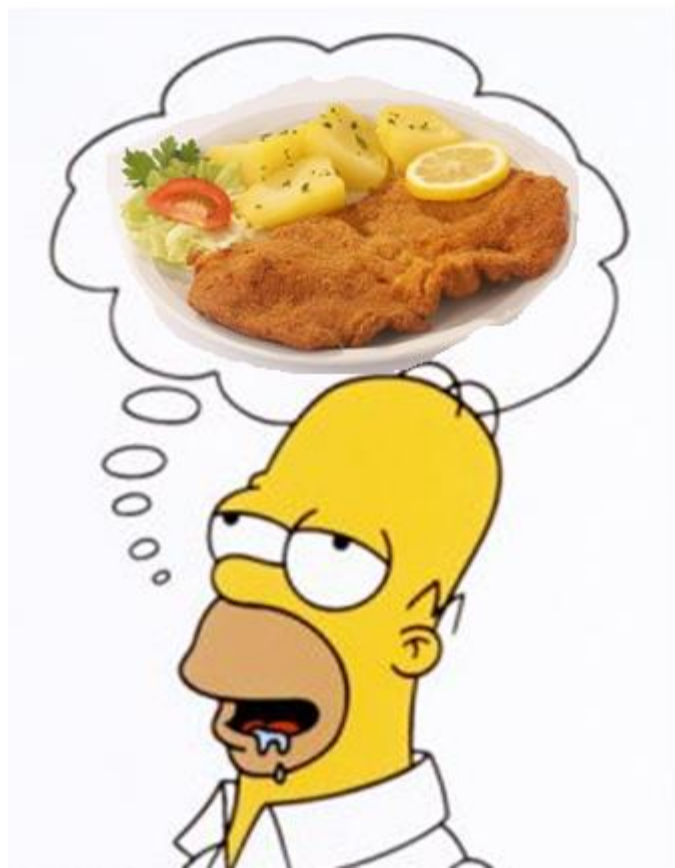# Linked Data Visualization

Laura Po, PhD

Department of Engineering "Enzo Ferrari"

University of Modena and Reggio Emilia

Italy

3th Keystone Training School - *Keyword Search in Big Linked Data*

Institute for Software Technology and Interactive Systems, TU Wien, Austria

Goal of the Talk

- To provide practical skills required for exploring LOD sources

Outline

- The importance of visualization
- How a Linked Data Visualization Process can be defined
- Practical use of LOD/ RDF browsers and visualization toolkits

Why is visualization of Linked Data important?

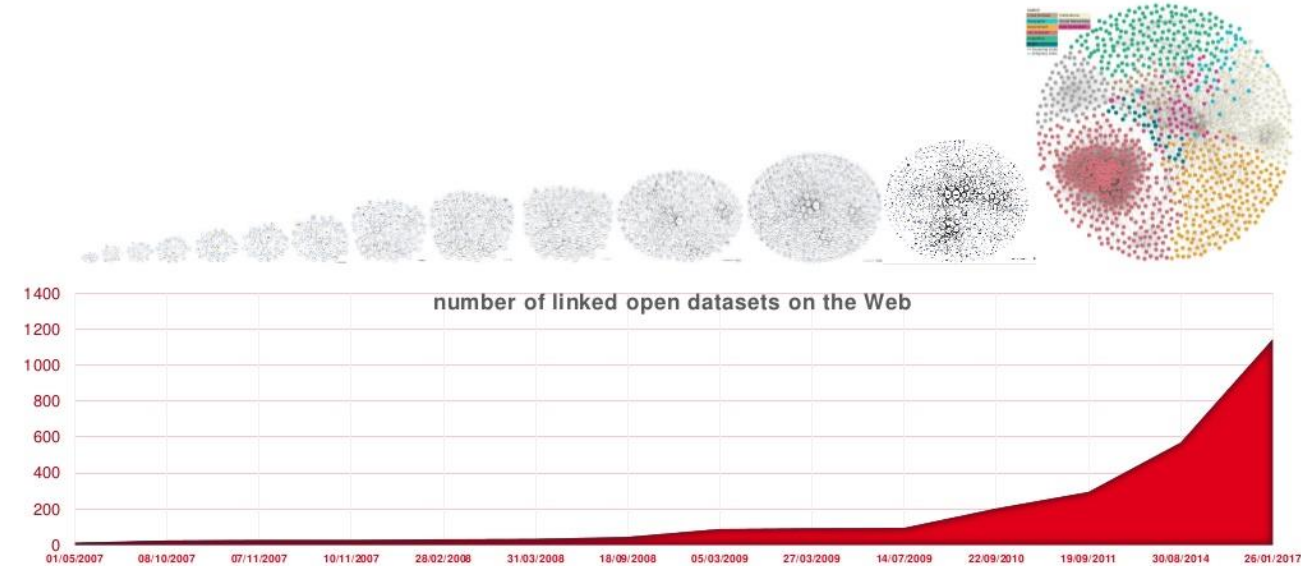- Large and Dynamic Data
  - Efficiently and effectively handle billions of objects within dynamic datsets

- Visual presentation and interaction issues
  - Offer ways to easly explore datasets
  - Proposing summaries and overviews
  - Incremental and progressive techniques

- Variety of Users and Tasks

BOLD – Big Open Linked Data
"The bigger the number, the harder it can be to visualise"

Bratsas et al (2016), Preface on special session "data impact: Big, open, linked data innovations" at 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP) https://doi.org/10.1109/SMAP.2016.7753368

Dwivedi et al, (2017) *Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural*, Inf Syst Front (2017) 19:197–212 https://link.springer.com/article/10.1007/s10796-016-9675-5

Saxena, (2017) *BOLD (Big and Open Linked Data): what's next?*, Library Hi Tech News, Vol. 34 Issue: 5, pp.10-13, https://doi.org/10.1108/LHTN-04-2017-0020

Craig, (2016), *BOLD: The power and potential of Big Open Linked Data*, Published on 11 Oct 2016 on the Thomson Reuters Blog https://blogs.thomsonreuters.com/answerson/bold-power-potential-big-open-linked-data/

Why visualize data instead of provide statistic analysis?

