

1849 - Datenorientierte Systemanalyse

11/06/2014

Axel Polleres

Stundenwiederholung

1) Extend your Web interface with

- At least one form to insert, delete, or update data.
- At least one report that displays at least one diagram

Abgabe: bis 6.6.2014
(naechste Woche ist keine VO!)

2) Familiarize yourself with SPARQL: go through the examples...

Come up with a few own interesting queries on DBpedia or another public SPARQL endpoint:

- Use the SPARQL endpoint at
 - <http://live.dbpedia.org/sparql>
- Other Open RDF Data SPARQL endpoints you could use:
 - <http://open-data.europa.eu/en/linked-data> ... Linked Data from the European Open Data Portal
 - <http://worldbank.270a.info/sparql> ... Linked data from WorldBank
- Even if you don't manage to query the data you wanted: Formulate the query you intend in natural language (e.g. "*Football players born in Vienna after 1950*") and we will try to work towards the solution together next time.

3) Start thinking about the final project [...]

In der nächsten Stundenwiederholung sollten Sie zeigen können:

- Eine "draft"-Version Ihres Abschluss-Projekts, incl. Eingabemasken und Diagramm.
- Ein eigenes validierendes RDF file
- SPARQL-Abfragen mit den in der heutigen Stunde präsentierten features (FILTER, UNION, OPTIONAL, aggregates)

Overview

- Was sich leider nicht mehr ausgegangen ist:
 - Das Statistik-Paket "R" (a "teaser" example)
- Wrap-up & Outlook
- Possible Bachelor thesis topics

Data Analytics with R:

- Das Statistik-Paket "R" – Grundlagen
 - Tutorial:
<http://mitloehner.net/lehre/~mitloehn/lehre/rbasics/rbasics.html>
 - Lots of material available at <http://www.r-project.org/>

Start:

```
$ R  
> help.start()
```

- Ein Beispiel: Process RDF & Linked Data with R

Active (who played in 2013 squad) strikers, their names, dates of birth and goals, according to dbpedia:

How many goals do strikers listed on Dbpedia score on average?
How's the distribution?

```
SELECT DISTINCT ?P ?Birthdate ?Name (sum(?G) AS ?Goals) WHERE {  
?S a <http://dbpedia.org/ontology/SoccerClubSeason> ;  
  <http://dbpedia.org/property/season> 2013 ;  
  <http://dbpedia.org/property/p> ?P .  
?P <http://dbpedia.org/property/goals> ?G .
```

```
{ SELECT DISTINCT ?P (sample(?N) as ?Name) (sample (?B) as ?Birthdate ) WHERE {  
  ?P <http://dbpedia.org/property/name> ?N ;  
  <http://dbpedia.org/ontology/birthDate> ?B;  
  <http://dbpedia.org/property/position> <http://dbpedia.org/resource/  
Forward_(association_football)>  
  FILTER( datatype(?B) = <http://www.w3.org/2001/XMLSchema#date> ) }  
GROUP BY ?P }  
FILTER ( isnumeric(?G) )  
}  
GROUP BY ?P ?P ?Birthdate ?Name
```

Import this into our database!

- We know that already!
 - Enter the query in <http://live.dbpedia.org/sparql/>
 - Or get it as CSV:
 - store the query as "strickers.rq"
 - `curl "http://live.dbpedia.org/sparql" -F 'query=@strickers.rq' -H 'Accept: text/csv' -o strickers.csv`
 - `psql`
 - `CREATE TABLE strickers (Player varchar(100), Birthdate date, Name varchar(100), Goals integer);`
 - `\COPY strickers FROM 'strickers.csv' WITH DELIMITER ',' CSV HEADER`
 - `SELECT * FROM strickers;`

Daten mit R verarbeiten:

- R aufrufen
- Mit R Verbindung zur Datenbank aufbauen und mit folgender SQL-Abfrage Tore und Alter der einzelnen Spieler abfragen:
 - `SELECT Name, 2014-(EXTRACT(YEAR FROM birthdate)) as Age, goals FROM strikers;`
- Siehe:
<http://mitloehner.net/lehre/rbasics/rbasics.html>
Section „R and Database Connection“

