



# SEO4OLAP – Search Engine Optimized Presentation of Statistical Linked Data

Daniel Breucker, Benedikt Kämpgen, Andreas Harth

INFORMATIK, BDS DST 2016, 2016-09-30



30 Jahre FZI



# Motivation

- G8 Open Data Charter pushes open statistics
- Large amounts of numeric data already available
- In theory: Important information at your fingertips
  - Computations
  - Comparisons
  - ...

Eurostat - Data Explorer

appsso.eurostat.ec.europa.eu/nui/show.do?wai=true&dataset=lfsi\_emp\_a

Apps For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#)

eurostat

View Table Select Data

Explanatory texts (metadata) Information Download Preview

**Employment (main characteristics and rates) - annual averages**

Last update: 21-01-2016

Table Customization

Cell Formatting  1.234,56  1,234.56  1 234.56

Hide empty lines  Hide flags/footnotes

TIME GEO SEX INDIC EM

Total Total employment (resident population concept - LFS)

Update

GEO	TIME	2005	2006	2007	2008	2009	2010
European Union (28 countries)		212,063.2	216,155.8	220,363.1	222,875.5	218,952.2	216,865.8
European Union (27 countries)		210,490.3	214,569.4	218,629.3	221,104.9	217,195.2	215,175.6
European Union (15 countries)		168,928.6	171,934.7	174,916.2	176,615.2	173,438.9	172,789.1
Euro area (19 countries)		139,775.0	142,542.7	145,354.0	146,758.6	143,819.6	142,942.0
Euro area (18 countries)		138,340.6	141,113.7	143,902.4	145,331.5	142,502.2	141,694.3
Euro area (17 countries)		137,368.2	140,082.8	142,845.0	144,276.6	141,593.7	140,843.6
Belgium		4,235.4	4,264.0	4,380.3	4,445.9	4,420.7	4,488.7
Bulgaria		2,981.9	3,110.0	3,252.6	3,360.7	3,253.6	3,075.3 (b)
Czech Republic		4,764.0	4,828.1	4,922.0	5,002.5	4,934.3	4,885.2
Denmark		2,752.4	2,805.4	2,803.7	2,852.8	2,770.5	2,706.1
Germany (until 1990 former territory of the FRG)		36,361.6 (b)	37,172.3	37,988.7	38,541.5	38,471.1	38,737.8
Estonia		615.6	651.7	657.6	656.0	593.9	568.0
Ireland		1,952.0	2,043.7	2,143.1 (b)	2,128.4	1,961.3	1,882.2
Greece		4,443.5	4,527.5	4,564.0	4,610.5	4,556.0	4,389.8
Spain		19,207.0 (b)	19,939.1	20,579.9	20,469.7	19,106.9	18,724.5
France		24,984.2	25,150.1	25,586.7	25,926.4	25,674.4	25,730.9
Croatia		1,572.9	1,586.3	1,733.7	1,770.5	1,757.0	1,690.2
Italy		22,407.0	22,757.6	22,894.4	23,090.3	22,698.7	22,526.9
Cyprus		348.0	357.3	377.9	382.9	382.9 (b)	395.2
Latvia		972.3	1,030.9	1,057.4	1,054.9	908.5	850.7
Lithuania		1,434.4	1,428.9	1,451.5	1,427.1	1,317.4	1,247.7
Luxembourg		193.6	195.3	202.9 (b)	202.4	217.2	220.8
Hungary		3,901.5	3,928.4	3,902.0	3,848.3	3,747.7	3,732.4
Malta		149.4 (b)	151.2	155.4	158.6	159.5	162.6
Netherlands		8,110.9	8,260.9	8,463.5	8,592.7	8,596.1	8,370.2 (b)
Austria		3,747.4	3,826.0	3,923.6	3,994.1	3,982.2	4,016.7

ec.europa.eu/eurostat/cache/metadata/en/lfsi\_act\_a\_esms.htm

# Problem

- How to find the right number for a specific question?
- Standard way = search engines
- Numeric data often not properly indexed
- Reasons
  - Deep web
  - CSV, SDMX...
  - JavaScript
  - ...

The screenshot shows a Google search results page for the query "employment germany 2014". The browser's address bar shows the URL "https://www.google.de/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=employment germany 2014". The search bar contains the text "employment germany 2014". Below the search bar, there are navigation tabs for "All", "Images", "News", "Videos", "Shopping", "More", and "Search tools". The search results are displayed below, showing "About 120.000.000 results (0,47 seconds)". The first result is a PDF document titled "September 2014 OECD Employment Outlook 2014" from the OECD website. The second result is another PDF document titled "EMPLOYMENT PLAN 2014 GERMANY" from the German government website. The third result is a data visualization titled "Germany Unemployment Rate | 1950-2016 | Data | Chart ..." from tradingeconomics.com. The fourth result is a webpage titled "National economy & environment - Labour market - Federal ..." from the German Federal Statistical Office (Destatis).

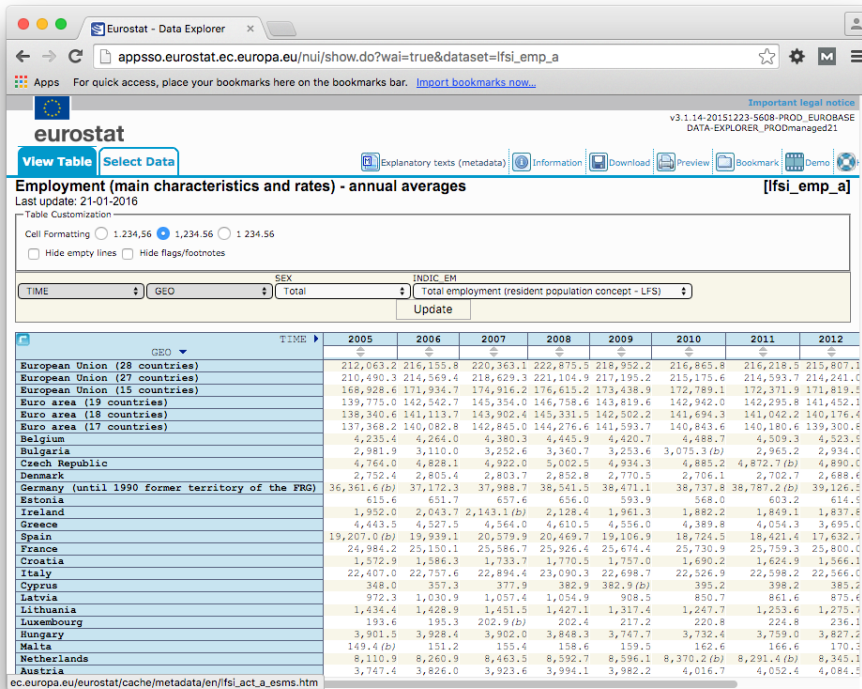
# At the same time... Statistical Linked Data