- Anscombe's quartet of datasets having similar statistical properties but appearing very different when plotted

http://en.wikipedia.org/wiki/Anscombe's_quartet
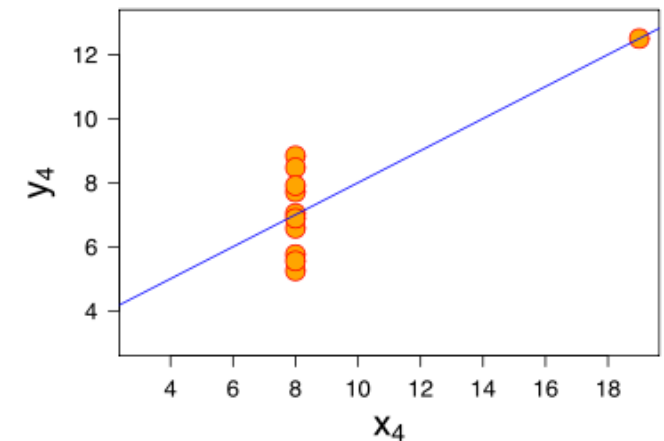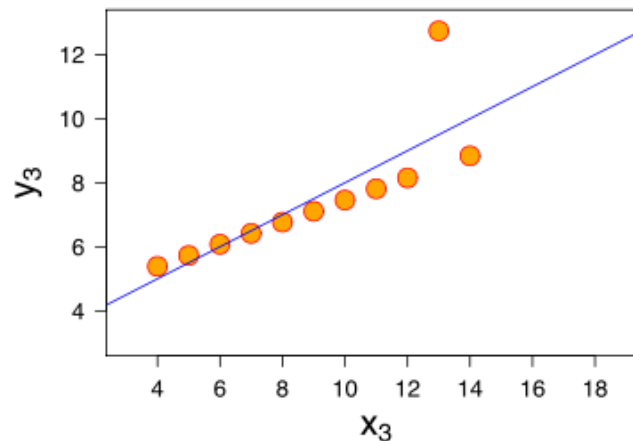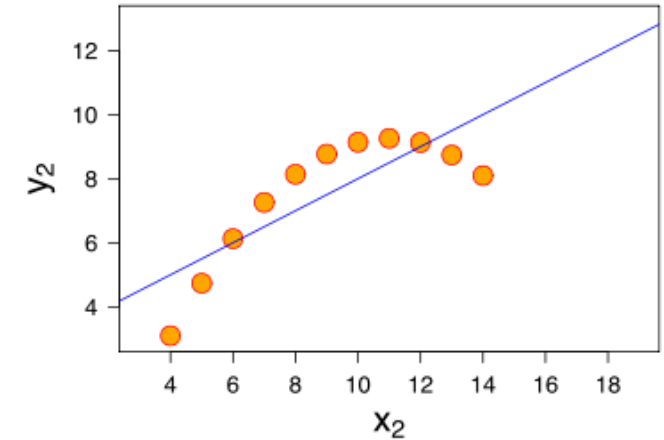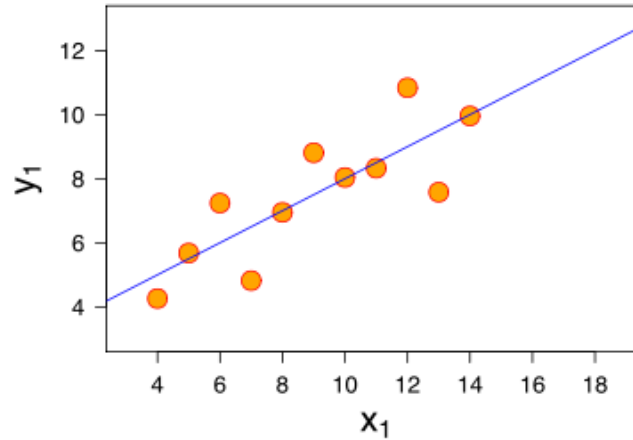
Users

