QuerioCity: A Linked Data Platform for Urban Information Management

*Vanessa Lopez, **Spyros Kotoulas**, Marco Luca Sbodio, Martin Stephenson, Aris Gkoulalas-Divanis, Pol Mac Aonghusa*

This presentation also contains work from:

SPUD: Semantic Processing of Urban Data

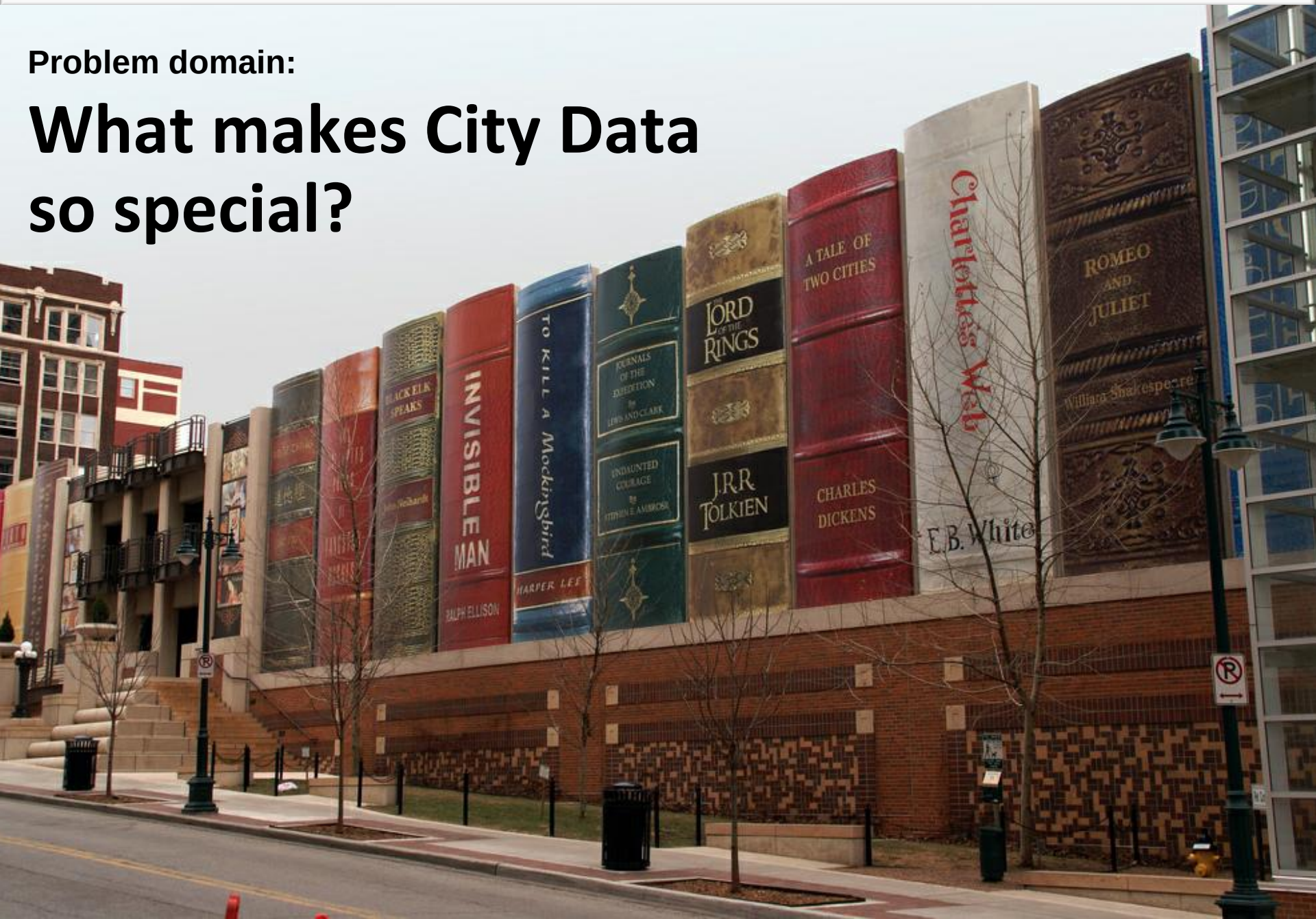
***Spyros Kotoulas**, Vanessa Lopez, Raymond Lloyd, Marco Luca Sbodio, Freddy Lecue, Martin Stephenson, Elizabeth Daly, Veli Bicer, Aris Gkoulalas-Divanis, Giusy Di Lorenzo, Anika Schumann, Denis Patterson, and Pol Mac Aonghusa*

Main question:

How can we go from Raw Data to Insight into the operation of a City, with minimal effort?