*How to publish numeric data on the Web?*

**"20,512"**



# Problem



Important legal notice  
v3.1.14-20151223-5608-PROD. EUROBASE  
DATA-EXPLORER\_PRODmanaged21

View Table Select Data

Explanatory texts (metadata) Information Download Preview Bookmark Demo

Employment (main characteristics and rates) - annual averages [lfsi\_emp\_a]

Last update: 21-01-2016

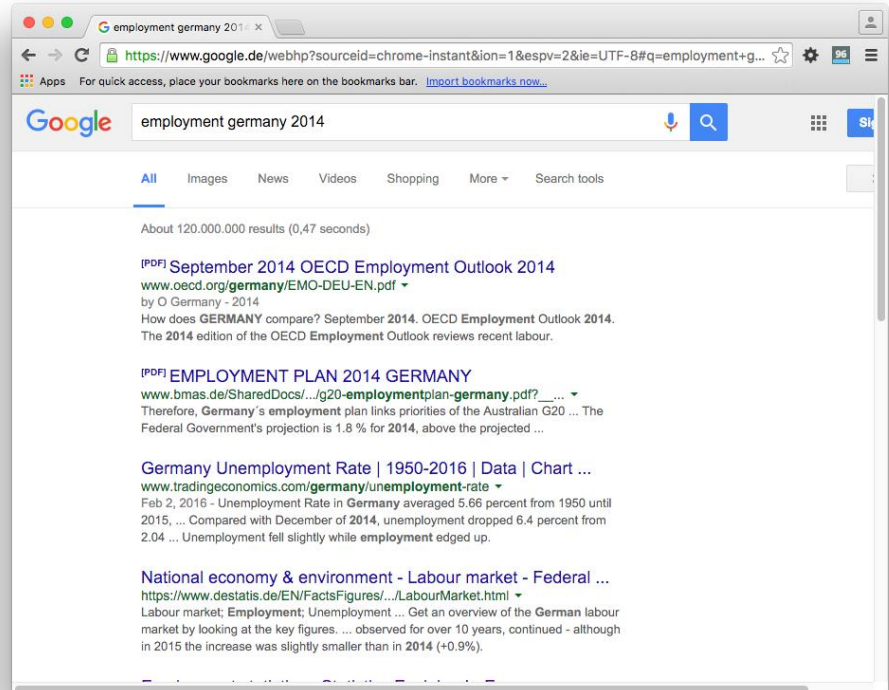
Table Customization

Cell Formatting: 1. 234.56 2. 1,234.56 3. 234.56

Hide empty lines Hide flags/footnotes

SEX: Total INDIC EM: Total employment (resident population concept - LFS)

GEO	2005	2006	2007	2008	2009	2010	2011	2012
European Union (28 countries)	212,063.2	216,155.8	220,363.1	222,875.5	218,952.2	216,865.8	216,218.5	215,807.1
European Union (27 countries)	210,490.3	214,569.4	218,629.3	221,104.9	217,193.2	215,175.6	214,593.7	214,241.0
European Union (15 countries)	168,928.6	174,934.7	174,916.2	176,615.2	173,438.9	172,789.1	172,371.9	171,819.2
Euro area (19 countries)	139,775.0	142,542.7	145,354.4	146,758.6	143,819.6	142,942.0	142,295.8	141,452.1
Euro area (18 countries)	138,340.6	141,113.7	143,902.4	145,331.5	142,502.2	141,694.3	141,042.2	140,176.4
Euro area (17 countries)	137,368.2	140,082.8	142,845.0	144,276.6	141,593.7	140,843.6	140,180.6	139,300.5
Belgium	4,235.4	4,266.0	4,380.3	4,445.8	4,420.7	4,488.7	4,509.3	4,525.5
Bulgaria	2,981.9	3,110.0	3,252.6	3,360.7	3,253.6	3,075.3(b)	2,965.2	2,934.0
Czech Republic	4,764.0	4,828.1	4,922.0	5,002.3	4,934.3	4,885.2	4,872.7(b)	4,890.0
Denmark	2,752.4	2,805.4	2,803.7	2,852.8	2,770.5	2,706.1	2,702.7	2,688.4
Germany (until 1990 former territory of the FRG)	36,361.6(b)	37,172.3	37,988.7	38,541.5	38,671.1	38,737.8	38,787.2(b)	39,126.5
Estonia	615.6	651.7	657.6	656.0	593.9	568.0	603.2	614.5
Ireland	1,952.0	2,043.7	2,143.1(b)	2,128.4	1,961.3	1,882.2	1,849.1	1,837.8
Greece	4,443.5	4,527.5	4,564.0	4,610.5	4,556.0	4,389.8	4,054.3	3,695.0
Spain	19,207.0(b)	19,939.1	20,579.9	20,469.7	19,106.9	18,724.5	18,421.4	17,632.3
France	24,984.2	25,150.1	25,586.7	25,926.4	25,674.4	25,730.9	25,759.3	25,800.0
Croatia	1,572.9	1,586.3	1,733.7	1,770.5	1,757.0	1,690.2	1,624.9	1,566.1
Italy	22,407.0	22,757.6	22,894.4	23,090.3	22,698.7	22,526.9	22,398.2	22,566.0
Cyprus	348.0	357.3	371.9	382.9	382.9(b)	395.2	398.2	385.2
Latvia	972.3	1,030.9	1,057.4	1,054.9	905.5	850.7	861.6	875.6
Lithuania	1,434.4	1,428.9	1,451.5	1,427.1	1,317.4	1,247.7	1,253.6	1,275.7
Luxembourg	193.6	195.3	202.9(b)	202.4	217.2	220.8	224.8	236.1
Hungary	3,901.3	3,928.4	3,902.0	3,848.3	3,747.7	3,732.4	3,759.0	3,827.2
Malta	149.4(b)	151.2	155.4	158.6	159.5	162.6	166.6	170.3
Netherlands	8,110.9	8,260.9	8,463.5	8,592.7	8,596.1	8,370.2(b)	8,291.4(b)	8,345.1
Austria	3,747.4	3,826.0	3,923.6	3,994.1	3,982.2	4,016.7	4,052.4	4,084.5



employment germany 2014

About 120,000,000 results (0,47 seconds)

[PDF] September 2014 OECD Employment Outlook 2014  
www.oecd.org/germany/EMO-DEU-EN.pdf  
by O Germany - 2014  
How does GERMANY compare? September 2014. OECD Employment Outlook 2014.  
The 2014 edition of the OECD Employment Outlook reviews recent labour.