PRODUCERS

Consumers

domain expert

Lay-users

Technical expert

LOD Visualization

- LOD simplifies accessing and integrating data from different sources
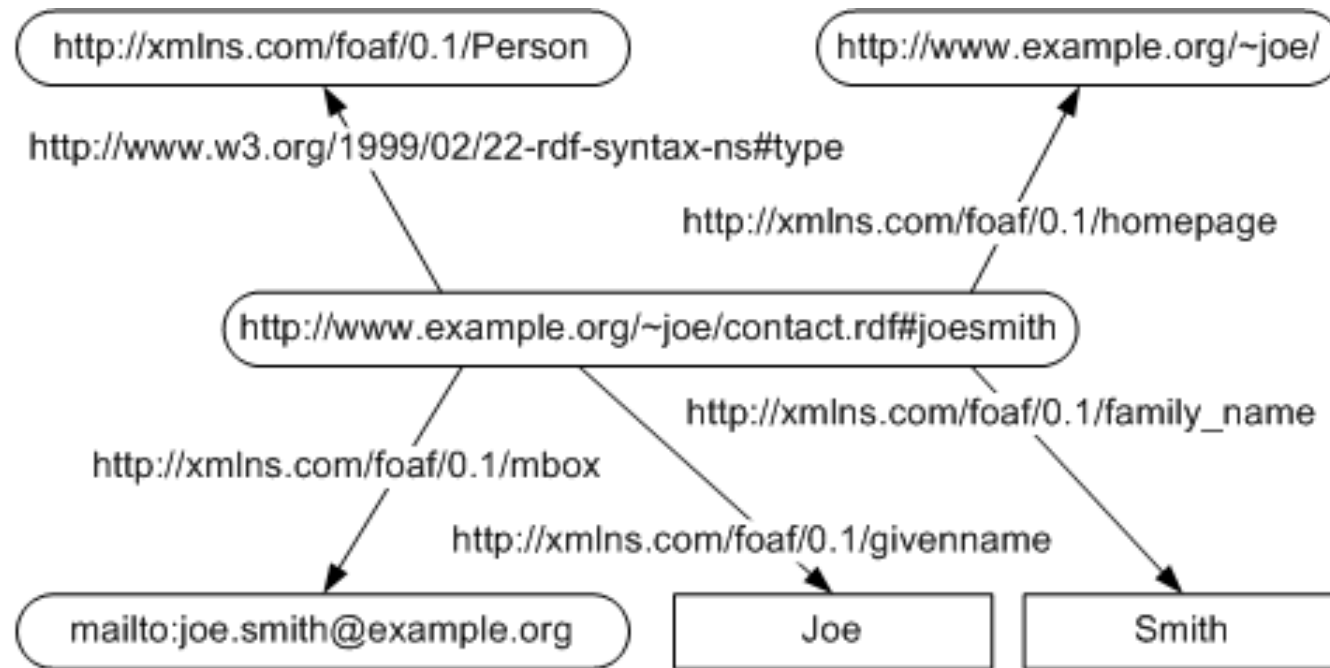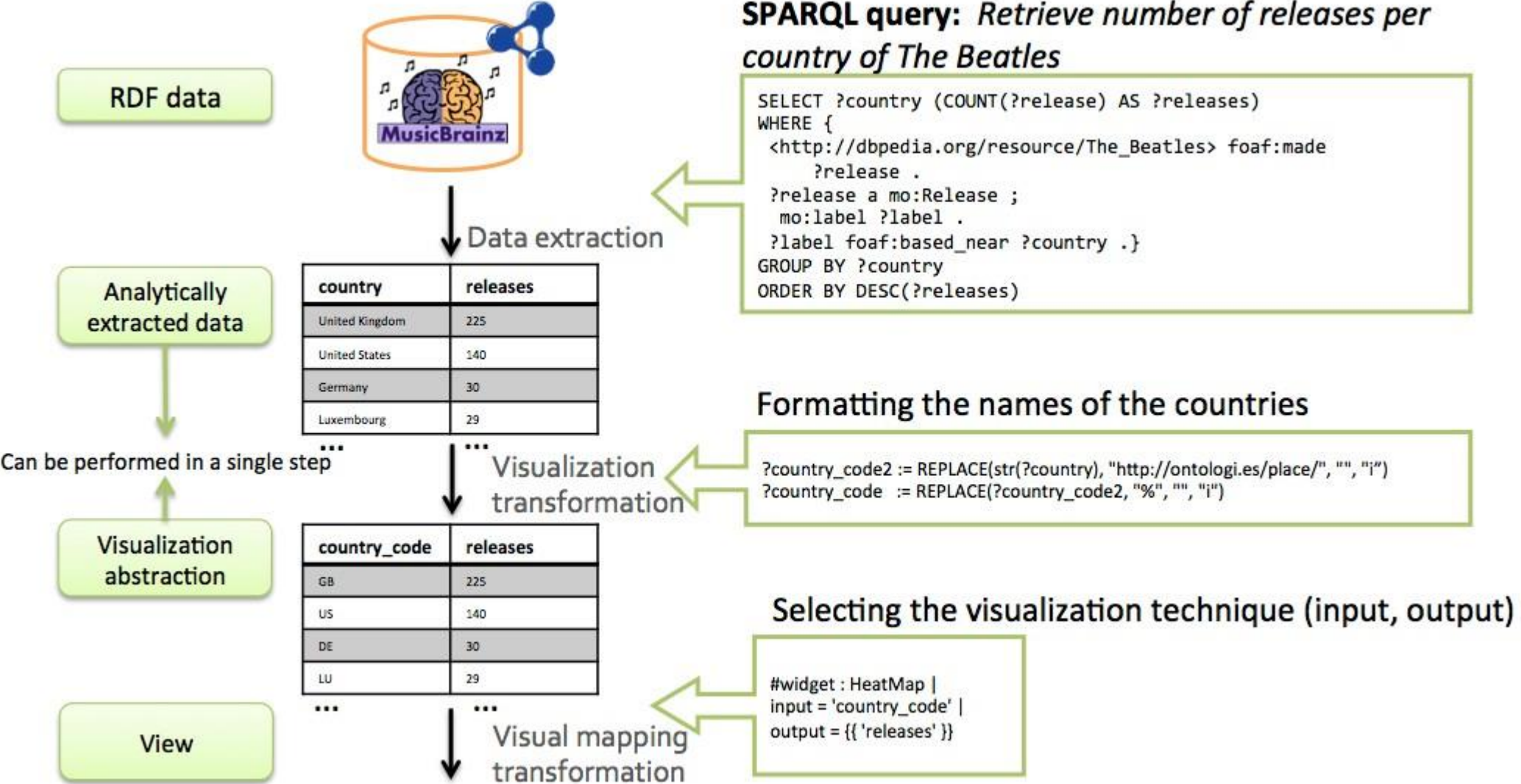- SPARQL makes it easy to select from, and analyse the data

- It's natural to visualise the data as graphs (networks)
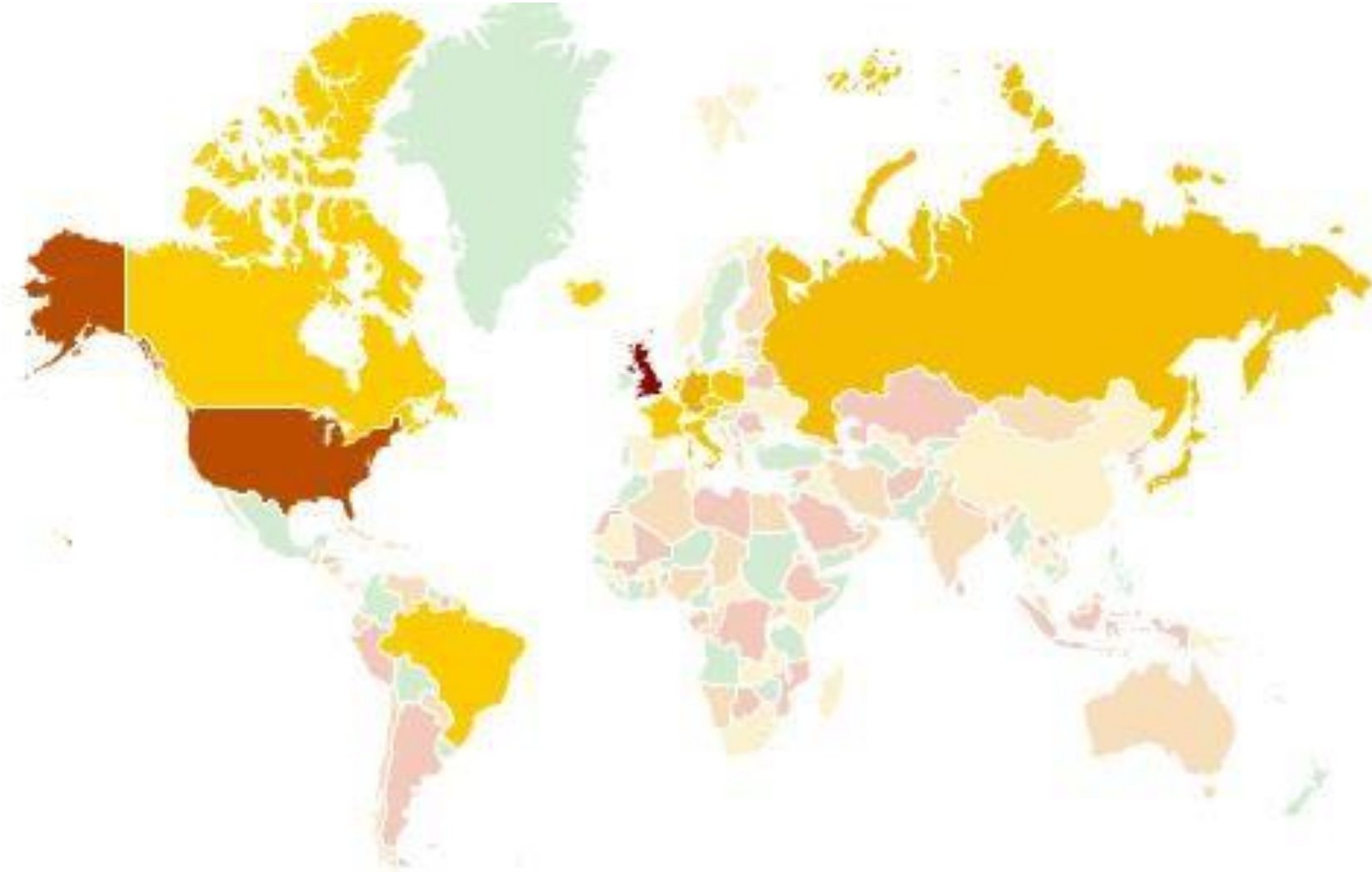 … but other forms of visualisation also possible