Sample R session:

```
ssh -X balrog
```

← wichtig, wenn Sie die Grafik aus R anzeigen wollen, müssen Sie -X verwenden!

```
cd www
```

← alternativ, wenn Sie im www Verzeichnis arbeiten, können Sie sich generierte Grafiken über den Browser anzeigen lassen

```
R
```

```
library(RPostgreSQL)
```

```
drv <- dbDriver("PostgreSQL")
```

```
con <- dbConnect(drv, user="apollere", password="apollere", dbname="apollere", host="localhost")
```

```
strikers <- dbGetQuery(con, "SELECT Name, 2014-(EXTRACT(YEAR FROM birthdate)) as Age, goals FROM strikers;")
```

```
attach(strikers)
```

```
goals
```

```
age
```

```
jpeg('goals-by-age.jpg')
```

← Wenn Sie diese Zeile weglassen, wird die Grafik direkt am Bildschirm ausgegeben

```
plot(x=age, y=goals)
```

```
dev.off()
```

← Plot-Grafik wird in Datei geschrieben

```
max(goals)
```

```
jpeg('hist-goals.jpg')
```

```
hist(goals)
```

```
dev.off()
```

```
sd(goals)
```

```
sd(age)
```

```
mean(age)
```

```
jpeg('hist-age.jpg')
```

```
hist(age)
```

```
dev.off()
```

```
t.test(age, mu = 28)
```


What did we just do with that script?

- Do the goals per striker follow a normal distribution? No...
- Does the age of strikers follow a normal distribution? More or less...
- We checked whether assuming a mean(age) of 28 for strikers justified by the data? (t-test) ...
Answer was: No

Connecting the dots... ... Might be useful for your final project (optional!!!):

- Generating Reports using R
 - Use
 - `sink(outputfile)` e.g. `sink("report.html")`
 - `cat()` `Cat("<h2>Report</h2>")`, e.g.
 - See examples at <http://mitloehner.net/lehre/datsys/reports.html>
 - Generate reports in HTML or RTF
 - For more convenient/sophisticated RTF file generation from R there's a package:
 - `install.packages(rtf)`
 - See <http://cran.r-project.org/web/packages/rtf/vignettes/rtf.pdf>
- Calling R from a Web interface
<http://cran.r-project.org/doc/FAQ/R-FAQ.html#R-Web-Interfaces>

Generate an HTML File...

- For some formatting stuff we'll use the Hmisc package, which we need to install/load:

```
> install.packages("Hmisc")  
> library("Hmisc")
```

Note: We have R version 2.12.1 (2010-12-16) installed, so for the the rtf package, you probably need to install some older versions of packages (compatible with that version of R) from source:

```
$ wget http://cran.r-project.org/src/contrib/Archive/Hmisc/Hmisc_3.9-0.tar.gz  
$ R  
> install.packages("Hmisc_3.9-0.tar.gz", repos = NULL, type="source")
```

For an example: Get the file

https://ai.wu.ac.at/~polleres/teaching/DOSA_2014/20140611/GenerateHTML.r

- Download this file
- Call the following from your commandline:
 - `R --no-save --silent < GenerateHTML.r`
- Generates a file `Output.html` that contains a report in HTML.

Generate an RFT File...

Installing the rtf package

Note: We have R version 2.12.1 (2010-12-16) installed, so for the the rtf package, you probably need to install some older versions of packages (compatible with that version of R):

```
$ wget http://cran.fhcrc.org/src/contrib/Archive/R.methodsS3/R.methodsS3_1.2.1.tar.gz
$ wget http://cran.fhcrc.org/src/contrib/Archive/R.oo/R.oo_1.7.5.tar.gz
$ wget http://cran.fhcrc.org/src/contrib/Archive/rtf/rtf_0.4-3.tar.gz
$ R
> install.packages("R.methodsS3_1.2.1.tar.gz",repos = NULL, type="source")
> install.packages("R.oo_1.7.5.tar.gz",repos = NULL, type="source")
> install.packages("rtf_0.4-3.tar.gz",repos = NULL, type="source")
```

