>

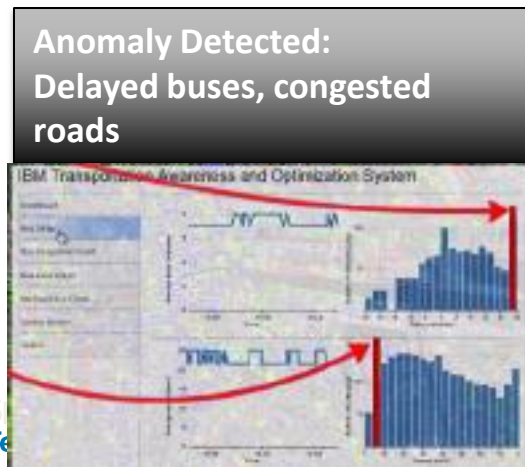
^b NRA - National Roads Authority <http://www.nrtraffic.ie/weather>

^c <http://dublinked.ie/>

^d <https://www.eventbrite.com/api>

^e <http://api.eventful.com>

^f <http://dbpedia.org/>



Detection to
Diagnosis?

Diagnosis:
A music concert next to
Canal Road at 3PM

Applications

*“True genius resides in the capacity for
evaluation of uncertain, hazardous, and
conflicting information.”*

Winston Churchill



Stream Data example

- **Context-based CCTV Camera Selection**

[Tallevi et al, ISWC'13]

- 100's CCTV cameras in Dublin.

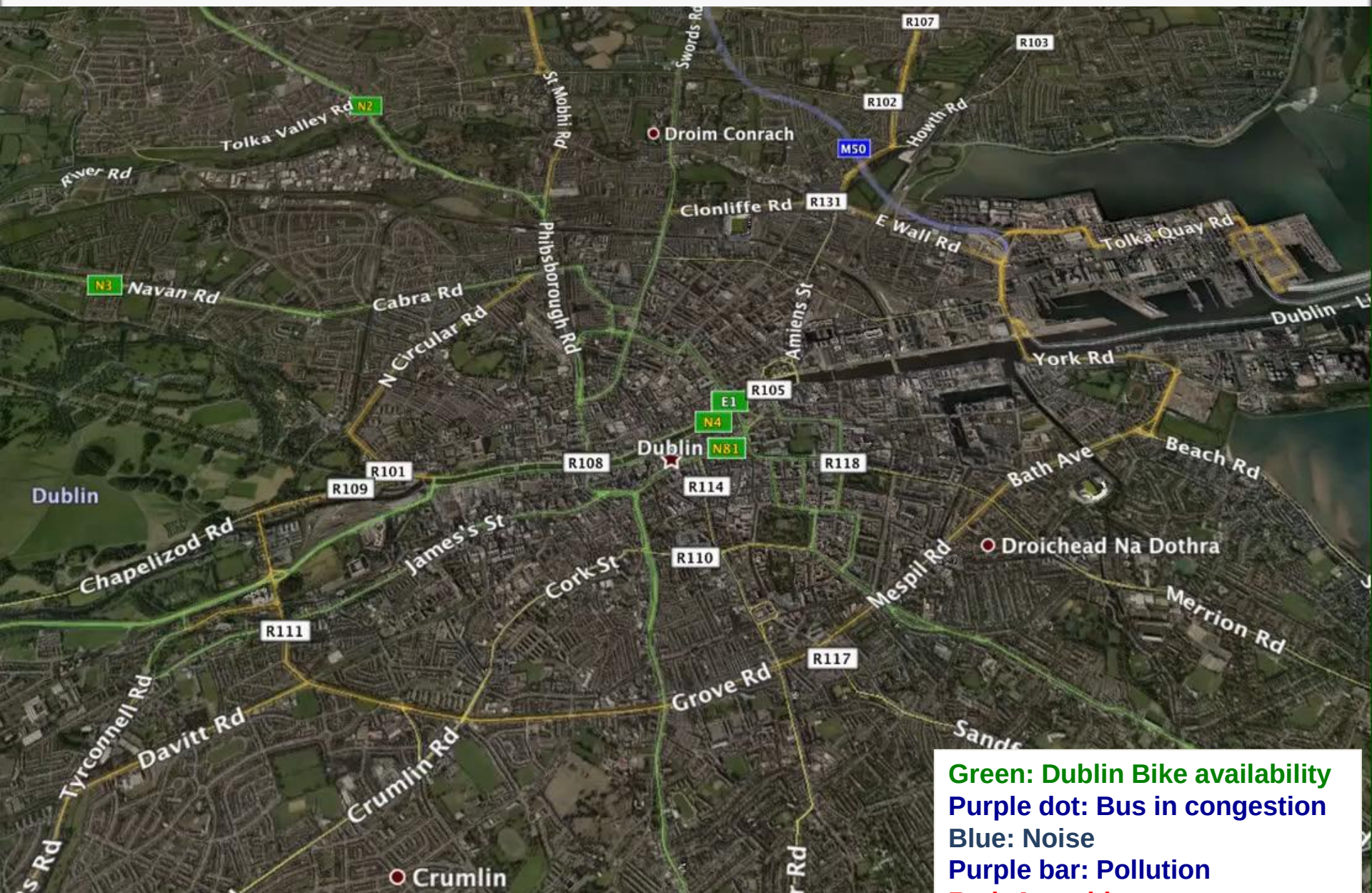
- Live and static context:

- Traffic
- Noise
- Pollution
- Amenities
- ...



- Continuous SPARQL interpreter, with extensions for heterogeneous data and execution engine on top of Infosphere Streams
- Live fusion of information to select **top-k most interesting cameras based on context.**





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

Fusing Data Streams from Dublin City to Select Surveillance Cameras

Simone Tallevi-Diotalle, Spyros Kotoulas, Freddy Lecue

Social Cities

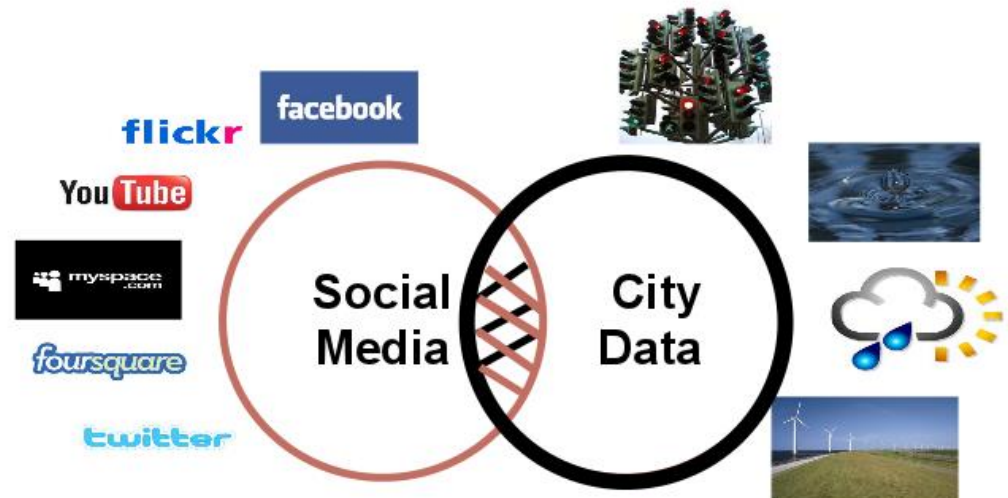
Our interaction with Cities is increasingly digital, these 'Citizen Signals' - including social media, human-system interactions and pervasive device traces - create a unique opportunity to close the loop between citizens and the City.