RDF Graph

Example of LOD visualization process

SPARQL query: *Retrieve number of releases per country of The Beatles*

```
SELECT ?country (COUNT(?release) AS ?releases)
WHERE {
 <http://dbpedia.org/resource/The_Beatles> foaf:made
     ?release .
 ?release a mo:Release ;
   mo:label ?label .
 ?label foaf:based_near ?country .}
GROUP BY ?country
ORDER BY DESC(?releases)
```

RDF data

Analytically extracted data

country	releases
United Kingdom	225
United States	140
Germany	30
Luxembourg	29
...	...

Data extraction

Formatting the names of the countries

```
?country_code2 := REPLACE(str(?country), "http://ontologi.es/place/", "", "i")
?country_code  := REPLACE(?country_code2, "%", "", "i")
```

Visualization transformation

Can be performed in a single step

Visualization abstraction

country_code	releases
GB	225
US	140
DE	30
LU	29
...	...

Selecting the visualization technique (input, output)

```
#widget : HeatMap |
input = 'country_code' |
output = {{ 'releases' }}
```

View

Visual mapping transformation

Heatmap visualization of The Beatles releases

LOD visualization systems

They can be classified in 6 categories

1. Browsers and Exploratory systems

2. Generic visualization systems

3. Domain vocabulary & device specific systems

4. Graph-based visualization systems

5. Ontology visualization systems

6. Visualization libraries

Bikakis and Sellis, (2016) *Exploration and Visualization in the Web of Big Linked Data: A Survey of the State of the Art.* Proceedings of the Workshops of the EDBT/ICDT 2016 Joint Conference, EDBT/ICDT Workshops 2016, Bordeaux, France, March 15, 2016 http://ceur-ws.org/Vol-1558/paper28.pdf

Evolution over time

System category / year		2001...	...2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Browsers	Textual			x		xx							
	Visual		x	x			xx						
	Faceted			xx	x		x						
	Other paradigms							xx	x	x			
Recommenders	Type/domain specific								xxx				
	Cross-type/domain									xx			
	Undisclosed/commercial										x	xx	
ESS	View-based										x	x	
	Algorithm-based							x	xx		x	x	x
Research focus													

SemWeb in the Scientific American   OWL is a W3C recom.   LOD & DBpedia   SPARQL is a W3C recom.

Marie and Gandon, (2014) *Survey of Linked Data Based Exploration Systems*, Proceedings of the 3rd International Workshop on Intelligent Exploration of Semantic Data (IESD 2014) co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, October 20, 2014
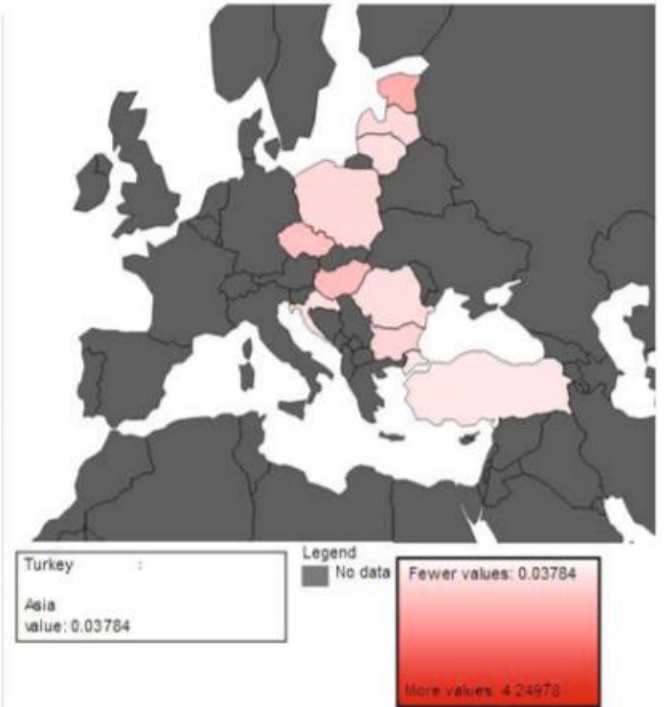http://ceur-ws.org/Vol-1279/iesd14_8.pdf

Exploratory search

- Exploratory search systems (ESS) forms a special category of seeking information on the Web with the purpose of revealing related information to the searcher along with retrievals of what have been searched for.