For an example: Get the file

https://ai.wu.ac.at/~polleres/teaching/DOSA_2014/20140611/GenerateRTF.r

- Download this file
- Call the following from your commandline:

```
R --no-save --silent < GenerateRTF.r
```

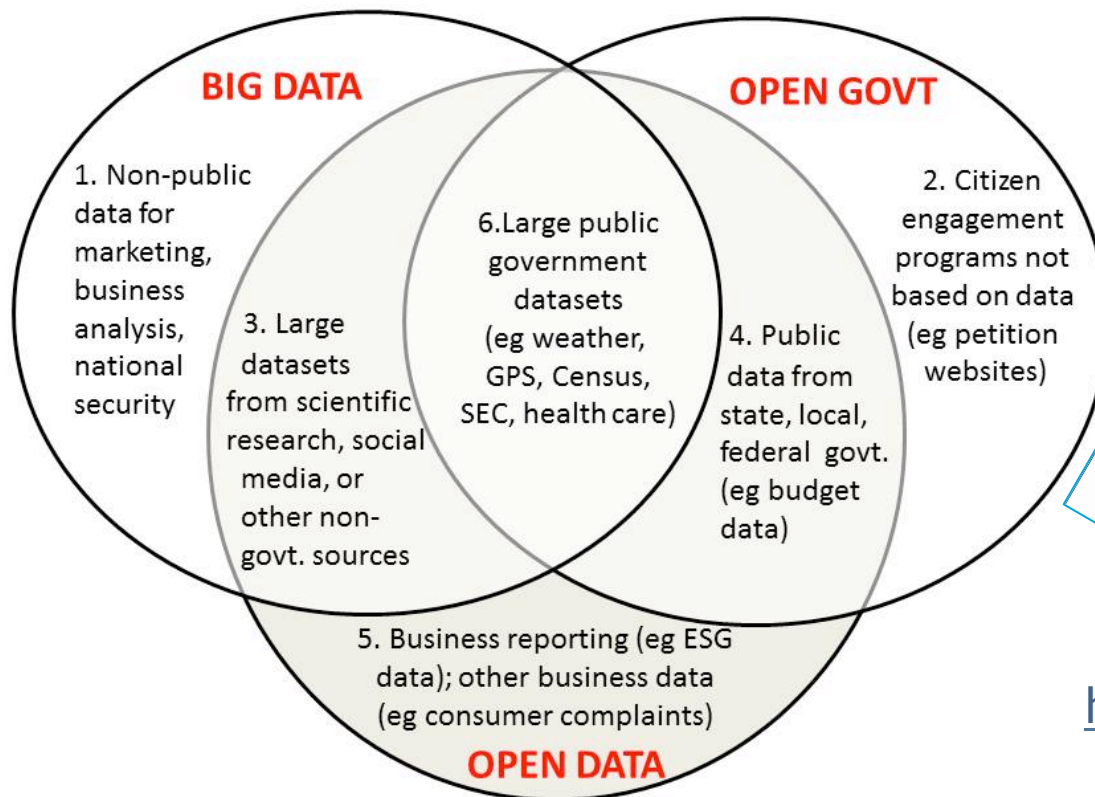
Generates a file Output.html that contains a report in Rich Text Format (RTF), can be opened & edited by most common word processors, e.g. Word.

Wrap-up: What did we learn

- Creating a Relational Database
- Querying a Relational Database
- Importing and integrating data from external sources (CSV, JSON, RDF)
- Generating Reports and run analytics
- Creating a Web interface for your database

Why is that all important?

- Big Data, Data Analytics, Data Science, Open Data, & Business Intelligence are "hot topics"



<http://www.bigdatavalue.eu/>

<http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/>

Why is that all important?

- EU & Austria are pushing Open Data!