Problem: Social Cities insights

How can we use these insights to improve City Operations and Planning? Can we harness citizen engagement & social media to augment traditional information sources?

Citizen generated data to study urban dynamics:
[Kling et al, SIGSPATIAL GIS'12]

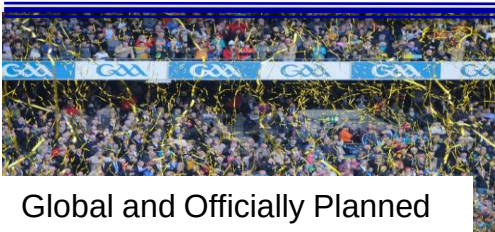
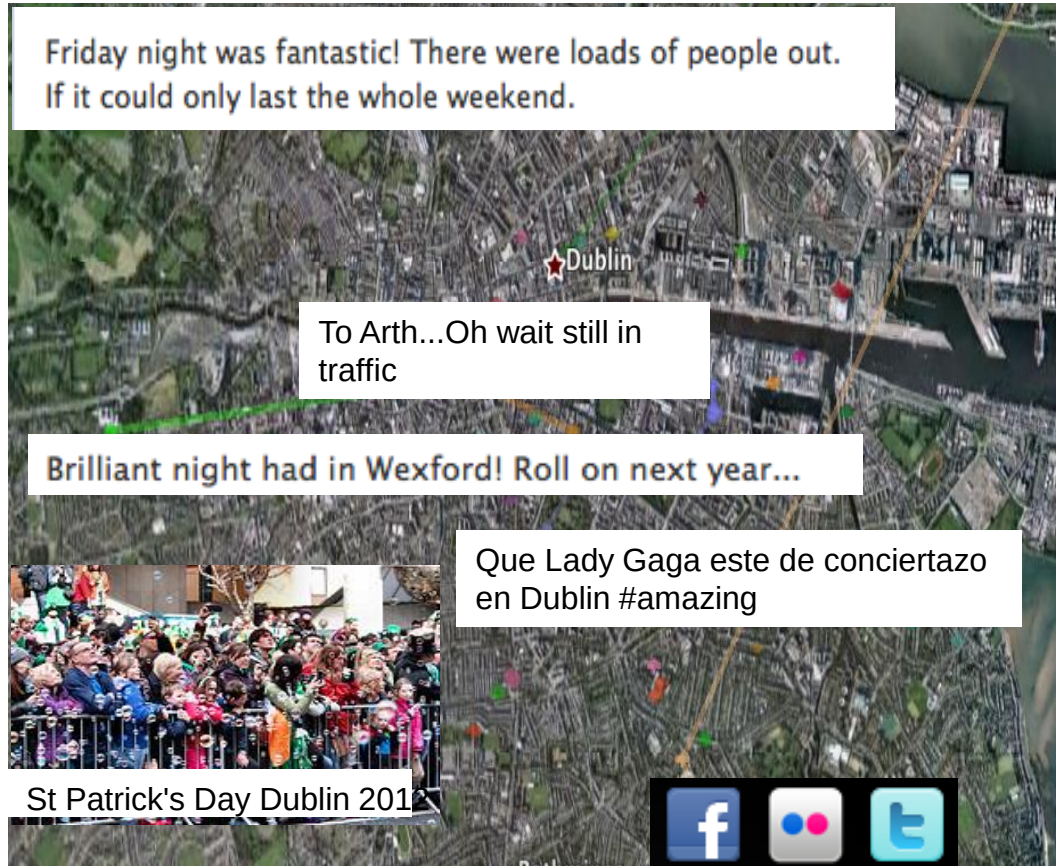
- Cluster urban areas based on topics
- Spatial-temporal topic distribution



Post-event analysis and characterization

[Di Lorenzo et al, MDM'13]

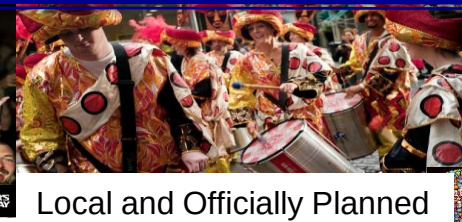
- Extract citizens' discussion topics and identify the relevant ones
- Discover correlation between discussion topics and events
- Study magnitude of events: what is their impact?
 - spatial/temporal profile;
 - estimate event's attendees;
 - mobility of event's attendees;
 - correlate their mobility patterns with the event evolution



Global and Officially Planned



Global and Unofficially Planned



Local and Officially Planned



Unofficially Planned

EXSED – Topics Extractors – Time Space

[Di Lorenzo et al, MDM'13]

Latent Dirichlet Allocation (LDA) principle

“Arts”	“Budgets”	“Children”	“Education”
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT

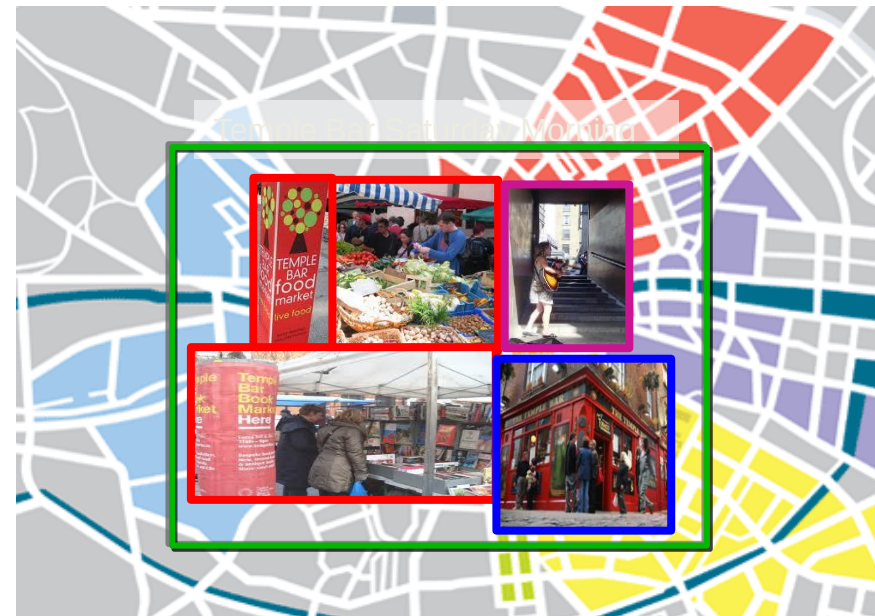
The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

Augmented trajectories from half million geo-located tweets from 11 Sep 2012 – 11 October 2012

userID	latitude	longitude	time	tags
--------	----------	-----------	------	------

LDA applied in a city scenario

Market	Music	Pub
food	nice	pub
soup	song	guinness
market	irish	temple
book	Busker	beer