Palagi, et al. (2017), *A Survey of Definitions and Models of Exploratory Search.* Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics. ACM, http://doi.acm.org/10.1145/3038462.3038465
Marie,(2015) , *Linked data based exploratory search*. PhD Thesis, Université Nice Sophia, Antipolis, https://tel.archives-ouvertes.fr/tel-01130622

Classification and Comparisons

Tool Name	Versions		Source Visualization		Querying						Result Visualization	
	Type	Last Update	Classes	Instances	over multiple sources	by key-words or sentences	visual building	visual representation	by SPARQL language	related queries	List	Advanced
LODlive	live demo [6], desktop	2012	only for the selected classes	only for the selected classes	yes	no	no	no	no	no	no	yes
LODmilla	online [7]	2014	only for the selected classes	only for the selected classes	yes	no	no	no	no	no	no	yes
gFacet	live demo [1], desktop	2011	only for the selected classes	only for the selected classes	no	yes	yes	yes	no	no	yes	no
LOD Visualization	online [4]	2012	yes	no	no	no	yes	yes	no	no	yes	no
iSPARQL	online [2], desktop	2014	no	no	yes	no	no	yes	yes	no	yes	yes
SPARKLIS	online [9]	2014	no	no	no	yes	yes	yes	no	yes	yes	no
PAYOLA	live demo [8], desktop	2013	no	yes	no	no	yes	no	yes	no	yes	yes
LODatio	online [5]	2013	no	no	yes	no	no	no	yes	yes	yes	no
Linked Data Query Wizard	online [3]	2014	no	no	no	yes	yes	yes	no	no	yes	yes
Visualbox	demo [10], desktop	2013	no	no	yes	no	no	no	yes	yes	yes	yes
Affective Graph	—	2014	only for the selected classes	no	no	no	yes	no	yes	no	yes	no
LODeX	online	2014	yes	no	no	no	yes	yes	yes	no	yes	yes

CODE Linked Data Query Wizard

https://code.know-center.tugraz.at/search

http://code-research.eu/
EU project 2012-2014

Hoefler, Patrick, et al. "Linked Data Query Wizard: A Novel Interface for Accessing SPARQL Endpoints." *LDOW*. 2014.

TEST

- Using CODE Linked Data Query Wizard search for "Johann Strauss II" within the Dbpedia source

- Explore the result

- Add columns that show some property like "birth place", "given name", "music composer of ", …

Label ▾	▼ "Johann Strauss II" ▾	Type ▾		Birth date ▾	Birth place ▾	IsPrimaryTopicOf ▾	Music by of ▾	Music composer of ▾	Musician of (old) ▾	GivenName ▾
Johann Strauss, II						Johann_Strauss,_II ▾				
Johann Strauss II		Agent ▾		1825-10-25	Neubau ▾	Johann_Strauss_II ▾	Paradise Found (musical) ▾	A Night in Venice (1934 film) ▾	A Corny Concerto ▾	Johann
		Artist109812338 ▾			Vienna, Austria ▾		The Great Waltz ▾	A Night in Venice (1953 film) ▾	Johann Mouse ▾	
		CausalAgent100007347 ▾						Boudu Saved from Drowning ▾		
		Composer109947232 ▾						Oh... Rosalinda!! ▾		
		Creator109614315 ▾						The Reluctant Sadist ▾		
		Entertainer109616922 ▾						Vienna Blood (film) ▾		
		Human ▾						Voices of Spring (1933 film) ▾		
		Human ▾								
		LivingThing100004258 ▾								
		Musician110339966 ▾								
		Musician110340312 ▾								
		NaturalPerson ▾								
		Object100002684 ▾								
		Organism100004475 ▾								
		Performer110415638 ▾								

TEST
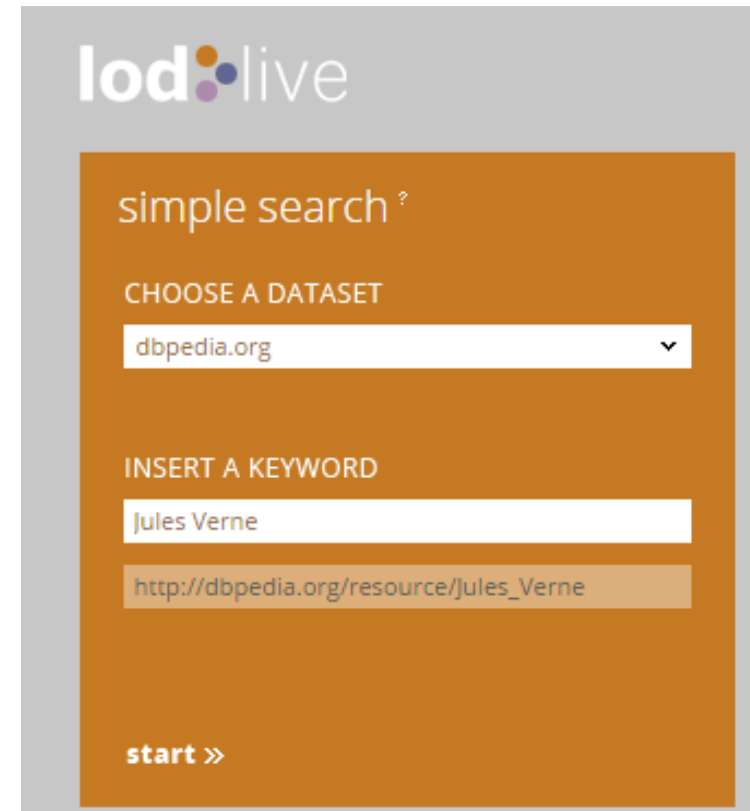
- Navigate the Linked Jazz cloud
- Change the visualization option (fized, similar, gender, dynamic)

LOD live

LodLive project provides a demonstration of the use of Linked Data standards (RDF, SPARQL) to browse RDF resources. The application aims to spread linked data principles using a simple and friendly interface with reusable techniques.