[PDF] EMPLOYMENT PLAN 2014 GERMANY  
www.bmas.de/SharedDocs/...g20-employmentplan-germany.pdf?...  
Therefore, Germany's employment plan links priorities of the Australian G20 ... The Federal Government's projection is 1.8 % for 2014, above the projected ...

Germany Unemployment Rate | 1950-2016 | Data | Chart ...  
www.tradingeconomics.com/germany/unemployment-rate  
Feb 2, 2016 - Unemployment rate in Germany averaged 5.66 percent from 1950 until 2015, ... Compared with December of 2014, unemployment dropped 6.4 percent from 2.04 ... Unemployment fell slightly while employment edged up.

National economy & environment - Labour market - Federal ...  
https://www.destatis.de/EN/FactsFigures/...LabourMarket.html  
Labour market; Employment; Unemployment ... Get an overview of the German labour market by looking at the key figures. ... observed for over 10 years, continued - although in 2015 the increase was slightly smaller than in 2014 (+0.9%).

## Research questions

- Can we semi-automatically build a version of arbitrary statistics on the web that is indexed by search engines?
- Is it true that the more specific a query, the better does our version in comparison to existing pages?

# Outline

- Motivation/Problem
- **Related Work**
- Approach
- Evaluation
- Discussions
- Conclusions

# Related Work

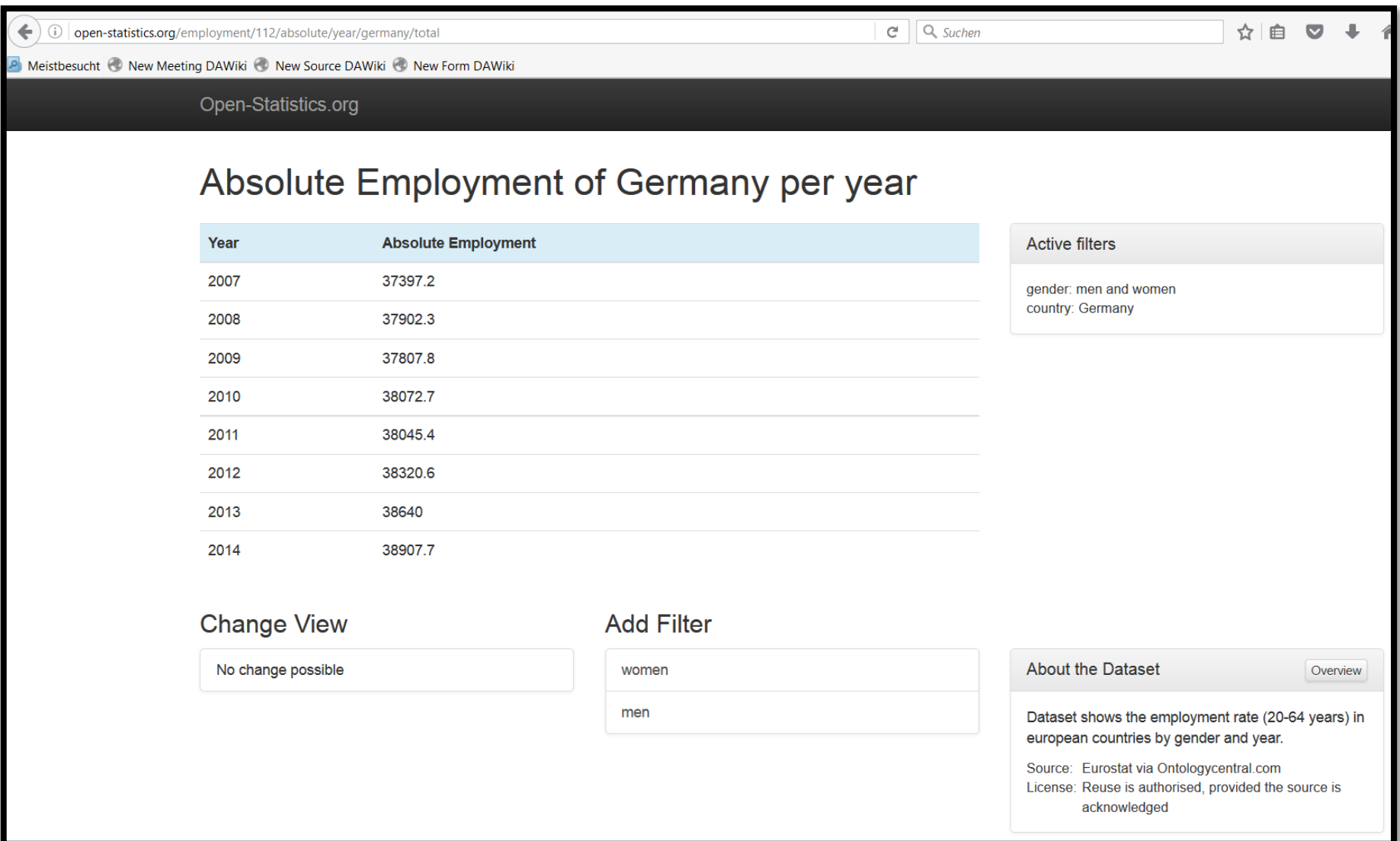
- We generate **search engine optimized pages**
  - Statistical Linked Data visualisation: CODE Query Wizard [Hoe13], CubeViz [SMM+12]
- We use **semantic annotations**
  - Automatic semantic annotation of documents: (automatic) annotation of websites with schema.org [VE13,Mik15]
- We do not do **off-page SEO** (e.g., promoting incoming links)
  - SEO: Web 2.0 techniques [Mal09], empirical evaluations of SEO techniques [SCC13]
- We make **transparent the sources** of numeric data
  - Commercial websites with numeric information: Wolfram Alpha, Tradingeconomics, Statista...



# Outline

- Motivation/Problem
- Related Work
- **Approach**
- Evaluation
- Discussions
- Conclusions

# Approach: For every possible "QB view" a search engine optimized landing page



open-statistics.org/employment/112/absolute/year/germany/total

Meistbesucht New Meeting DAWiki New Source DAWiki New Form DAWiki

## Absolute Employment of Germany per year

Year	Absolute Employment
2007	37397.2
2008	37902.3
2009	37807.8
2010	38072.7
2011	38045.4
2012	38320.6
2013	38640
2014	38907.7

