

Knowledge Graphs have unfinished business...



WU
WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

@AxelPolleres

<http://polleres.net/presentations>

Panel Questions - Short answers:

- **What are the problems of current KG approaches?**

Knowledge Graphs have unfinished business...

i.e., we can't just go along, without solving some core problems
(that don't go away by rebranding Semantic Web and Linked Data)!

- **What are your visions or thoughts about KG applications in academia and industry**

Decentralised Data Eco-Systems beyond centralised KGs!

- **What are the future directions for KG**

Natural **Data** Understanding, i.e. "NoNLP"!

Contextualisation ... still ;-)

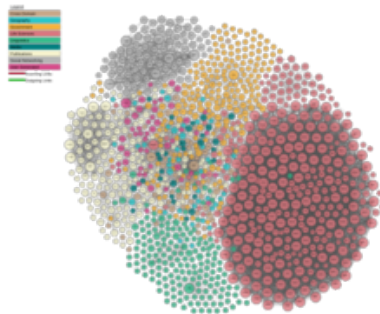
Scalable Linked Data Processing... now for real!

Knowledge Graphs have unfinished business...

... with **Linked Open Data**

more specifically, with **BOTH**

Linked Data



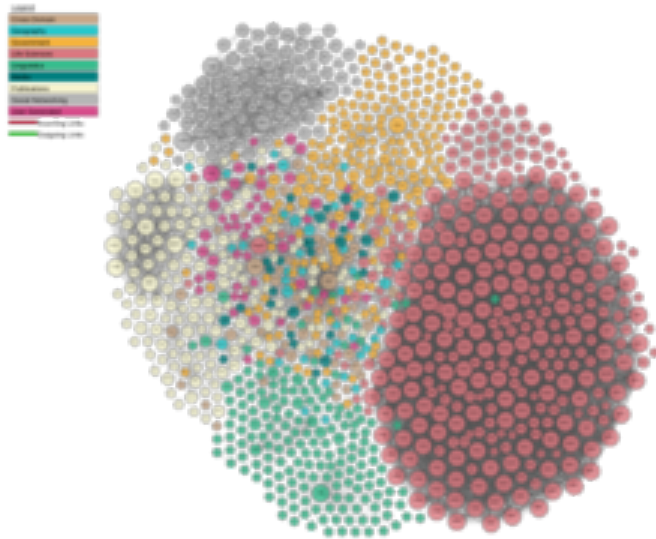
AND

Open Data



Knowledge Graphs and Linked Data

Beyond Dbpedia and Wikidata... we had the vision of a **Web of Open KGs** that would be *interlinked* and *queryable*



Did it produce more than this lovely picture?

Yet: we are left with tons of issues [1]!

... e.g.
linkage between KGs still sparse [2],
but also others, e.g. availability & sustainability
(beyond a PhD lifetime) of Open KGs!

[1] Axel Polleres, Maulik Rajendra Kamdar, Javier D. Fernández, Tania Tudorache, and Mark A. Musen. A more decentralized vision for linked data. 11(1):101--113, January 2020. SWJ 10-years special issue. [[http](#)]

[2] Armin Haller, Javier D. Fernández, Maulik R. Kamdar, and Axel Polleres. What are links in linked open data? a characterization and evaluation of links between knowledge graphs on the web. *ACM Journal of Data and Information Quality (JDIQ)*, to appear, Pre-print available <https://epub.wu.ac.at/7193/>.

Knowledge Graphs and Open Data

Open Data had a boost some years ago and is still trending, esp. for Government Data & Transparency!

Yet: hardly any of this data is linked/linkable or even findable!



A good start, but not much better either!

The screenshot shows the 'data.gv.at' website with a search bar containing 'Leopoldstadt'. Below the search bar, there are search results for 'Suchergebnisse - Daten & Dokumente'. The first result is 'Stadtplan von Anton Behsel 1825' dated 29.03.2019, with a link to 'Wien'. Other results include 'Stadt Wien/Magistratsabteilung 8 - Wiener Stadt- und Land...' and 'Georeferenzierter Stadtplan der Inneren Stadt und...'. There are also options for 'WMS', 'GIF', 'JPEG', and 'PNG' formats. At the bottom, there are links for 'Alle Datensätze', 'Alle Dokumente', and 'Alle News'.