Problem domain:

What makes City Data so special?



Big City Data (4 Vs)

+ Openness + Number of datasets!

Volume

- Lots of relevant information

Velocity

- Streams
- Frequent updates

Variety

- Different models
- Different file formats

Veracity

- Diverse sources
- Difficult to do assess quality

Process view

Publishing

- Upload
- Annotation

Organization

- Indexing
- Cataloguing
- Dataset discovery

Extraction

- Querying
- Semantic Search

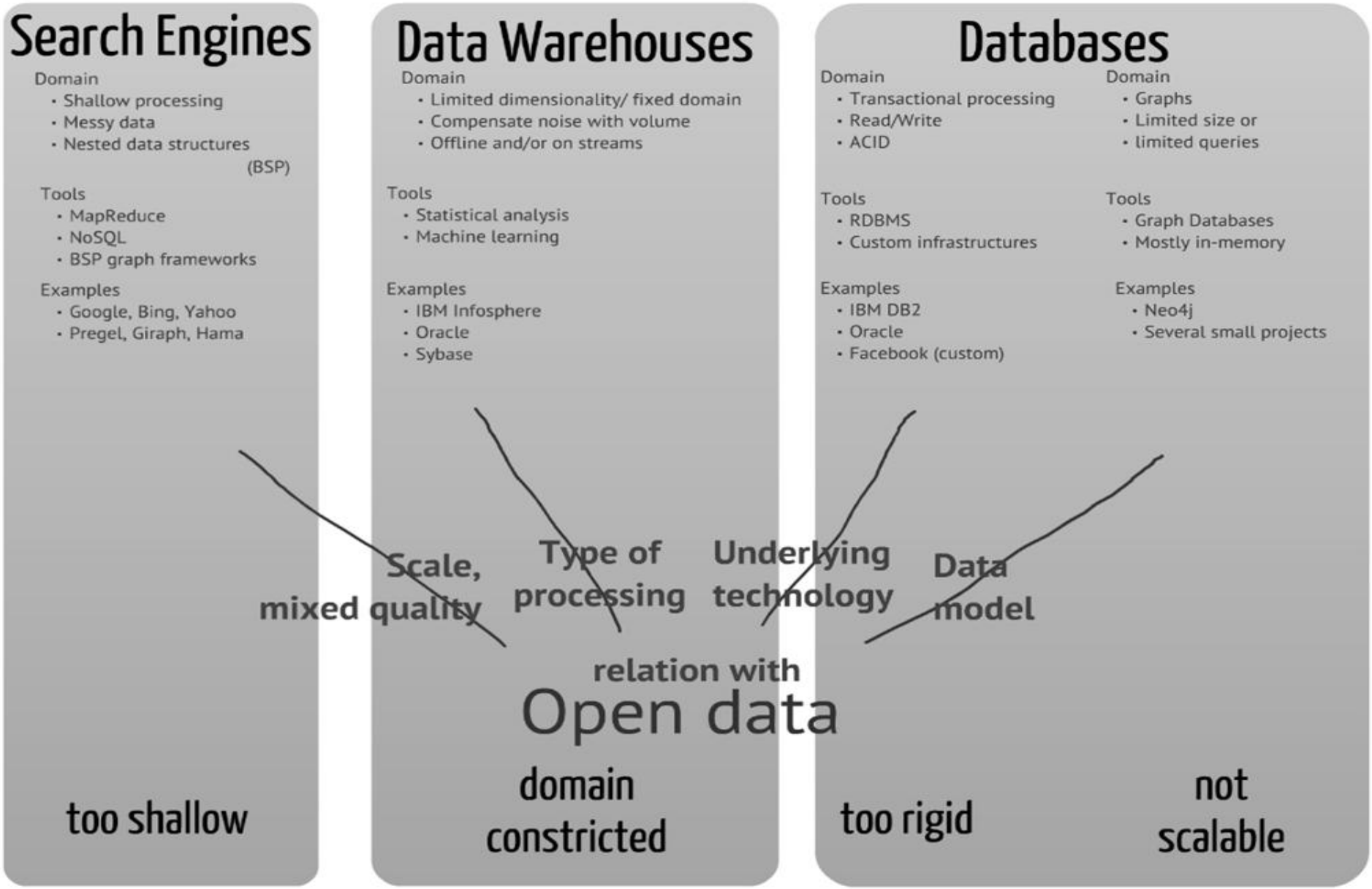
Usage

- Tool integration
- Visualization
- Analysis

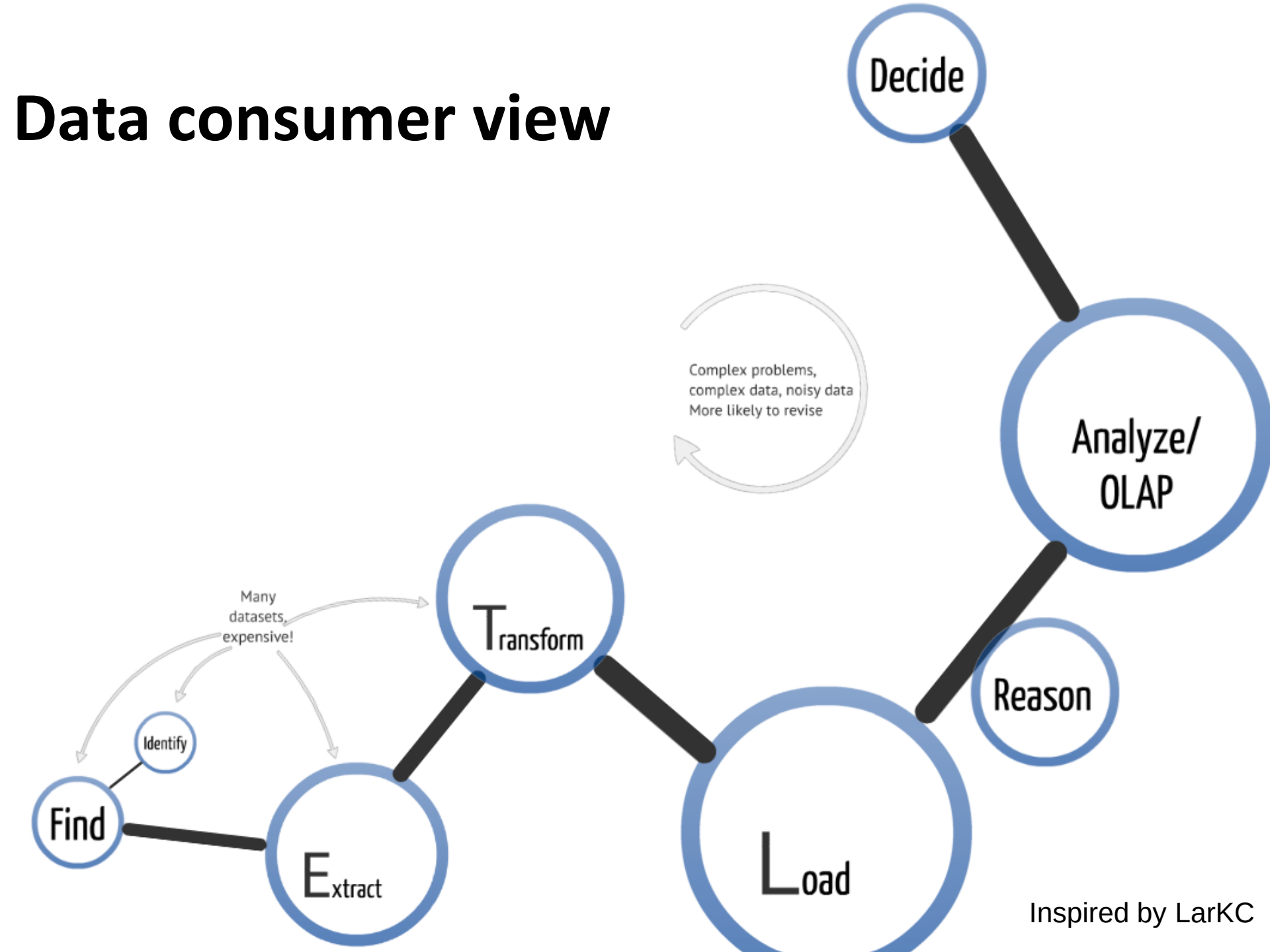
Ubiquitous aspects: Provenance, Governance, Performance, Security, Privacy