THE WORLD BANK
Open Data

wien.at **Open Government Data**
Offene Daten für Wien

UNdata

Open Data Berlin

london.gov.uk
london datastore

Opening up Europe's public data

DBpedia

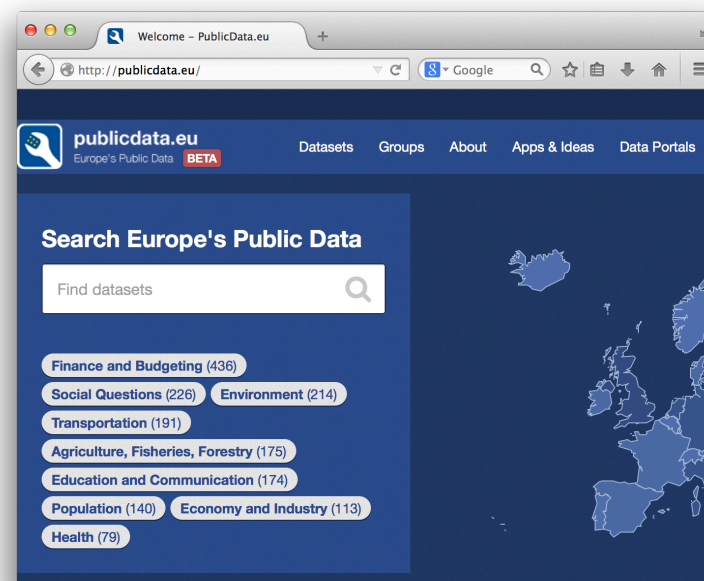
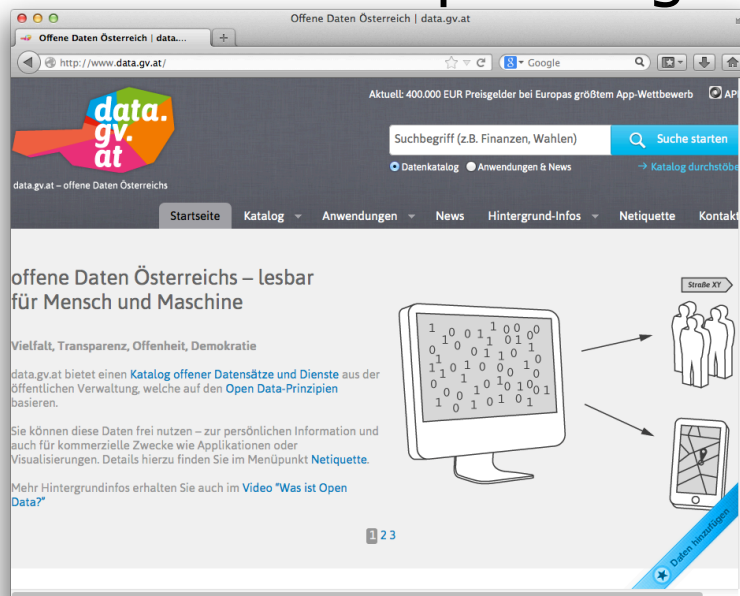
OpenStreetMap

European Commission
Open Data Portal Beta

DIRECTIVE 2007/2/EC
INSPIRE

Why is that all important?

- More and more Open Data available: increasingly in standard formats like RDF! EU & Austria are pushing Open Data!



The Austrian Open Government Data portal just won the UN Public Service Award 2014!

From Data comes Intelligence & Knowledge!



IBM Watson

24 October 2014: Invited Talk
by Chris Welty from IBM Watson
Research @ **WU !!!!**

http://en.wikipedia.org/wiki/Chris_Welty



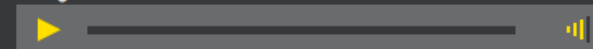
Google_{UK} Knowledge Graph

<http://fm4.orf.at/stories/1740490/>

Computer passes Turing Test

A computer has convinced members of a panel of expert judges that it is human, in what the creators are saying is the first example of its kind of a computer passing the so called Turing Test. Thomas Eiter of Vienna's University of Technology gives his analysis.

Turing test



Another example: Google Knowledge Graph

[News for venice italy](#)

[Tornado tears through parts of Venice, Italy \(VIDEOS\)](#)

[Washington Post \(blog\)](#) - 16 hours ago

A rare tornado (or waterspout, when over water) swept over several islands (Lido, Sant'Elena and Sant'Erasmo) off **Venice's** lagoon earlier ...