Active filters

- gender: men and women
- country: Germany

Change View

No change possible

Add Filter

- women
- men

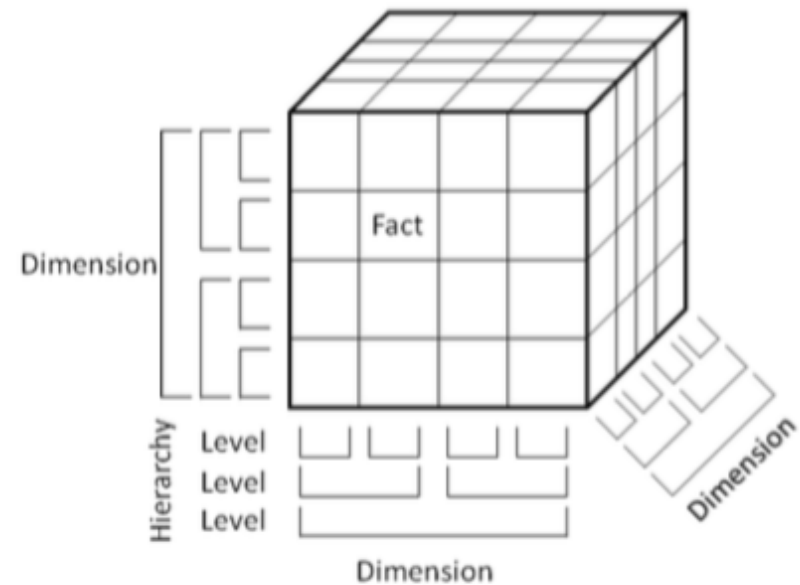
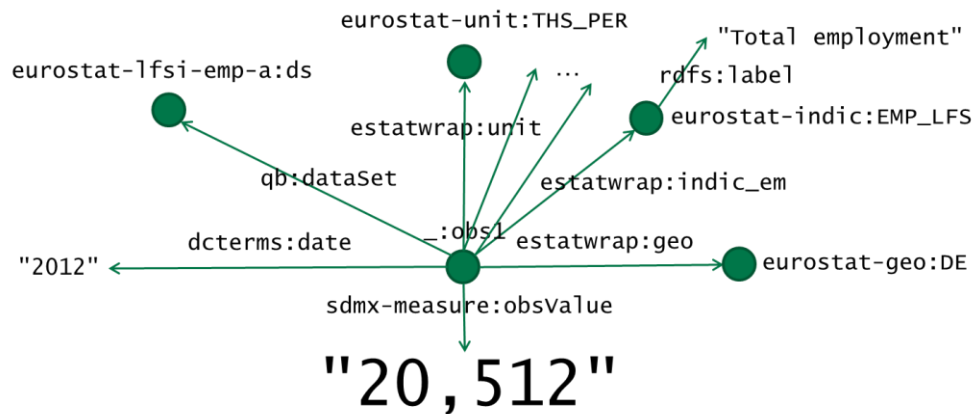
About the Dataset [Overview](#)

Dataset shows the employment rate (20-64 years) in european countries by gender and year.

Source: Eurostat via Ontologycentral.com  
License: Reuse is authorised, provided the source is acknowledged