Google Dataset Search: Building a search engine for datasets in an open Web ecosystem

Natasha Noy
noy@google.com
Google AI
Mountain View, California

Matthew Burgess
matburg@google.com
Google AI
Mountain View, California

Dan Brickley
danbri@google.com
Google
Mountain View, California

The screenshot shows the Google Dataset Search interface with the search term 'Leopoldstadt'. The results list several datasets from Statista:

- Mitarbeiter der Erste Group nach Ländern 2018** (de.statista.com) with a button to 'Explore at de.statista.com'.
- Cost-Income-Ratio der Erste Group bis 2018** (de.statista.com).
- Betriebserträge und -aufwendungen der Erste Group** (de.statista.com).

Metadata for the first result includes: 'Dataset provided by Statista', 'Time period covered 2018', and 'Area covered Österreich'.

Knowledge Graphs and Open Data

Open Data had a boost some years ago and is still trending, esp. for Government Data & Transparency!

Yet: hardly any of this data is linked/linkable or even findable!

- (Open) Knowledge Graphs can be used to **link Open Data!**
 - challenges: Coverage, numeric data, OD Quality/KG Quality

Temporal filters

Filter results by timespan: Off Title & description CSV columns

1/2010 1/2020

Filter pattern

Apply Filter

Linz

Republic of Austria > Oberösterreich > Linz Stadt > Linz

Spatial entity or Full-text results

Hotspot - Standorte - Hotspot Standorte [Stadl Linz](#)
POI's (Points of Interest) für Hotspot (freies, kostenloses WiFi) in der Stadt Linz. Die Koordinaten sind im im EPSG-Codes WGS84 verfügbar. <http://data.gv.at>

Nummer	Latitude	Longitude	Name	Kurztext	Start im Jahr	Ende im Jahr	Stadt	Postleitzahl
4007	48.304793	14.299414	Hotspot Linz - Rotes...	Hier ist nur einer v...	2013	0	Linz	4020

14 oberösterreichischen Gemeinden

Gemeindennummer	Gemeindenname	Ordentliche Einnahme...	Ordentliche Ausgaben	Außere Einnahme
40101	Linz	628704196,3	718773006,9	131859

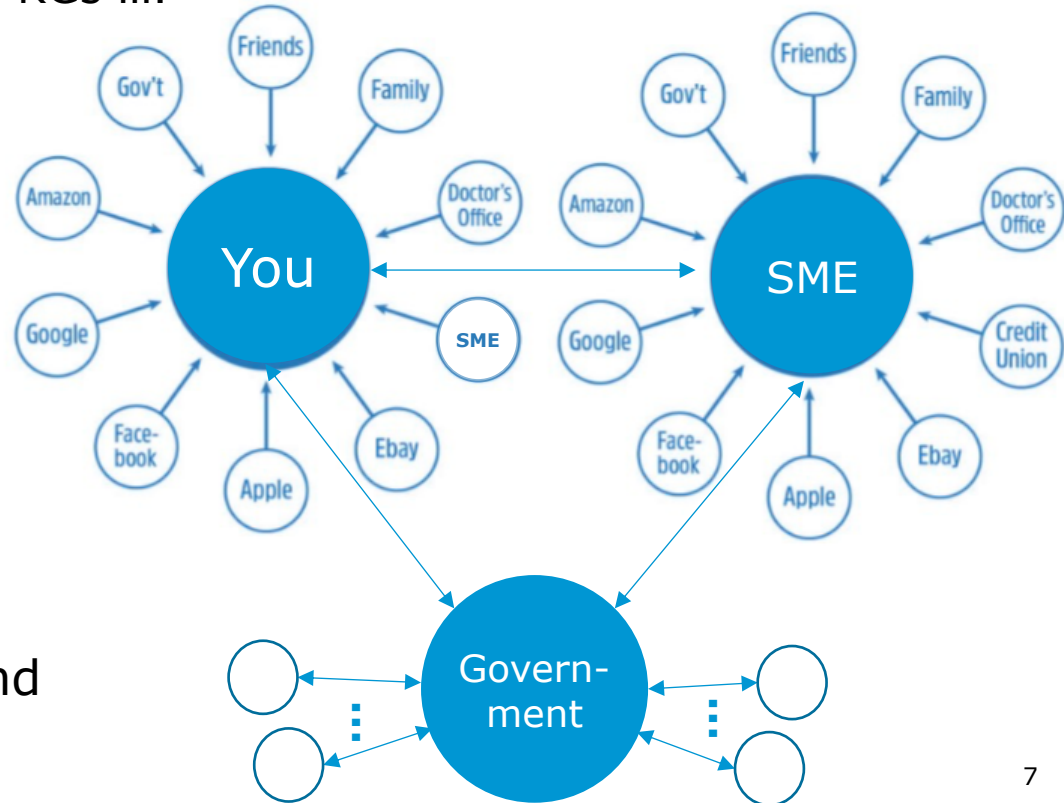
Keynote "How do Linked Data, Open Data, and Knowledge Graphs interplay?" at DEXA 2019, August 2019, Linz, Austria.