http://en.lodlive.it/
http://en.lodlive.it/?http://dbpedia.org/resource/Jules_Verne
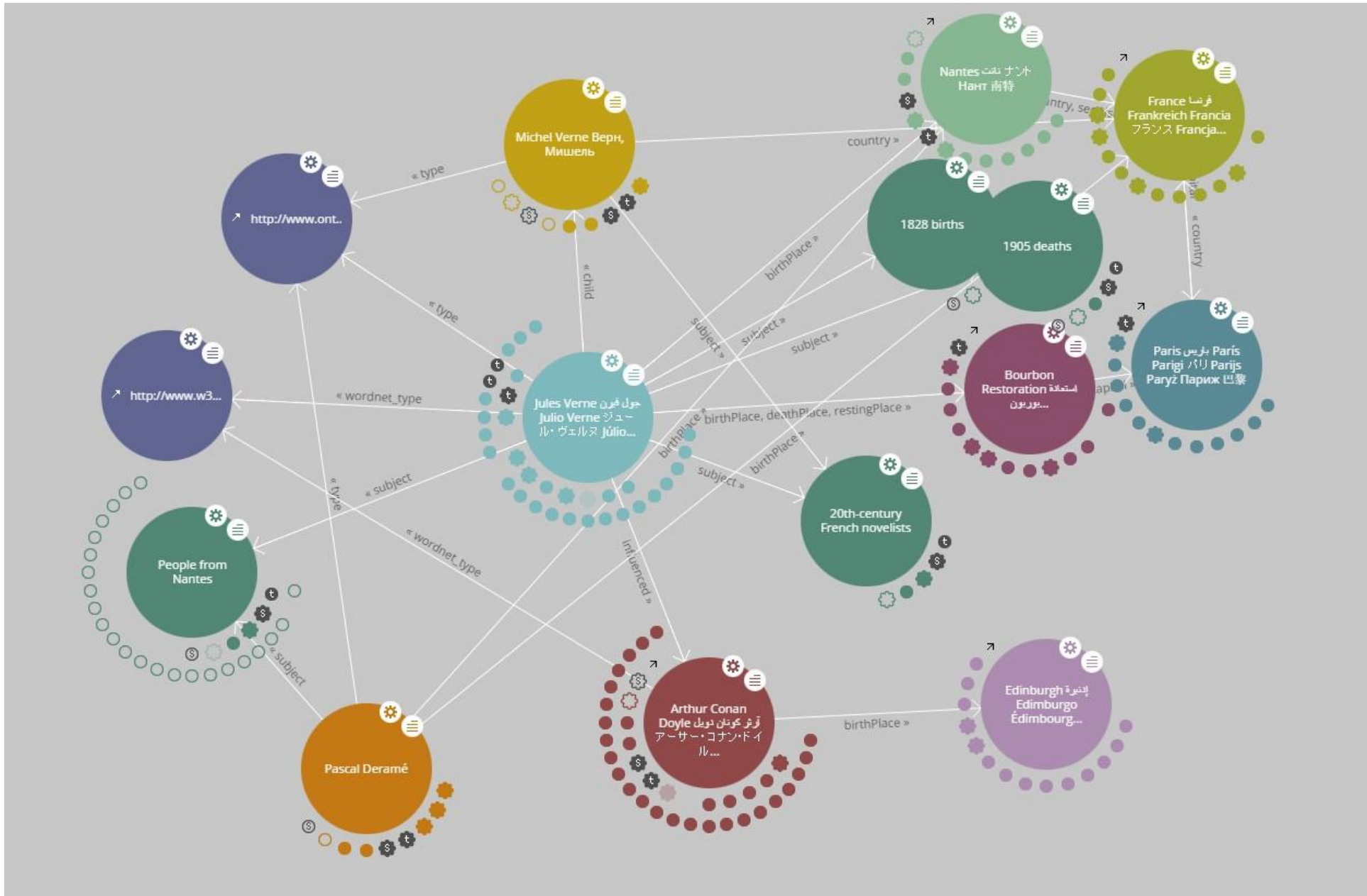
TEST

By using LodLive online to explore dbpedia resources, search for Johann
Strauss II http://en.lodlive.it/
 - who is he?
 - where was he born? where did he died?
 - Is he the son of Johann Strauss?
 - find which type are associated to him
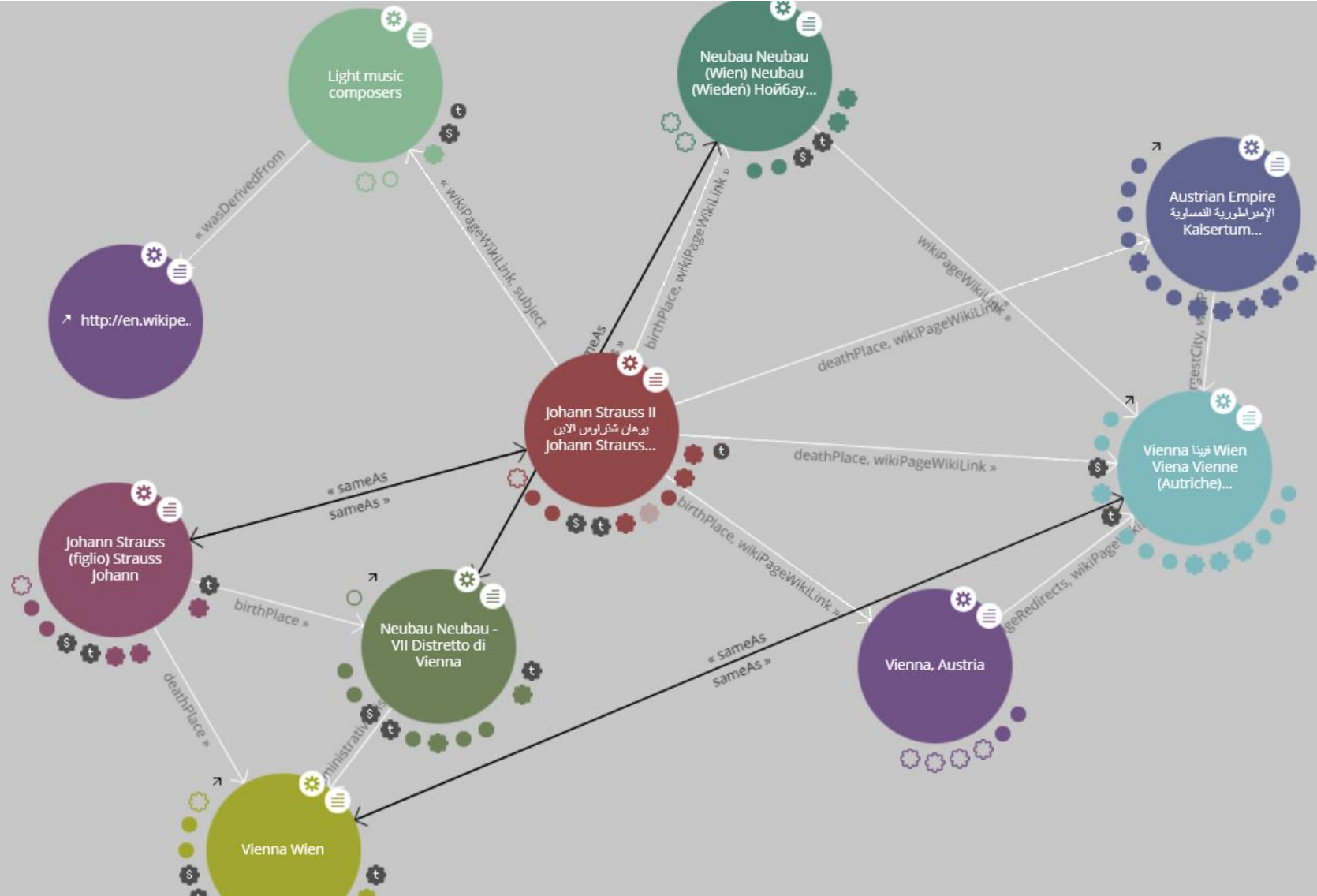
type
Thing Person Person Agent NaturalPerson
Q215627 Q24229398 Q5 Agent Person
WikicatRomanticComposers Artist109812338
CausalAgent100007347 Composer109947232
Creator109614315 Entertainer109616922
LivingThing100004258 Musician110339966
Musician110340312 Object100002684
Organism100004475 Performer110415638
Person100007846 PhysicalEntity100001930
Violinist110754578 Violist110755080
Whole100003553 YagoLegalActor
YagoLegalActorGeo Wikicat19th-
centuryAustrianPeople Wikicat19th-
centuryClassicalComposers Wikicat19th-
centuryComposers
WikicatAustrianClassicalComposers
WikicatAustrianComposers
WikicatAustrianPeopleOfJewishDescent
WikicatAustrianViolinists WikicatBalletComposers
WikicatVienneseComposers WikicatViolists
WikicatOperaComposers
WikicatPeopleFromNeubau
WikicatPeopleFromVienna