# QB View?

QB view = OLAP subcube query [KOH12] = Diced X Sliced X Freed

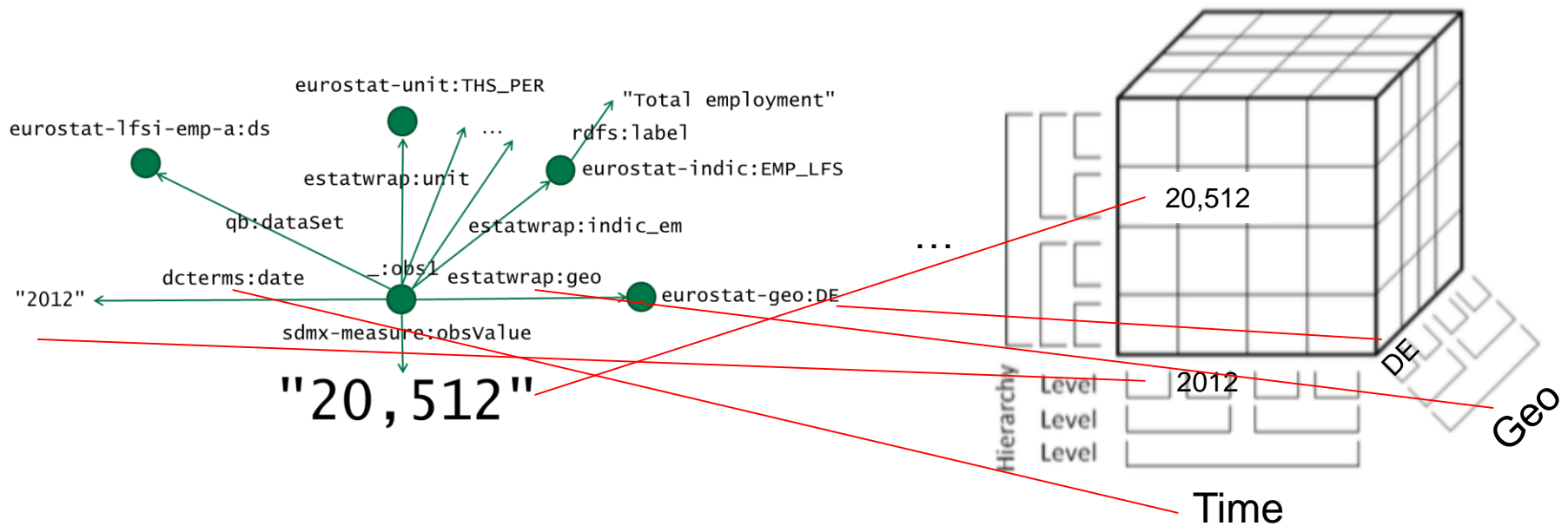


Statistical Linked Data

Data Cube

# QB View?

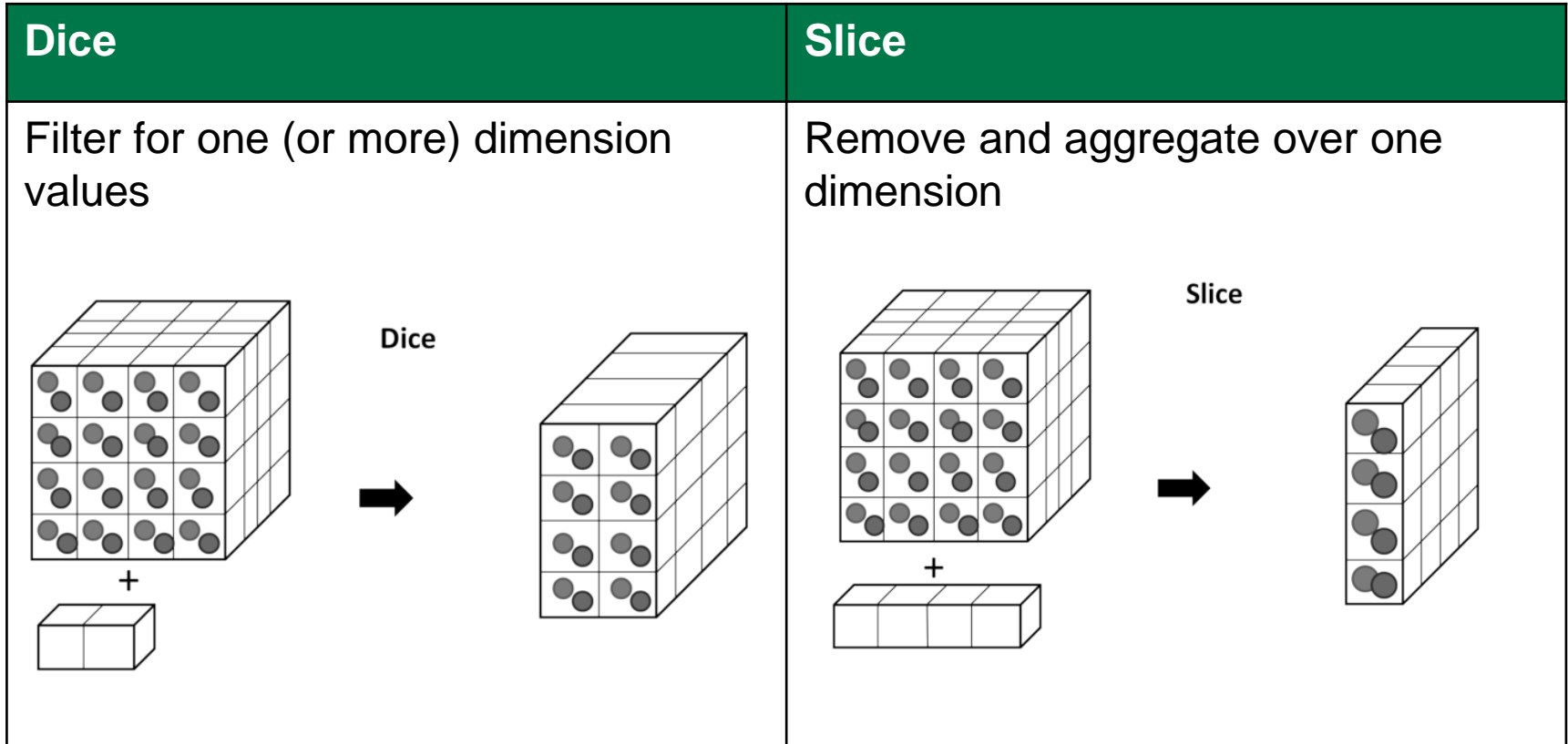
QB view = OLAP subcube query [KOH12] = Diced X Sliced X Freed



Statistical Linked Data

Data Cube

# QB View? (2)



# QB View? (3)

- OLAP subcube query [KOH12]: Diced x Sliced x Freed
- Example:  $(\{(\text{Unit}, \text{THS\_PER}), (\text{Indicator}, \text{Total employment}), \dots\}, \{\text{Sex}, \text{Geo}\}, \{\text{Time}\})$

Dice	Slice	Free
Unit = Thousand persons Indicator = Total employment Age = 15-64 years Geo = DE	Sex = ALL	Time

Time	Value
2007	37,397
2008	37,902
...	...

Numeric data on the landing page

# For every possible "QB view"?

- # QB view = 
$$\prod_{i=1}^n (d_i + 2)$$
- with
  - $n$  : number of dimensions
  - $d_i$  : number of values of dimension  $i$
  - 2 : additional values per dimension for "Sliced" and "Freed"
- Example:
  - 6 dimensions (time, geo, age, unit, sex, indicator)
  - 10 values for time, 42 (geo), 2 (age), 2 (unit), 2 (sex), 2 (indicator)
  - $12 \times 44 \times 4 \times 4 \times 4 \times 4 = 135,168$

# Search engine optimized landing page?

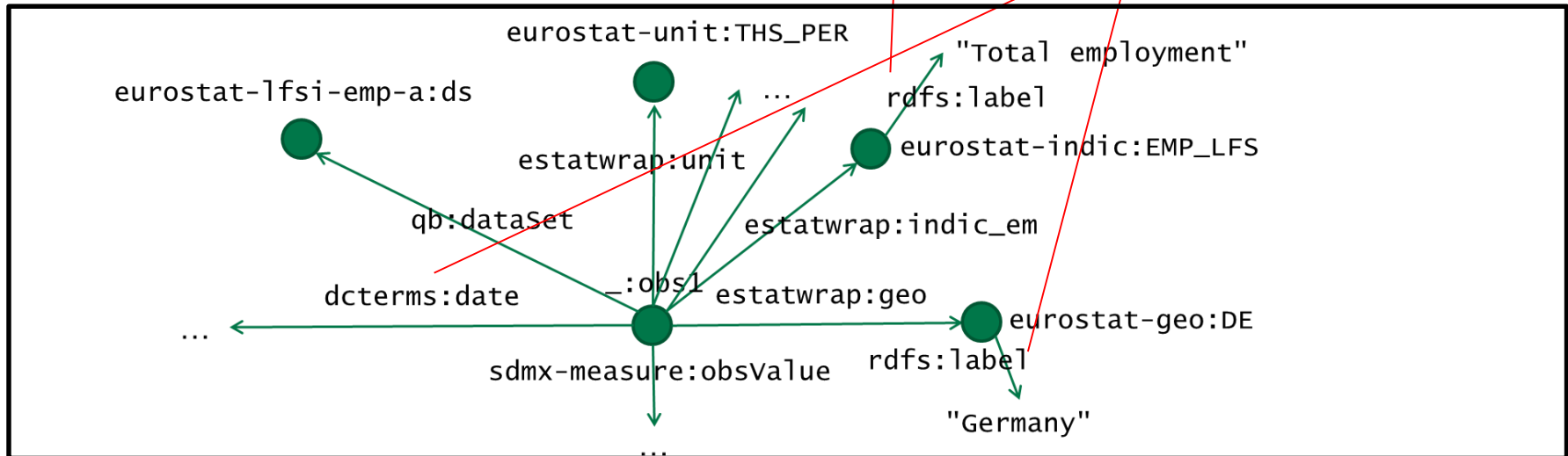
Verbalisation (using templates)

```
<title> ... </title>
<h1>{{diced}}{{freed}}</h1>
...
```



```
<title> ... </title>
<h1>Employment of Germany per year</h1>
...
```