polleres.net/presentations/20190827DEXA_keynote.pdf

What are your vision or thoughts about KG applications in academia and industry

We need to build on our strength!

- Rather than (more) centralised KGs
 - Make a **Decentralised** Data Eco-Systems work!
- enable trusted data sharing
- enable FAIR principles



Applications galore!

e.g.: bring together
personal data spaces and
industrial data spaces!

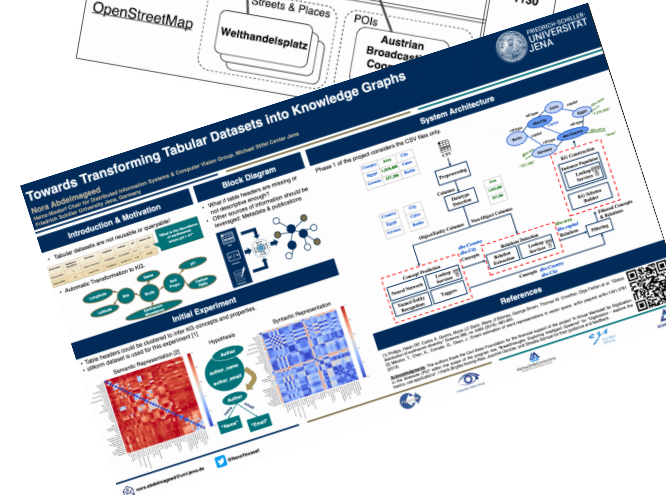
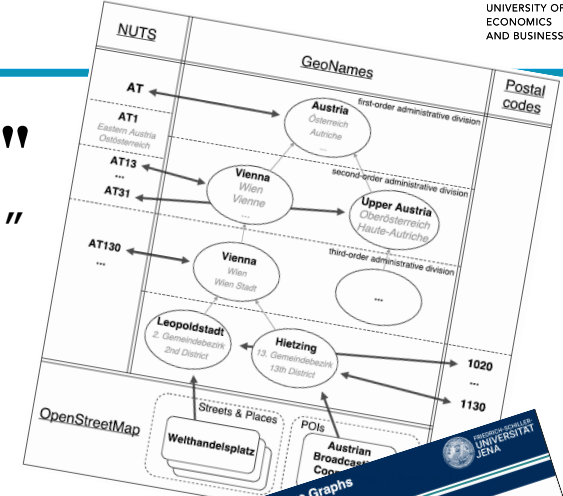
What are the future directions for KG 1/3

Towards "Natural Data Understanding"

"Natural Data" = "Structured Data as it occurs in the wild"

- Heterogenous data assets
- From different sources and origins
- Different Formats
- Different Semantics
- Sparse descriptive metadata
- Raw/not necessarily for human consumption

→ **(Still)** Hard to search and integrate!



[3] Sebastian Neumaier and Axel Polleres. Enabling spatio-temporal search in open data. *Journal of Web Semantics (JWS)*, 55, March 2019.