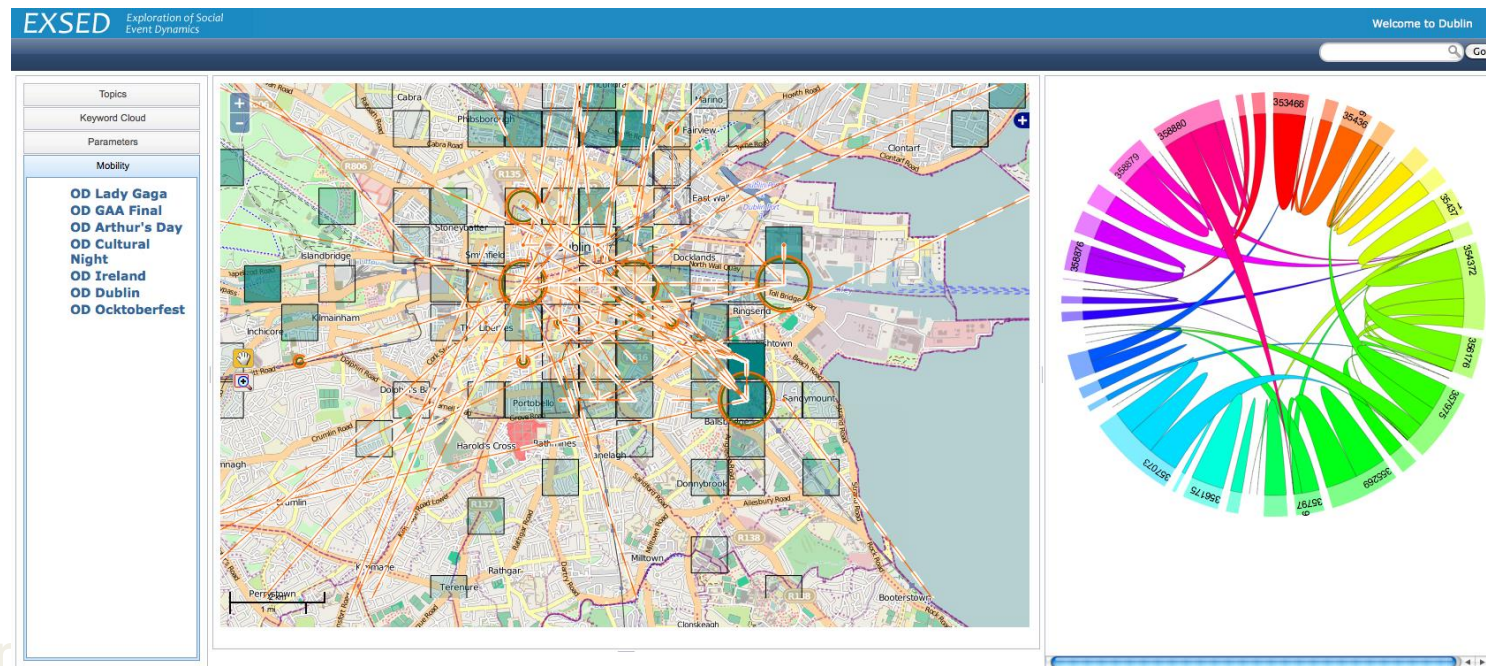


EXSED – Event evolution

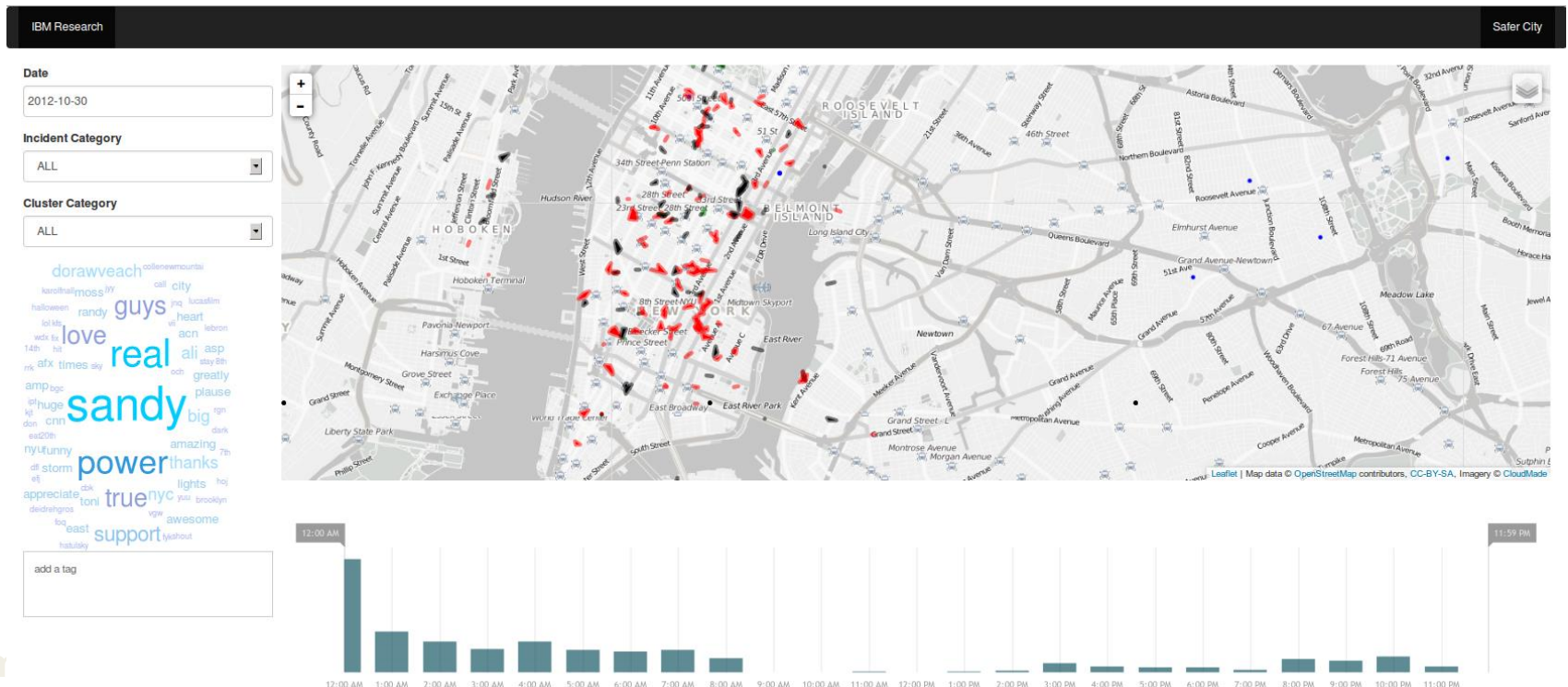
[Di Lorenzo et al, MDM'13]

- Filtering techniques – determining important places
 - Averaging location among consecutive measurements within a given spatial and temporal window [trajectory miner]
- Spatial and temporal profiles of a topic
- Mobility origin-destination matrix for event's attendees. Correlate mobility patterns with event evolution

Mobility exploration view:



- Identify and analyze public safety related incidents from social media
- Based on spatio-temporal clustering algorithm
- Improve situational awareness for potentially-unreported activities happening in a city.