http://kreusch-sheet-music.net/eng/index.php?p...

LODmilla

Micsik, András, Sándor Turbucz, and Zoltán Tóth. "Exploring publication metadata graphs with the LODmilla browser and editor." *International Journal on Digital Libraries* 16.1 (2015): 15-24.

Micsik, András, Sándor Turbucz, and Zoltán Tóth. "Browsing and traversing linked data with lodmilla." *ERCIM News* 2014.96 (2014): 35-36.

TEST

Using LODMilla search and add the following node from Dbpedia:

- Johann Strauss II
- Vienna
- The Blue Danube
- Austria
- Johann Strauss I
- Wolfgang Amadeus Mozart
- Composer
- Musician

- Look at the connections between nodes

LODEX

It is a tool for producing a representative summary of a Linked open Data (LOD) source starting from scratch, thus supporting users in exploring and understanding the contents of a dataset.

LODeX extracts statistical indexes that uses to build the representative summary, by quering the SPARQL endpoint of a LOD source.

- LODeX 2.0 (http://www.dbgroup.unimo.it/lodex2) includes the possibility to compose visual queries by selecting objects from the representative summary of a LOD source
- LODeX Cluster (http://www.dbgroup.unimo.it/lodex2/testCluster) provides a more concise schema for huge datasets

LODeX Architecture
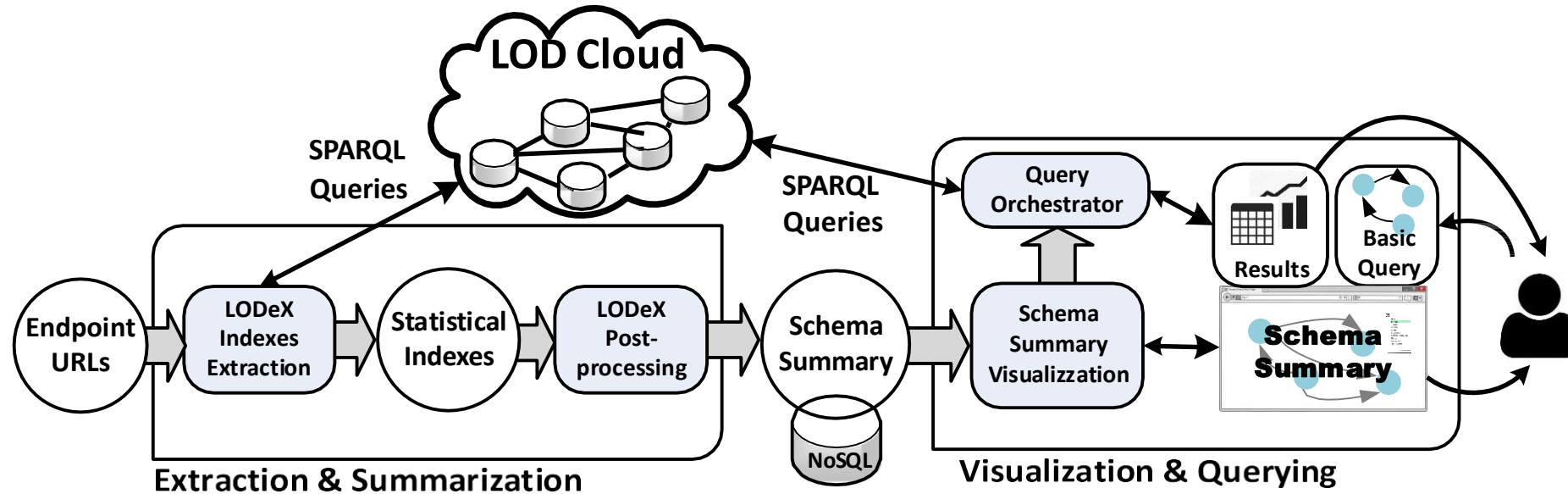
Benedetti, et al. (2015), Exposing the Underlying Schema of LOD Sources.   2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) (IEEE) ISBN: 9781467396189

Benedetti, et al. (2015), Visual Querying LOD sources with LODeX. Proceedings of the 8th International Conference on Knowledge Capture (ACM)

Benedetti, et al. (2015), LODeX: A tool for Visual Querying Linked Open Data. Proceedings of the ISWC 2015 Posters & Demonstrations Track @ (ISWC 2015), n. volume 1486