[4] Sebastian Neumaier, Jürgen Umbrich, Josiane Parreira, and Axel Polleres. Multi-level semantic labelling of numerical values. (ISWC2016)

What are the future directions for KG 2/3

- Scalable Linked Data Processing... now for real!
 - (SPARQL EP, TPF, SMART-KG --> there's something moving here!

Amr Azzam, Ruben Taelman, Axel Polleres

Abstract

A multitude of Linked Data Fragments (LDF) server interfaces have been proposed to expose Knowledge Graphs (KGs) on the Web. Each interface leads to different trade-offs when clients execute queries over them, such as how query execution effort is distributed between server and client. There is *no single silver bullet that works best everywhere*. Each of these interfaces has diverse characteristics that vary the performance based on server load, client resources, and network bandwidth. In this paper, we lay out the high-level ideas to introduce a hybrid LDF framework that can expose multiple interfaces based on a server-side cost model. In addition, we sketch a negotiation protocol through which clients can determine desirable interfaces during query planning using a client-side cost model.

Linked Data Fragments Spectrum

Linked Data Fragment Framework (LDF) Proposed to design new mixes of trade-offs.

Interface	Availability	Client Cost	Server Cost	NW Load
Data Dump	High	High	High	High
smart-KG WebCon 2020	High	Avg.	Avg.	Avg.
Triple Pattern Fragment (TPF) ISWC 2014	High	Low	High	High
SaGe WebConf 2019	Avg.	Low	Low	Low
SPARQL Endpoint	Low	Low	Low	Low

Cost-based Hybrid Framework

The goal of our framework is to expose

```
GetInterfaces(q, metrics, interfaces, GetValue, GetThreshold)
allowedInterfaces = []
// ...
```

- Contextualisation ... still ;-)
 - next step: Represent and query at scale contextualised KGs
 - Provenance, Policy-constrained Graphs/Triples, Property Graphs, SPARQL*

What are the future directions for KG 3/3

• *Managing context makes things even harder! → but is also doable!*

- However, Wikidata has more complex info: (**temporal** context, **provenance**,...)

Which cities in the UK have reached 1M in which year?

... Can I query that with SPARQL? **Yes!**

The screenshot shows the Wikidata Query Service interface. At the top, there are navigation buttons for 'Examples', 'Help', and 'More tools'. Below is a SPARQL query editor with a query that filters for cities with a population greater than 1,000,000 and groups them by city. The results are displayed in a table with columns for population, point in time, and determination method. The first result shows a population of 8,416,535±0 in 2012, determined by estimation, with a reference URL from the UK's Office for National Statistics. The second result shows a population of 1,011,157±0 in 1801, determined by census, with a reference URL from the Vision of Britain project.

```
1
2 SELECT ?city (min(?time) as ?year) WHERE {
3   ?city wdt:P31/wdt:P279* wd:Q515.
4   ?city wdt:P17 wd:Q38 .
5   ?city p:P1082 ?statement .
6   ?statement <http://www.wikidata.org/prop/statement/value/P1082> ?value .
7   ?statement <http://www.wikidata.org/prop/qualifier/P585> ?time .
8   ?value <http://wikiba.se/ontology#quantityAmount> ?population .
9   FILTER (?population > 1000000 )
10  } GROUP BY ?city
```

population	point in time	determination method	reference URL
8,416,535±0	2012	estimation	http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-england-and-wales/mid-2012/mid-2012-population-estimates-for-england-and-wales.html
1,011,157±0	1801	census	http://www.visionofbritain.org.uk/data_cube_page.jsp?data_theme=T_POP&data_cube=N_TOT_POP&u_id=10097836&c_id=10001043&add=N

BTW, seemingly not yet doable in

Which cities in the UK have reached 1M in which year?



Take-home summary

if I could predict the future, I wouldn't be here...

... rather an invitation (especially to young researchers):

p.s.: Don't **only** work on DL, I know you do anyway,
there's plenty of other stuff to do ;-)

There's still a lot of unfinished business and unfulfilled promises from Semantic Web, Linked Data, etc. that (despite "rebranding" into KGs)

- haven't been solved
- should and could be solved/tackled with more research!

Let's do this together!

<https://www.cost.eu/cost-action/distributed-knowledge-graphs/> COST Action

<https://knowgraphs.eu/> ITN