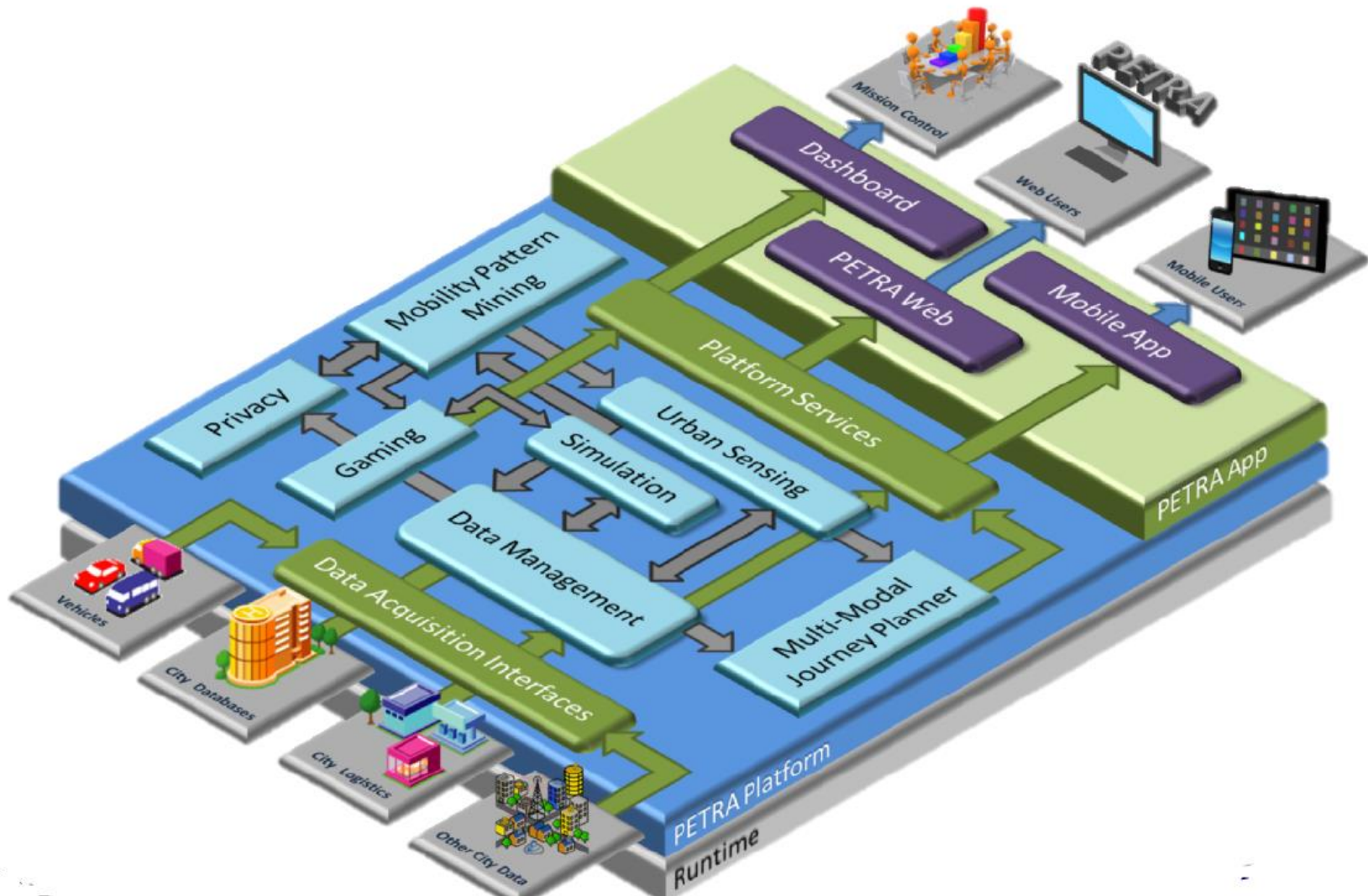


Managing Travels with PETRA: the Rome use case

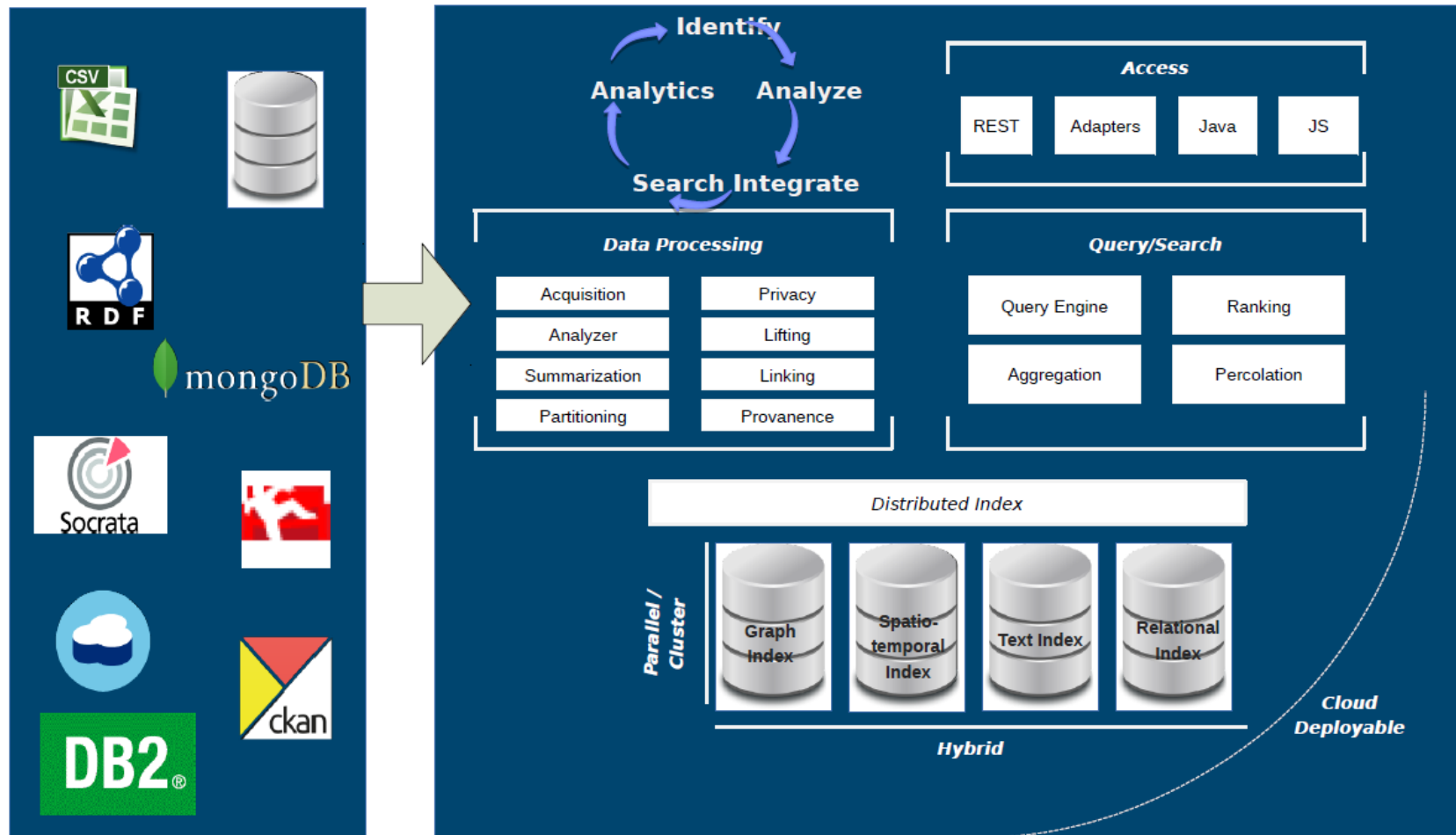
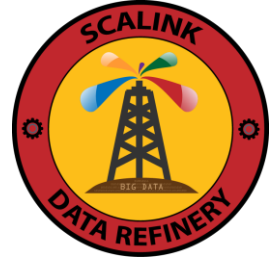
- PETRA FP7: Develop a platform connecting city mobility providers and controllers with the travelers in a way that information flows are optimized while respecting and supporting the individual freedom safety and security.
 - Integrated platform to enable the provision of citizen-centric, demand-adaptive city-wide transportation services.
 - Travelers will get mobile applications that facilitate them in making travel priorities and choices for route and modality.
- Our goal is to implement an independent module within the Petra platform which has the task of merging Roma data with KDDLab mobility patterns and providing them to the PETRA journey planner.