Tools: Reasoning, Information Retrieval, Machine Learning, Data Mining

Systems view



Data consumer view



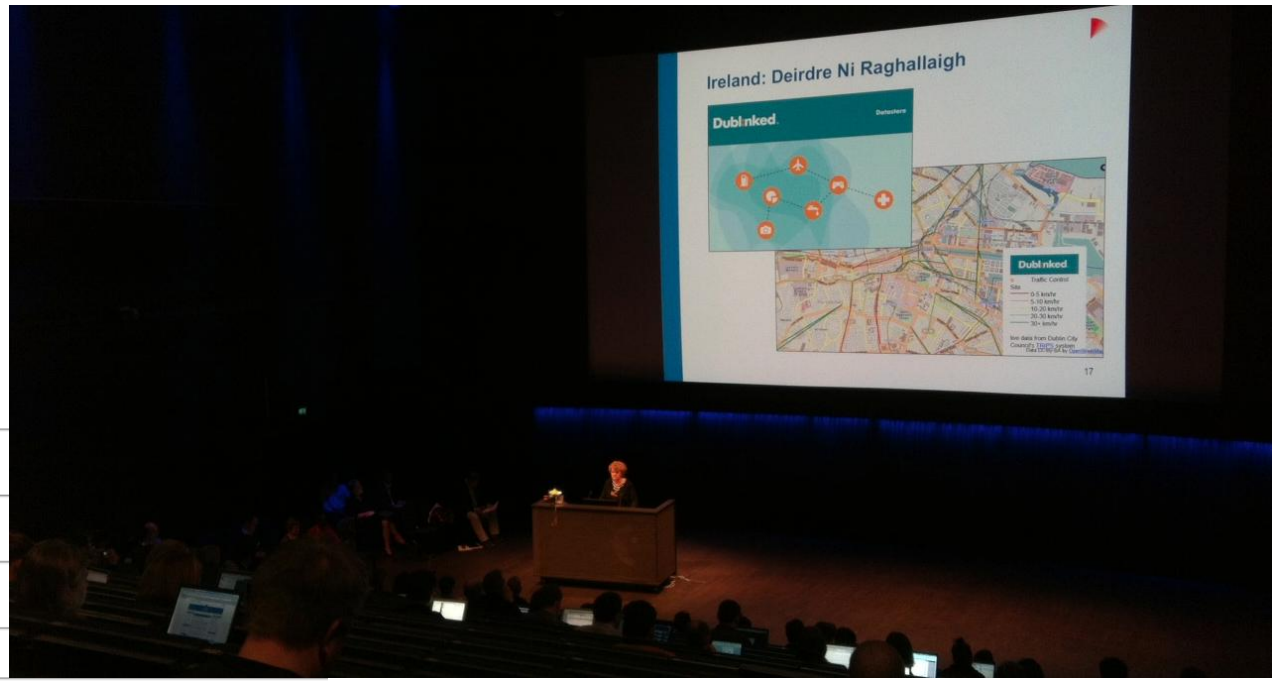
Inspired by LarKC

Real-World City Data: Dublinked.ie

Dublinked™ Datastore Apps Activities Meml

46 Countries
Ireland, US, UK, Germany, Poland, France, Spain, Italy, NL

249 Cities
Irish, Bialystok, Kensington



Traffic Volumes- City Centre Bridges Count Data 2007-2010	
Traffic Volumes- Canal Cordon Count Data 2008-2010	
Fats& Oils & Greases (FOG) Licences Register	
On Street Disabled Parking Bay in Dublin City Council area	
2011-2017 Dublin City Development Plan - Parking Zones	
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	

Corporation

SmarterCities Technology Centre

An example from Dublinked

No common schema

No explicit semantics

No common reference

No common vocabulary

5	400[illegible]	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	4012	BLANCHARDSTOWN-DELWOOD	1689	1405
14	401[illegible]	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
20	4019	DONABATE	2492	1923
21	4020	DUBBER	1369	210
22	4021	GARRISTOWN	371	359
23	4022	HOLLYWOOD	305	286
24	4023	HOLLYSTOWN	278	208

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)

Structure is not declared, this is no more than a matrix

100's times

SmarterCities Technology Centre

But we have the solution:
Linked Data
Right?