HTML



Statistical Linked Data



# Search engine optimized landing page? (2)

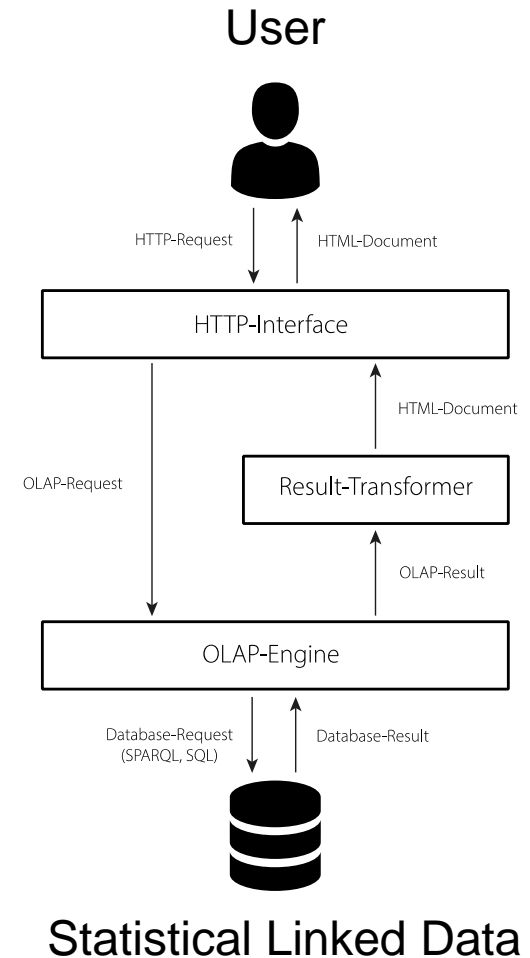
- URI schema
  - $(\{(\text{Unit}, \text{THS\_PER}), (\text{Indicator}, \text{Total employment}), \dots\}, \{\text{Sex}, \text{Geo}\}, \{\text{Time}\})$
  - HTTP Request  
`http://baseUri/endpoint?dataset=id1&freed=id2&diced=id3...`
  - URL Scheme  
`http://baseUri/id1/pattern/id1/id2/id3/...` (pattern defines how many freed dimensions)
  - Bsp:  
[http://example.org/lfsi-emp-a/1/date/total\\_employment/germany](http://example.org/lfsi-emp-a/1/date/total_employment/germany)
- No off-page optimisation (e.g., promoting incoming links)

# Outline

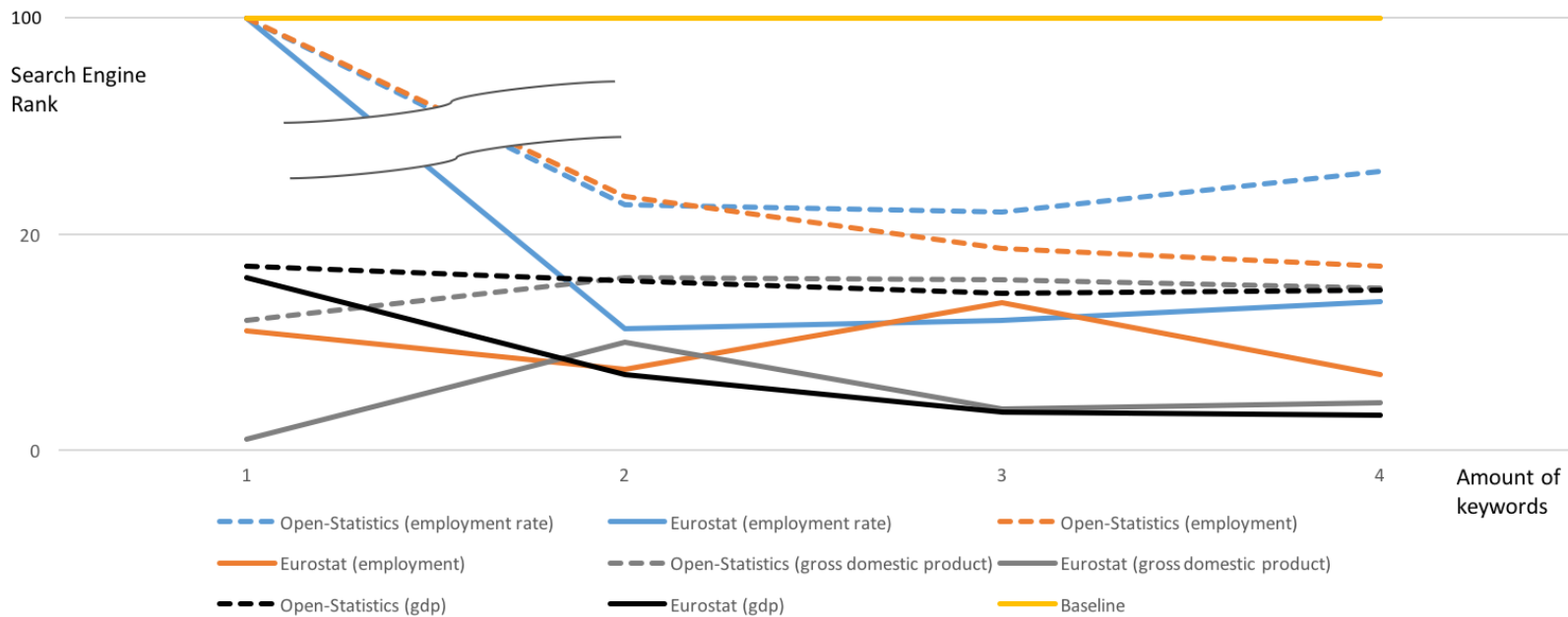
- Motivation/Problem
- Related Work
- Approach
- **Evaluation**
- Discussions
- Conclusions

# Evaluation

- Hypotheses
  - We can build a version of arbitrary Statistical Linked Data that is indexed by search engines
  - The more specific a query, the better does our version in comparison to existing pages
  
- Implementation
  - Website: <http://open-statistics.org/>
  - HTTP-Interface+Result-Transformer: <https://github.com/dbreucker/seo4olap-evaluation/>
  - Statistical Linked Data: Estatwrap: <http://estatwrap.ontologycentral.com/>
  - OLAP-Engine: OLAP4LD [KH14]
  
- Evaluation
  - 2 datasets: Employment in the EU - annual averages, Gross Domestic Product (GDP)
  - Restricted number of pages: max. 2x free dimensions, max. 2x diced dimensions
  - Time: 614 pages generated in 30min