PETRA Architecture

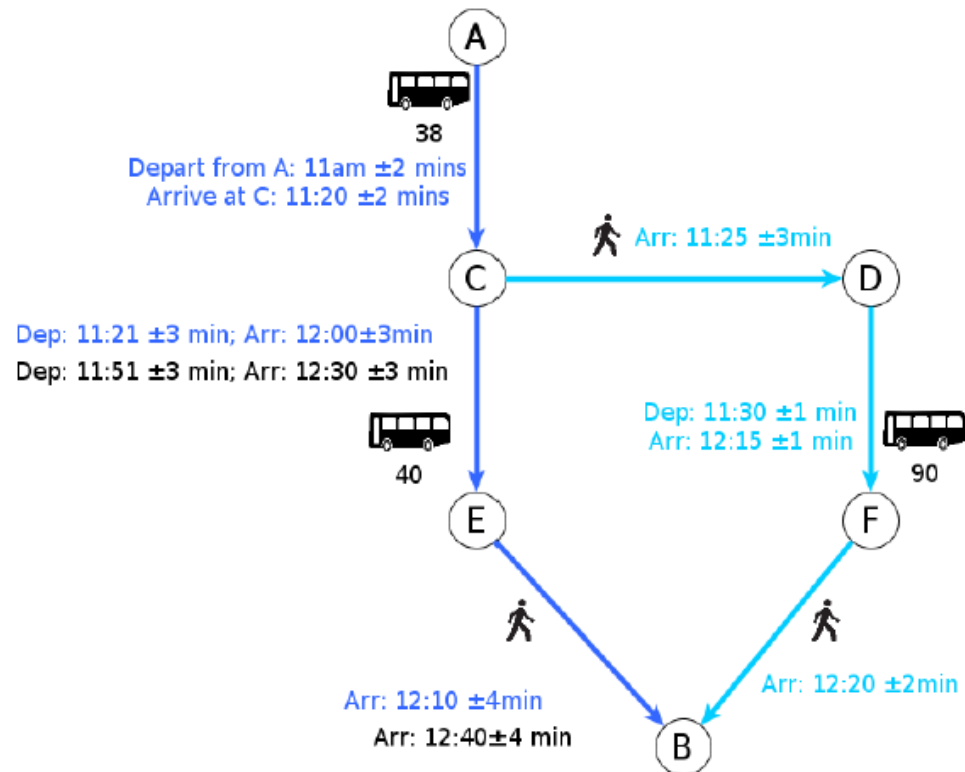


PETRA Data Management



Multi-modal Journey Planner

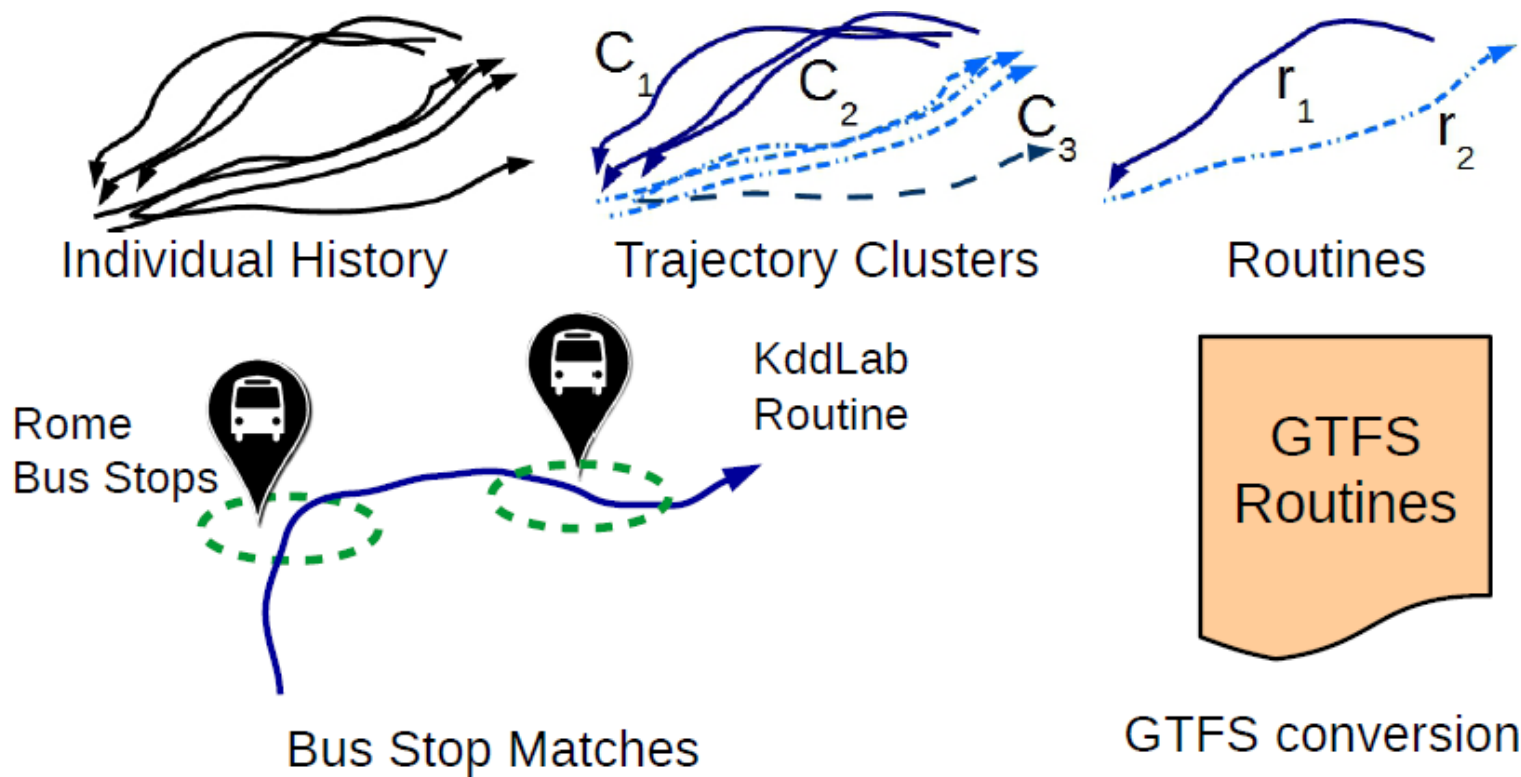
- Multi-modal travel
 - Combining diverse transport modes in one journey
 - One way of fighting congestions in cities
- Deterministic planning is the de-facto standard in deployed systems
- Real transportation networks feature several kinds of uncertainties (e.g. arrival times of public transport, congestions, etc)
- Using risk edging journey planner it is possible to optimize the users' journeys



	Linear trajectory A → C → E → B	Linear trajectory A → C → D → F → B	Contingent plan
Best case	12:10 ± 4min	12:20 ± 2min	12:10 ± 4min
Worst case	12:40 ± 4min	12:20 ± 2min	12:20 ± 2min