Linked Data Technologies

Analytics

SmarterCities Technology Centre

Linking Data is Expensive

(because data integration is expensive)

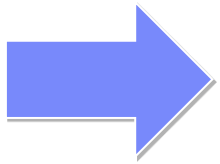
Linked Data Technologies

Analytics

Linked Data is Massive and Complicated

(because the domain is practically everything)

Linked Data Technologies



Analytics

SmarterCities Technology Centre

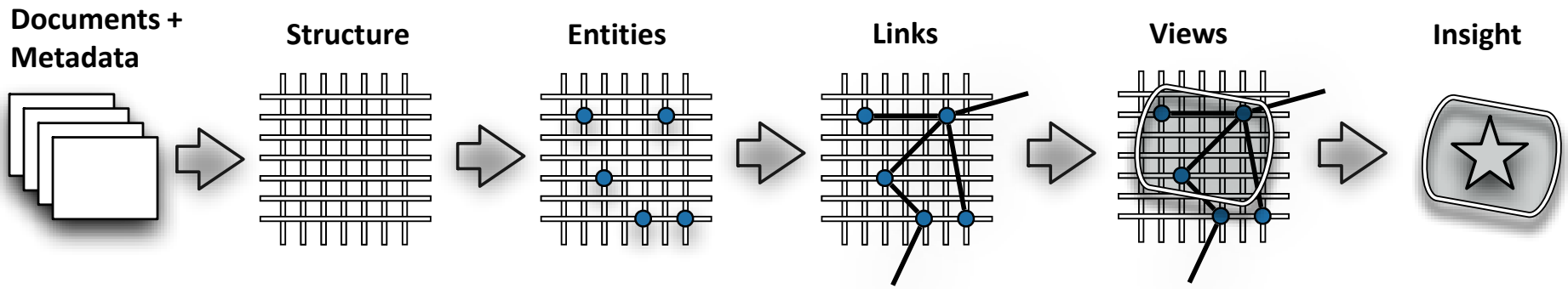
Return-on-investment is crucial

Linked Data Technologies

Analytics

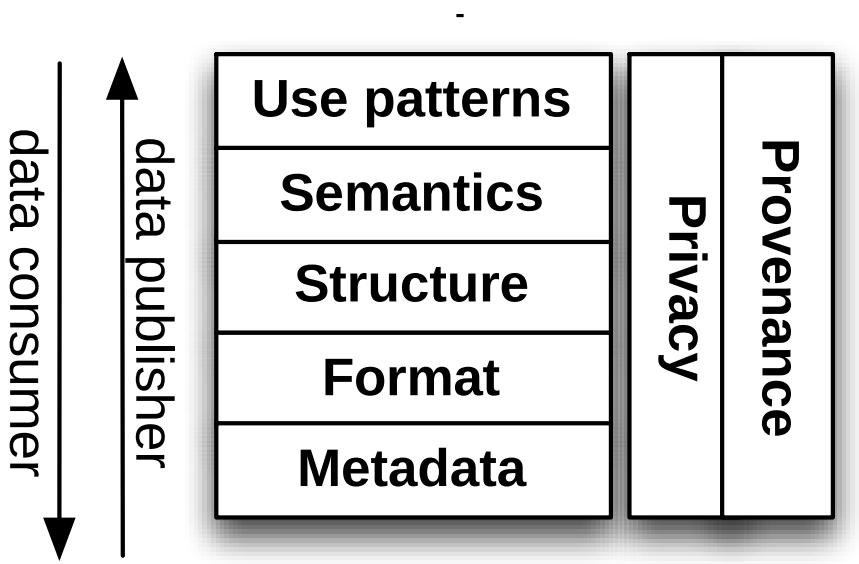
SmarterCities Technology Centre

General approach

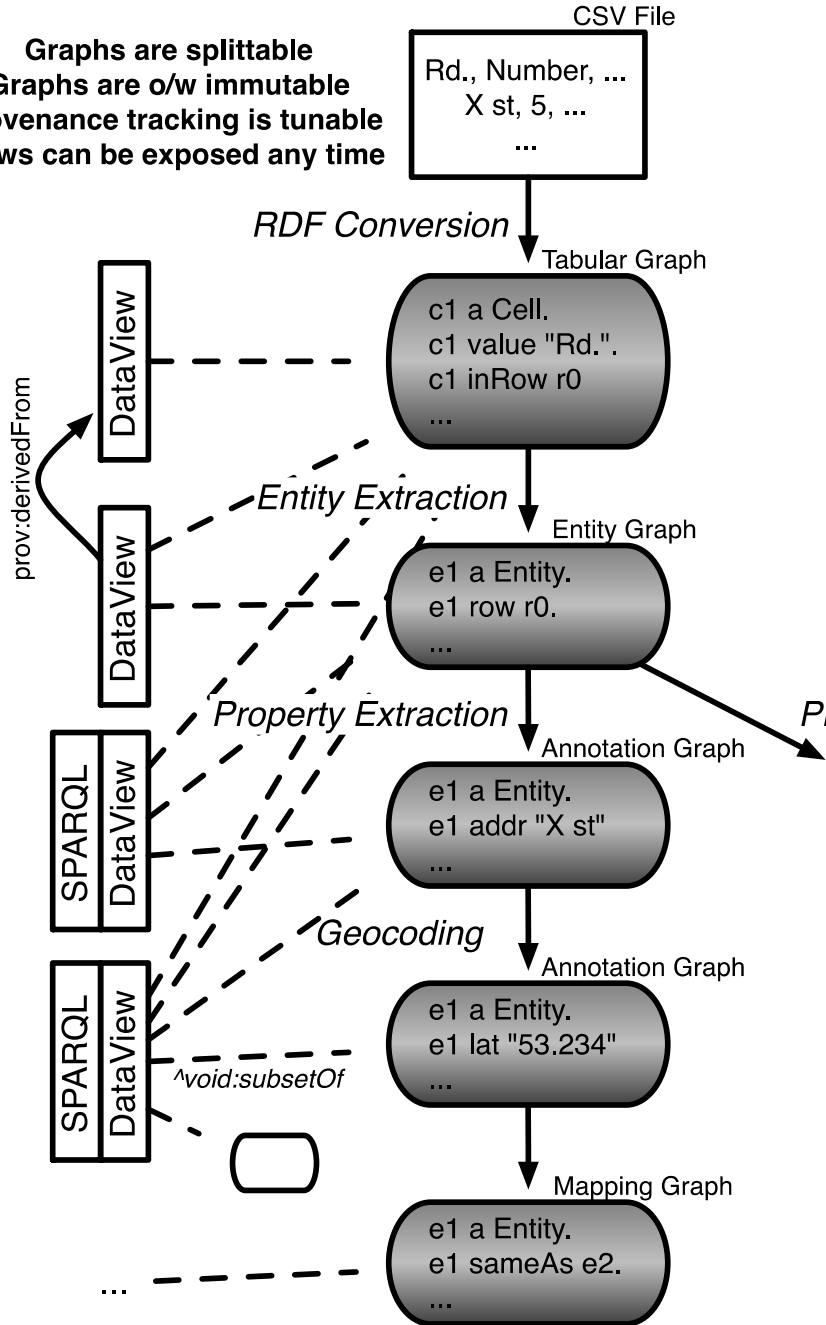


Pay-as-you-go, Gain-as-you-go

Information Management Outline:



Graphs are splittable
Graphs are o/w immutable
Provenance tracking is tunable
Views can be exposed any time



Semantic Processing of Urban Data

Business case in Dublin

- Why are ambulances late?

Sources of information

- 100's of datasets from four municipal authorities in Dublin
- Most static, some dynamic
- Social Media
- Linked Data

Domain of information

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

Semantic Processing of Urban Data (Process)

Publish and catalog information

- Link Metadata using domain Vocabularies (in our case, IPSV)
- Convert to simple RDF format

Extract entities

- Extract entities represented in RDF
- Link when we have high confidence (e.g. Latitude/Longitude, labels)
- Calculate link recommendations otherwise

Integrate

- Link (internally and externally)
- Extract semantics representation from input such as social Media

Extract interesting views

- Information is often hidden (e.g. location of hospitals is hidden in Fats & Greases licenses)

Semantic Diagnosis based on automated reasoning

- ***See presentation by Freddy Lecue at 14:00 (same room)***

Shameless Advertising

Poster & Demo session

Semantic Web Challenge submission

IEEE Internet Computing

Special Issue on Smart Cities

Currently **hiring @ IBM Research**

DEMO