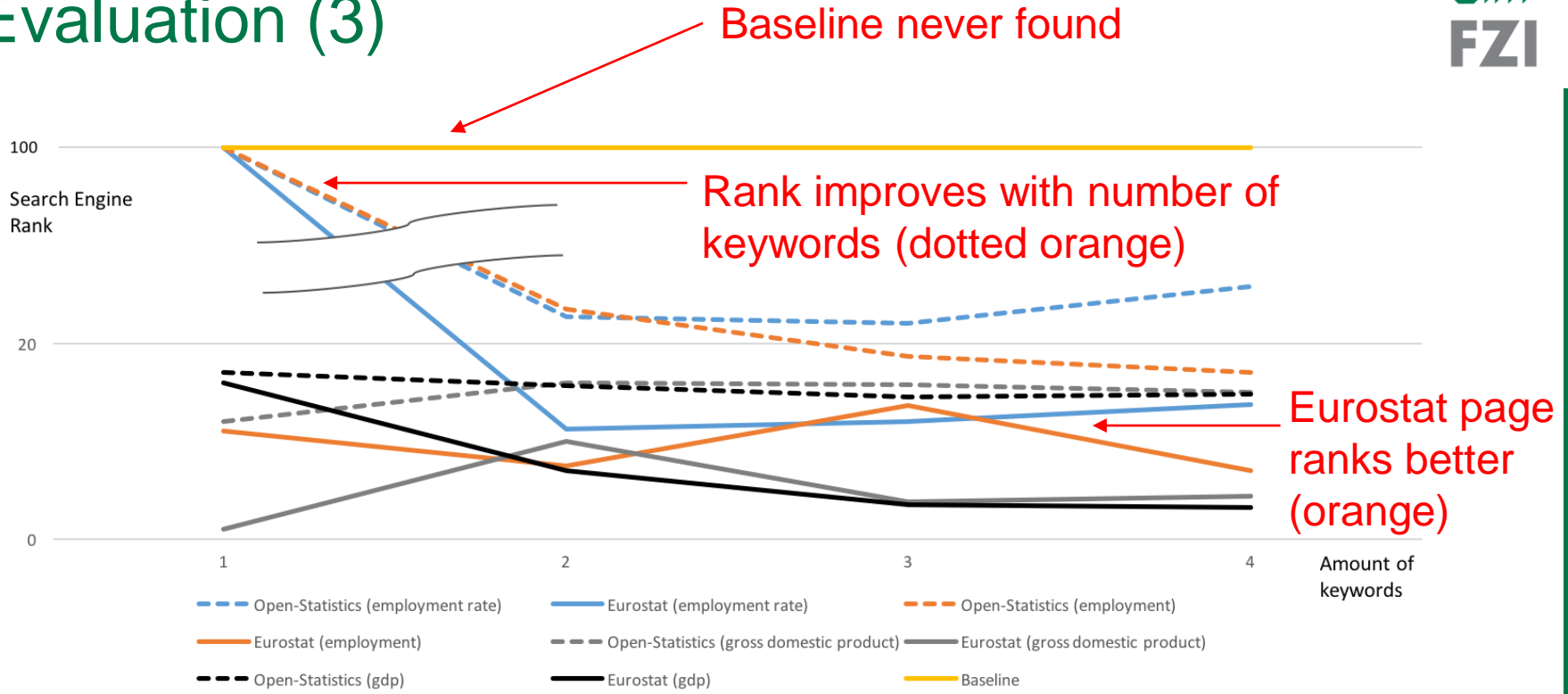


# Evaluation (3)



- Comparing Google rank of 3x page types using CuteRank tool
  - Baseline = Eurostat Pivot Pages
  - Eurostat = Landing pages of Eurostat (without actual numbers)
  - Open-Statistics = Our landing pages

# Evaluation (3)



- Comparing Google rank of 3x page types using CuteRank tool
  - Baseline = Eurostat Pivot Pages
  - Eurostat = Landing pages of Eurostat (without actual numbers)
  - Open-Statistics = Our landing pages

# Outline

- Motivation/Problem
- Related Work
- Approach
- Evaluation
- **Discussions**
- Conclusions

# Discussions

- About generation of pages...
  - **Measure dimension:** some dimensions refer to the indicator and should not be sliced since aggregation would be meaningless
  - **Slice member:** some dimensions contain special type members "total" that should not be included in aggregations
  - **Alternative labels** based on simple rules ("female" => "women")
  - **Unit** not always available in a machine readable format (e.g., ambiguous "ratio or thousand" simply in the textual documentation)
  - Some datasets only mention their **indicator** neither as a specific measure nor in a dimension but only in the title
  - **Schema.org** so far does not provide a way to annotate arbitrary numeric facts for search engines
- About the evaluation...
  - We can built a version of arbitrary Statistical Linked Data that is indexed by search engines – **Confirmed.**
  - The more specific a query, the better does our version in comparison to existing pages – **Valid for some pages but cannot be generally shown.**

# Outline

- Motivation/Problem
- Related Work
- Approach
- Evaluation
- Discussions
- **Conclusions**



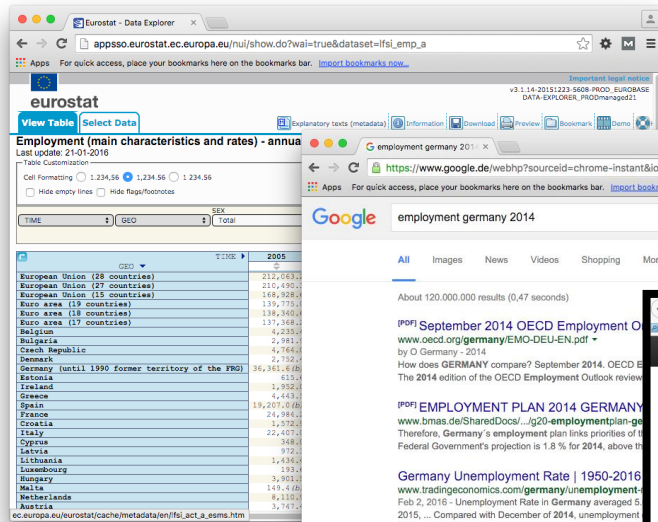
# Conclusions

## Summary

- We can automatically generate SEO landing pages for arbitrary statistics on the web using Statistical Linked Data
- Although we cannot compete with more prominent off-page optimized websites, some of our pages may eventually prove to be a useful source of specific numeric information for search engines and users

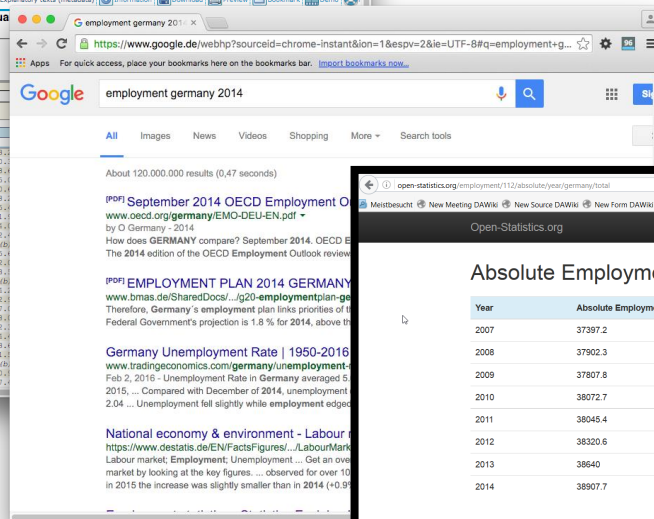
## Open work

- So far only two datasets published
  - What to do if we want to publish all 5,000 datasets of Eurostat and of other available Statistical Data sources?
- Is it possible to automatically provide additional information besides the specific numbers such as
  - definitions of indicators from DBpedia/Wikipedia
  - news articles mentioning the numbers
  - highlighting anomalies
  - showing related numbers (for comparisons)
- Is it possible to allow arbitrary combinations (e.g., derived number "GDP per capita")...
- ...and comparisons of numbers from the "Global Cube" of all available Statistical Linked Data on the web?