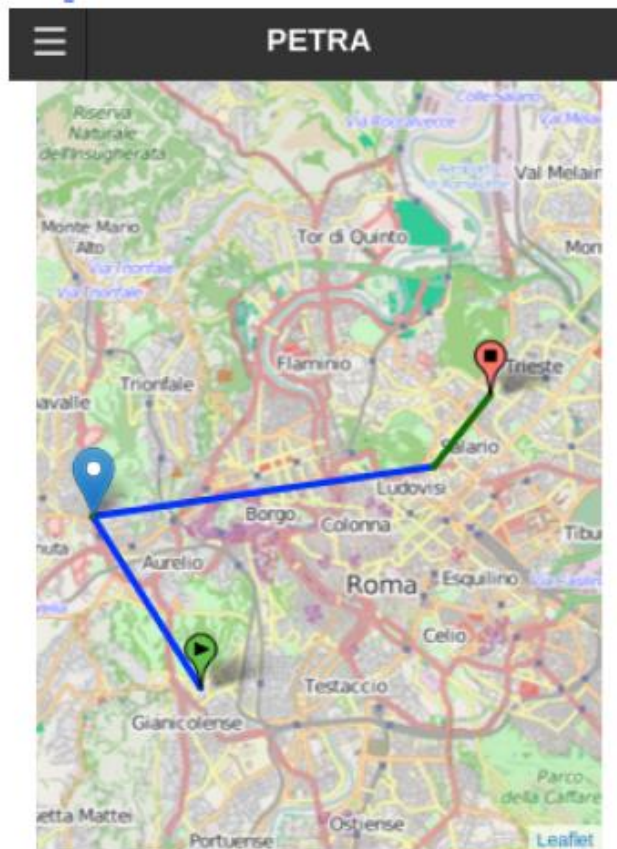
PETRA Carpooling

- **Main idea:** using systematic individual routines as “virtually” available bus lines (or public transport lines).
- **Mobility Profiles:** describe an abstraction in space and time of the systematic movements of a user.



Example of reactive journey planner

The original journey plan

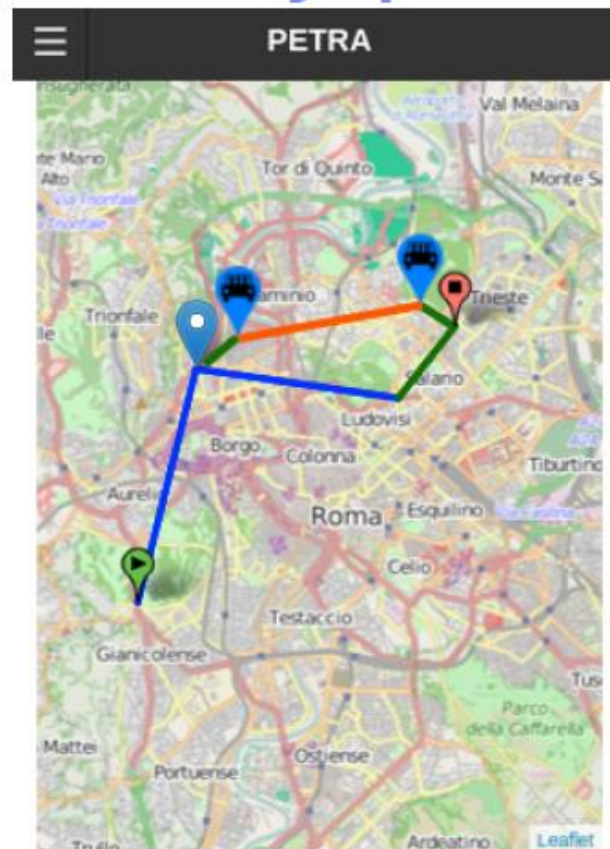


Choose one of (1):

Walk for 20.9552m to stop
CIRCONVALLAZIONE
CORNELIA (77806)

Verify

The adjusted journey plan



Choose one of (2):

Take 495 to stop
CORSO D'ITALIA- PO
(71405)

Verify

Cloudy Cities

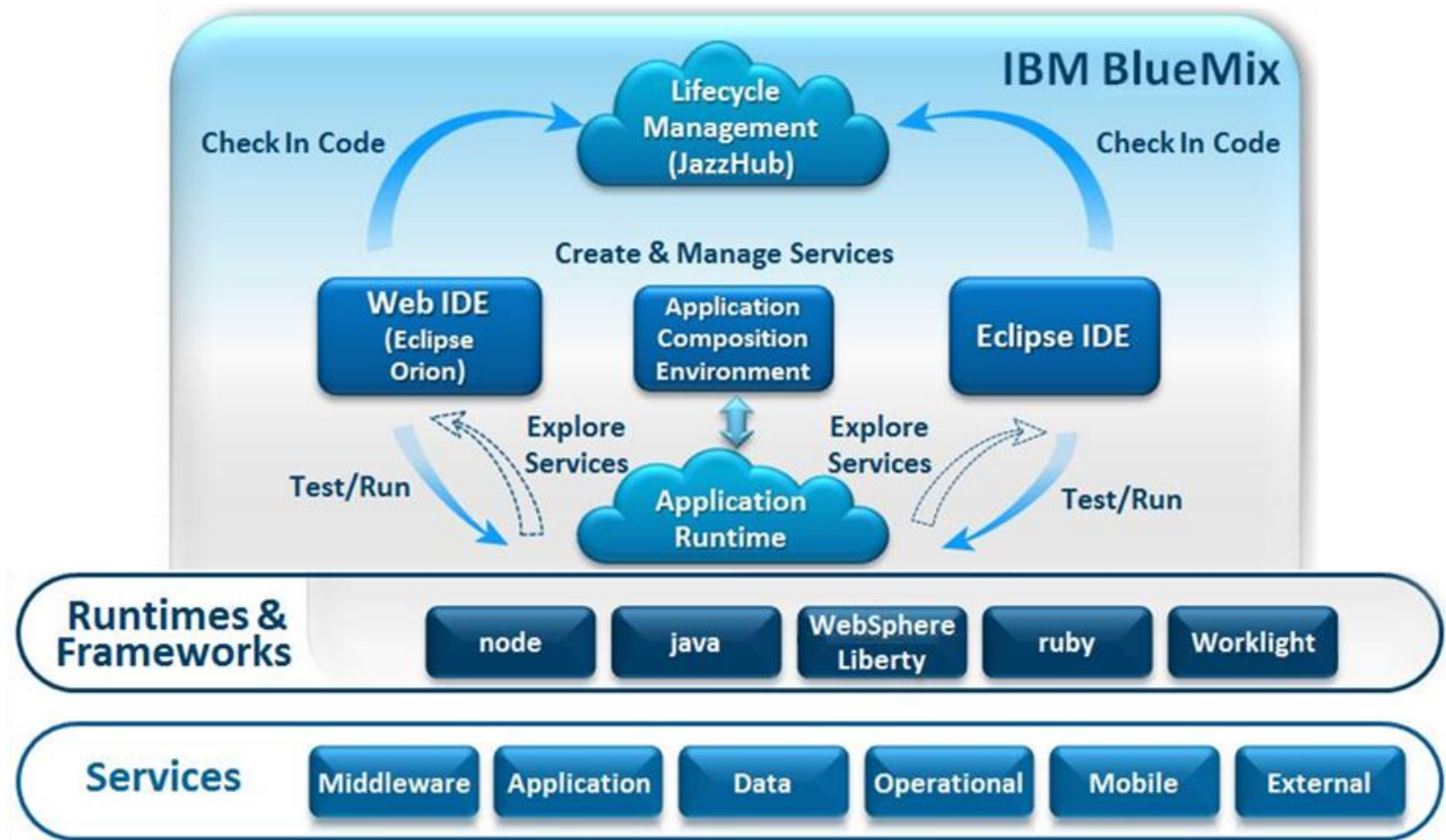
*“Without continual growth and progress,
such words as improvement, achievement,
and success have no meaning”*

Benjamin Franklin



BlueMix Overview

BlueMix is IBM's new PaaS solution that combines the power of Cloud Foundry with popular languages and IBM IaaS.