Benedetti, et al. (2014), A Visual Summary for Linked Open Data sources. ISWC 2014 Posters & Demo Track, Riva del Garda, Italy, ISSN: 1613-0073

Benedetti, et al. (2014), Online Index Extraction from Linked Open Data Sources. Second International Workshop on Linked Data for Information Extraction (LD4IE) @ (ISWC 2014), Riva del Garda, Italy, ISSN: 1613-0073

A runnin
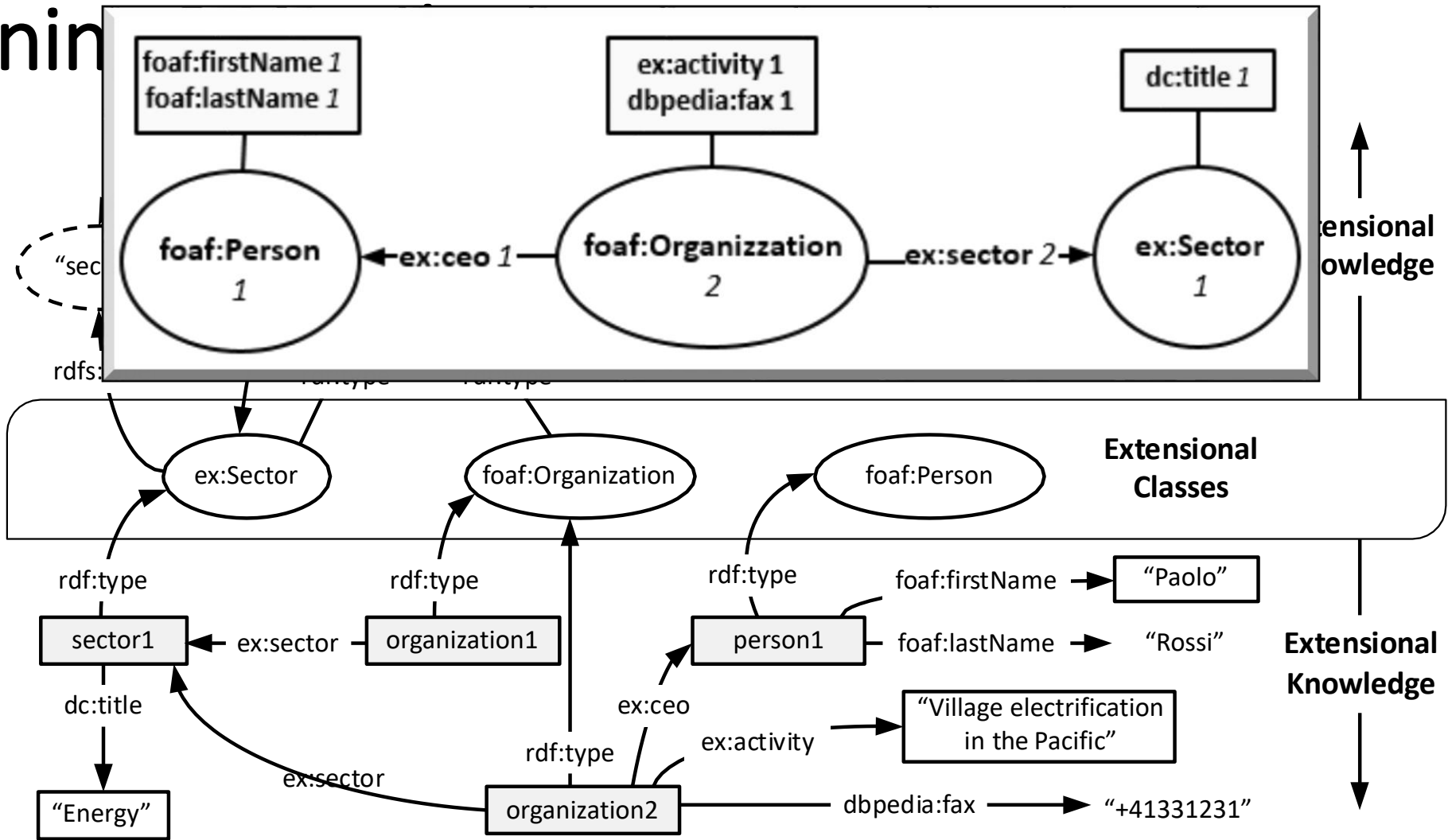
The information contained in the Intensional knowledge can be incomplete or absent

Schema Summary – Building a Visual Query

Linked Clean Energy Data (reegle.info)

Refinement Panel

TEST

By using Lodex http://www.dbgroup.unimore.it/lodex2/ find, navigate and explore the following datasets:

- **European Television Heritage**
  - How many classes it has? How many properties it has?
  - How many vocabulary are used?

- **Nobel Prizes**
  - How many vocabulary are used?
  - Define a query that select person (label, gender,name) that won a Nobel Prize , i.e.have an Award (year,label), add also the Category of the Award if it exists

Conclusions

- Analysis of the needs for visualization in the LOD context

- Practical use of some LOD browsers and visualization toolkits

- Navigation and exploration of some datasets and the construction of different visualizations

Actual limitations and challenges

- Most of the LOD visualization tools are still in-lab prototypes

- Lots tools allow the exploration of a limited list of datasets or have limitations in terms of size, format (SPARQL endpoint/RDF dumps) of the datasets they can explore

- SPARQL endpoints might be offline or have bad performance such as taking long time to respond to some queris.

- For dealing with BOLD, graph simplification is needed:
  - reducing size could be possible through filtering or aggregation

My vision

Aggregated View
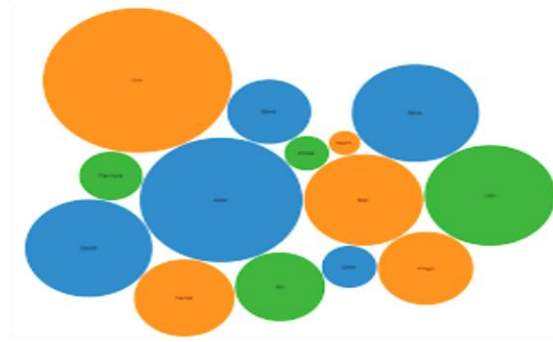
Schema Summary

RDF Graph

THANK YOU

Feel free to contact me at
laura.po@unimore.it

You can find more information on
my research and my group at
www.dbgroup.unimore.it

Slide are available on
http://www.slideshare.net/polaura