**eurostat**  
 Employment (main characteristics and rates) - annua  
 Last update: 21-01-2016  
 Table Customiser

TIME	SEX	Total
2005		
European Union (28 countries)		212,043.3
European Union (27 countries)		210,490.0
European Union (15 countries)		158,908.6
Euro area (19 countries)		139,775.4
Euro area (18 countries)		139,340.8
Euro area (17 countries)		137,368.1
Belgium		4,235.1
Bulgaria		2,981.1
Czech Republic		4,764.4
Denmark		2,752.4
Germany (until 1990 former territory of the FRG)		36,361.6
Estonia		415.2
Ireland		1,922.1
Greece		4,443.1
Spain		19,407.0
France		24,984.4
Croatia		3,978.8
Italy		22,407.1
Cyprus		548.1
Latvia		972.1
Lithuania		1,436.1
Luxembourg		1,591.1
Hungary		3,901.1
Malta		145.4
Netherlands		8,110.1
Austria		3,747.1



employment germany 2014

About 120,000,000 results (0,47 seconds)

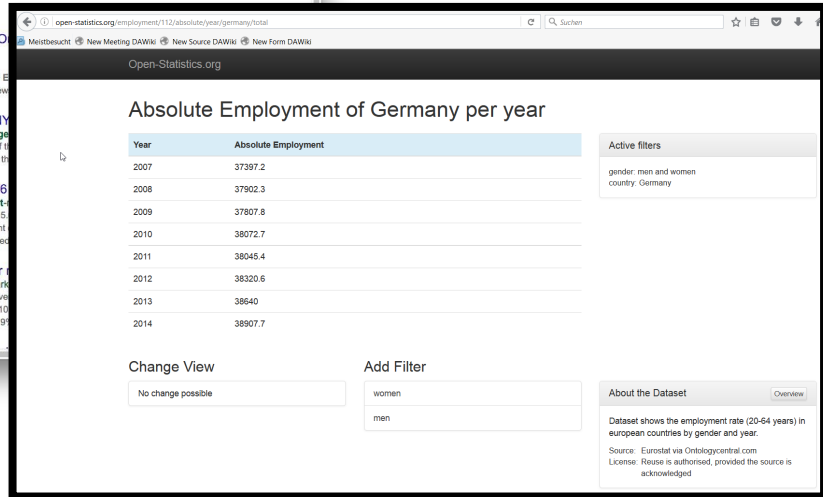
September 2014 OECD Employment Outlook  
[www.oecd.org/germany/EMO-DEU-EN.pdf](http://www.oecd.org/germany/EMO-DEU-EN.pdf)  
 by O Germany - 2014

How does GERMANY compare? September 2014. OECD Employment Outlook review  
 The 2014 edition of the OECD Employment Outlook review

EMPLOYMENT PLAN 2014 GERMANY  
[www.bmas.de/SharedDocs/.../g20-employmentplan-g...](http://www.bmas.de/SharedDocs/.../g20-employmentplan-g...)  
 Therefore, Germany's employment plan links priorities of the Federal Government's projection is 1.8 % for 2014, above the

Germany Unemployment Rate | 1950-2016  
[www.tradingeconomics.com/germany/unemployment-rate](http://www.tradingeconomics.com/germany/unemployment-rate)  
 Feb 2, 2016 - Unemployment Rate in Germany averaged 5.2015. ... Compared with December of 2014, unemployment 2.04 ... Unemployment fell slightly while employment edged

National economy & environment - Labour market  
<https://www.oestatistik.de/EN/FactsFigures/.../LabourMarket>  
 Labour market: Employment; Unemployment; ... Get an overview of the labour market by looking at the key figures ... observed for over 10 years. In 2015 the increase was slightly smaller than in 2014 (+0.9%)



Open-Statistics.org

### Absolute Employment of Germany per year

Year	Absolute Employment
2007	37397.2
2008	37902.3
2009	37807.8
2010	38072.7
2011	38045.4
2012	38320.6
2013	38640
2014	38907.7

Active filters  
 gender: men and women  
 country: Germany

Change View

Add Filter

About the Dataset  
 Dataset shows the employment rate (20-64 years) in european countries by gender and year.  
 Source: Eurostat via Ontologycentral.com  
 License: Reuse is authorised, provided the source is acknowledged

Thanks!

Kontakt:

Benedikt Kämpgen:

[kaempgen@fzi.de](mailto:kaempgen@fzi.de)

Andreas Harth:

[harth@kit.edu](mailto:harth@kit.edu)

# References

- [Hoe13] P. Hoefler, “Linked data interfaces for non-expert users,” in *The Semantic Web: Semantics and Big Data*. Springer, 2013, pp. 702–706.
- [KH14] Benedikt Kämpgen, Andreas Harth: *OLAP4LD - A Framework for Building Analysis Applications Over Governmental Statistics*. ESWC (Satellite Events) 2014: 389-394.
- [KOH12] Benedikt Kämpgen, Seán O’Riain, Andreas Harth: *Interacting with Statistical Linked Data via OLAP Operations*. ESWC (Satellite Events) 2012: 87-101.
- [Mal09] R. Malaga, “Web 2.0 Techniques for search engine optimization: Two case studies,” *Review of Business Research*, vol. 9, no. 1, pp. 132–139, 2009.
- [Mik15] P. Mika, “On Schema.org and Why It Matters for the Web,” *Internet Computing*, IEEE, vol. 19, no. 4, pp. 52–55, 2015.
- [SCC13] B.-Y. Shih, C.-Y. Chen, and Z.-S. Chen, “An empirical study of an internet marketing strategy for search engine optimization,” *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 23, no. 6, pp. 528–540, 2013.
- [SMM+12] P. E. R. Salas, M. Martin, F. M. D. Mota, S. Auer, K. Breitman, M. Casanova et al., “Publishing statistical data on the web,” in *Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on*. IEEE, 2012, pp. 285–292.
- [VE13] C. Veres and E. Elseth, “Schema.org for the Semantic Web with MaDaME.” in *I-SEMANTICS (Posters & Demos)*. Citeseer, 2013, pp. 11–15.