BlueMix Overview

BlueMix

BlueMix:

- Enables web and mobile applications to be rapidly and incrementally composed of services
- Offers scalability through quick provisioning through its **SoftLayer** cloud layer
- Supports fit-for-purpose programming models and services
- Delivers application changes continuously
- Embeds manageability of services and applications
- Provides optimized and elastic workloads
- Enables continuous availability



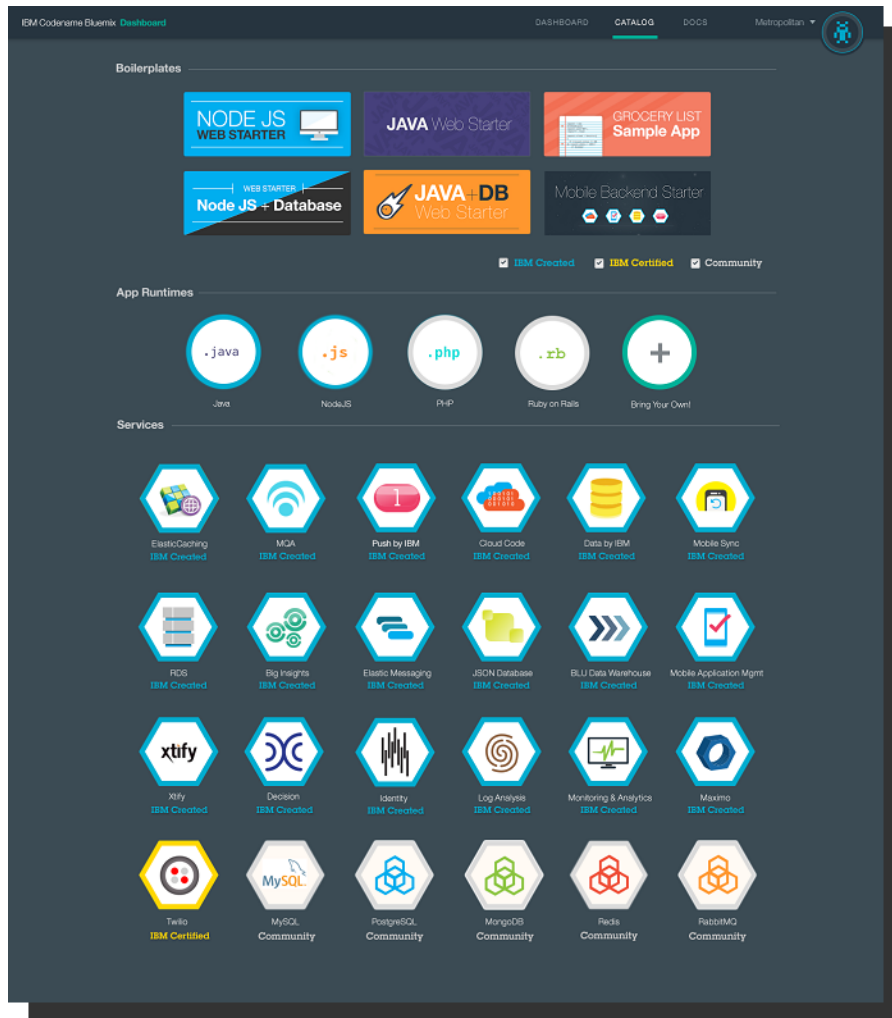
Example Scenarios

Scenarios	Examples
Systems of engagement applications (Mobile, social, Web) that integrate with systems of record applications within enterprises	▪ Mobile self-care application providing end-to-end service across channels to customers ▪ Facebook application that collects and transfers leads into enterprise lead management systems
Internet of Things Machine-to-Machine (M2M) scenarios involving events collection and front end interactive user interfaces	▪ Vehicle APIs to which consumer and fleet vehicles post real-time diagnostics data for driving analysis and insights ▪ Track and Trace applications
Rapid proof of concept development of systems of engagement applications in agile model	▪ Insurance quotes application used by field agents as a point of contact (POC) for collecting opinions and feedback
Cloud API development, deployment, and management	▪ Product catalog APIs exposed by a retail company for partner integration and third-party store fronts ▪ Enterprise APIs exposed for external consumption (consumers, partners, and more)
Self-contained systems of engagement applications composed from external or third-party APIs	▪ Digital taxi solution that integrates Google APIs and shows location co-ordinates of taxis

***Based on current functionality and features of BlueMix**



BlueMix User Interface



Run time

The developer can choose any language runtime or bring their own. Just upload your code and go.

DevOps

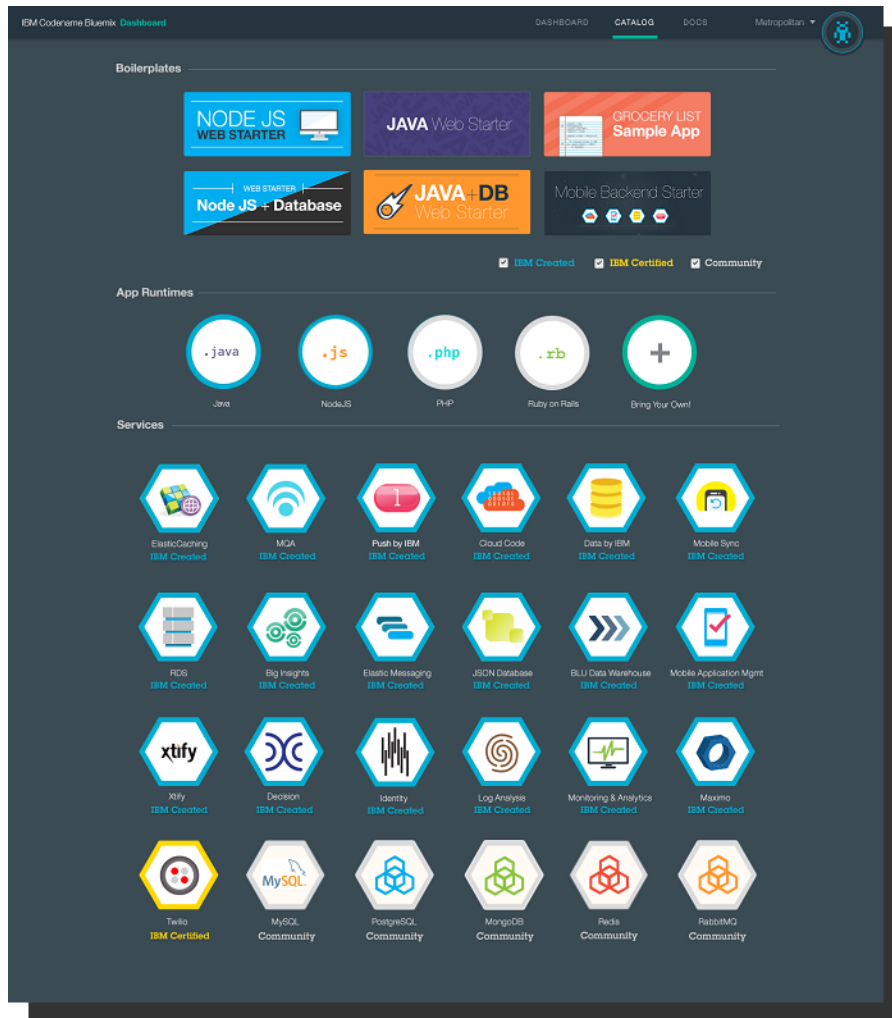
Development, monitoring, deployment, and logging tools allow the developer to run the entire application.

APIs and Services

A catalog of open source, IBM, and third-party APIs services allow a developer to stitch together an application in minutes.



BlueMix User Interface



Cloud Integration

Build hybrid environments. Connect to on-premises systems of record plus other public and private clouds. Expose your own APIs to your developers.

Extend SaaS Apps

Drop in SaaS App SDKs and extend to new use cases (for example, Mobile, Analytics, and web).



Wrap Up

Cities and City Data

- Majority of World population live in cities
- Cities are dynamic entities combining people, systems, infrastructure, businesses
- More and more city data becomes available enabling more insight
- City data is heterogeneous, multi-domain, noisy and big

Challenges

- Finding relevant information over large amounts of city data
- Addressing the 4Vs of Big City Data
- Addressing the end-user needs
- Addressing particular business use-cases

Applications

- Streaming Data
- Social Cities
 - Digital Age & Citizen Engagements
 - How to harness the social media data?
- Transportation
 - Journey Planning
 - Carpooling
- and much more....



References

- Marty Himmelstein, **Local search: The internet is the yellow pages**, IEEE Computer, 2005
- Klaus Berberich, Arnd C. Konig, Dimitrios Lymberopoulos, Peixiang Zhao, **Improving local search ranking through external logs**, SIGIR 2011.
- Hannu Kukka, Vassilis Kostakos, Timo Ojala, Johanna Ylipulli, Tiina Suopajarvi, Marko Jurmu, Simo Hosio, **This is not classified: everyday information seeking and encountering in smart urban spaces**, Personal and Ubiquitous Computing, 2013
- Spink, A., Wolfram, D., Jansen, M. B., & Saracevic, T. (2001). **Searching the web: The public and their queries**. *Journal of the American society for information science and technology*, 52(3), 226-234.
- Zhang, Wei Vivian, Benjamin Rey, Eugene Stipp, and Rosie Jones. **Geomodification in Query Rewriting**. In GIR. 2006.



References

Querio City / Urban Data

- **V. Lopez, S. Kotoulas, M. L. Sbodio, M. Stephenson, A. Gkoulalas-Divanis, P. Mac Aonghusa.** QuerioCity: A Linked Data Platform for Urban Information Management. In Use track at ISWC 2012.
- **V.Lopez, S.Kotoulas, M.L.Sbodio, R.Lloyd.** Guided exploration and integration of urban data. Hypertext'13.

Reasonable City

- **Freddy Lecue, Jeff Z, Pan.** Predicting Knowledge in an Ontology Stream. In Proc. of 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 Proc. IUI 2013
- **Freddy Lecue, Anika Schumann, Marco Luca Sbodio.** Applying Semantic Web Technologies for Diagnosing Road Traffic Congestions. In Proc. of ISWC 2012.

Social City

- **Elizabeth M. Daly, Giusy Di Lorenzo, Daniele Quercia, Michael Muller.** When the City Meets the Citizen. In Proc. of ICWSM 2012.
- **Giusy Di Lorenzo, Marco Luca Sbodio, Vanessa Lopez, Raymond Lloyd.** EXSED: an intelligence tool for Exploration of Social Event Dynamics. In Proc. of MDM 2013.

Stream City

- **Simone Tallevi, Spyros Kotoulas, Luca Foshini, Freddy Lecue, Antonio Corradi.** Real-time Urban Monitoring in Dublin using Semantic and Stream Technologies. In Use track at ISWC 2013

Care City

- **Spyros Kotoulas, Vanesa Lopez, Martin Stephenson et al.** Coordinating social care and health care using Semantic Web technologies. Demo session at ISWC 2013 (submitted)

SPUD: Semantic Processing of Urban Data – Demo: www.dublinked.ie/sandbox/SemanticWebChall
Kotoulas, Vanessa Lopez, Raymond Lloyd, Marco Luca Sbodio, Freddy Lecue, Martin Stephenson, Elizabeth Daly, Veli Bicer, Aris Gkoulalas-Divanis, Giusy Di Lorenzo, Anika Schumann, Denis Patterson, and Pol Mac Aonghusa



References

Processing and publishing Linked urban Data

- **[Maali, ESWC'12]** Maali, F., Cyganiak, R., Peristeras, V.: A publishing pipeline for linked government data. Proc. of ESWC, 2012.
- **[Datalift]** Schar_e, F., Atemezing, G., R., T., Gandon, F.e.a.: Enabling linked-data publication with the datalift platform. In (AAAI'12) Workshop on Semantic Cities, 2012
- **[TWC LOGD]** Ding. ,L., Lebo., T., Erickson, J.S. et al.: Twc logd: A portal for linked open government data ecosystems. Web Semantics, 2011.
- **IBM City Forward:** <http://cityforward.org>

Semantic Lifting

- **[RF123]** Han, L., Finin, T., Parr, C., Sachs, J., Joshi, A.: RDF123: From Spreadsheets to RDF. Proc. of ISWC 2008
- **[Csv2rdf4lod]** <http://data-gov.tw.rpi.edu/wiki/Csv2rdf4lod>
- **Skjæveland, M.G., Lian, E. H., Horrocks, I.** Publishing the Norwegian Petroleum Directorate's FactPages as Semantic Web Data. In use ISWC'13

Web Tables

- **Cafarella, M.J., Halevy, A., Madhavan, J.:** Structured data on the Web. Communications of the ACM, 2011
- **Sarma A., Fang, L., Gupta, N., Halevy, A., et al.:** Finding Related Tables, SIGMOD '12

Urban Dynamics

- **Kling f., Pozdnoukhov, A.:** When a city tells a story. In ACM SIGSPATIAL GIS, 2012

Evaluation Campaigns

- **Blanco et al.** Repeatable and Reliable Search System Evaluation using Crowd-Sourcing, SIGIR 2011
- **[QALD, JSW'13]** Lopez, V., Unger, C., Cimiano P., Motta, E.: Evaluating Question Answering over Linked Data, Journal Web Semantics 2013, <http://greententacle.techfak.uni-bielefeld.de/~cunger/qald/>



References

- Retrieval
 - Changsung Kang, Xuanhui Wang, Yi Chang, Belle Tseng, **Learning to rank with multi-aspect relevance for vertical search**, WSDM 2012
 - Nicholas D Lane, Dimitrios Lymberopoulos, Feng Zhao, Andrew T. Campbell, **Hapori: context-based local search for mobile phones using community behavioral modeling and similarity**, Ubicomp, 2010.
 - Klaus Berberich, Arnd C. König, Dimitrios Lymberopoulos, Peixiang Zhao, **Improving local search ranking through external logs**, SIGIR 2011.
 - Cheng, Zhiyuan, et al. **Toward traffic-driven location-based Web search**. CIKM, 2011.
 - Hristidis, Vagelis, Heasoo Hwang, and Yannis Papakonstantinou. **Authority-based keyword search in databases**. ACM Transactions on Database Systems (TODS) 33, no. 1 (2008): 1.
 - Li, Guoliang, Beng Chin Ooi, Jianhua Feng, Jianyong Wang, and Lizhu Zhou. **EASE: an effective 3-in-1 keyword search method for unstructured, semi-structured and structured data**. In *SIGMOD*, 2008.
 - Guo, Lin, Feng Shao, Chavdar Botev, and Jayavel Shanmugasundaram. **XRANK: ranked keyword search over XML documents**. In *SIGMOD*, 2003.
 - Bicer, Veli, Thanh Tran, and Radoslav Nedkov. **Ranking support for keyword search on structured data using relevance models**. In CIKM, 2011.