[Italy putting brakes on excitement](#)

[London Free Press](#) - 14 hours ago

[Italy could be hit by Spanish contagion](#)

[Economic Times](#) - 1 day ago

[Venice - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Venice

Venice (**Italian**: Venezia [veˈnɛttsja] (listen), Venetian: Venexia [veˈnesja] is a city in northeast Italy sited on a group of 118 small islands separated by canals ...

↳ [History of the Republic of Venice - Venice, Los Angeles - Grand Canal](#)

[Venice Vacations, Tourism and Venice, Italy Travel Reviews ...](#)

www.tripadvisor.com/Tourism-g187870-Venice_Veneto-Vacations.h...

Venice Vacations: With 130000 reviews of **Venice, Italy** travel resources, TripAdvisor is the source for Venice information.

[ItalyGuides.it: Virtual tour of Venice, Italy - travel information and city ...](#)

www.italyguides.it/us/venice_italy/venice_travel.htm

Venice tourism and travel information: transport, attractions, maps, travel advice, pictures, audio guides, airport information, activities, hotels and more in **Venice**, ...

[Official website of the Municipality of Venice - Comune di Venezia](#)

www.comune.venezia.it/flex/cm/pages/ServeBLOB.php/L/EN/.../1

Official website of the Municipality of **Venice, Italy**. News, information and tools available to citizens and visitors.

Venice



Venice is a city in northeast Italy sited on a group of 118 small islands separated by canals and linked by bridges. It is located in the marshy Venetian Lagoon which stretches along the shoreline between the mouths of the Po and the Piave Rivers. [Wikipedia](#)

Area: 159 sq miles (412 km²)

Weather: 72° F, Wind E at 5 mph, 50% Humidity

Local time: 12:13pm Wednesday (CEST)

Points of interest



Grand Canal of Venice



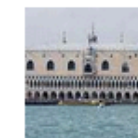
Piazza San Marco



Saint Mark's Basilica



Rialto Bridge



Doge's Palace

[Report a problem](#)

Why is that all important?

- As you might have realized yourself:
 - Mastering Data requires skills & further research!

Weiterführende Themen in der Forschung an unserem Institut:

- Possible Bachelor thesis topics:
 - WU "Open Data" Initiative - What insights can you gain from public data about your university?
 - Integrating Open Data from different sources and domains (Bachelor & MSc) e.g.:
 - "Sustainability" and "Quality of Life" related data from different Open Data Sources and presenting it in a Web interface.
 - Integrating & Analysing Music Data from Online Sources
 - Analysing Data Quality in Open Data Catalogs
 - Data Analysis for Optimizing Business Processes! New Project SHAPE

Siehe: <http://www.wu.ac.at/infobiz/team/pollerer>

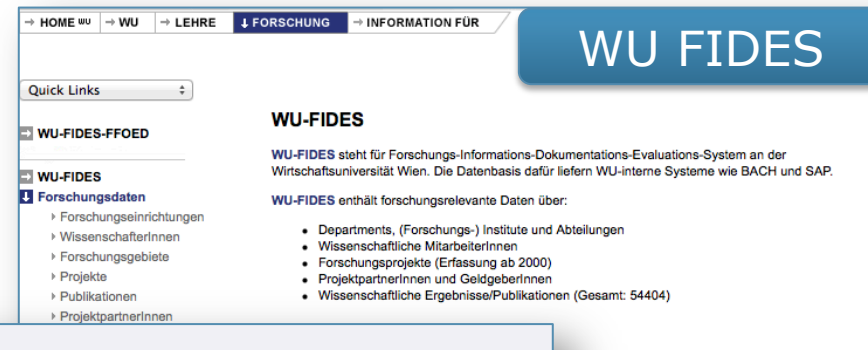
Topic1: Open Data @ WU

- Integrate public Data from WU and make it available in Standard Open Data formats:



WU Homepage

The screenshot shows the main navigation menu with categories like 'INFORMATION ÜBER', 'INFORMATION FÜR', and 'WU-NEWSLETTER'. A 'Veranstaltungskalender' (Event Calendar) for April 2014 is visible, along with a search bar and a 'Suche nach Thema sortiert' (Search by topic sorted) section with various filters.

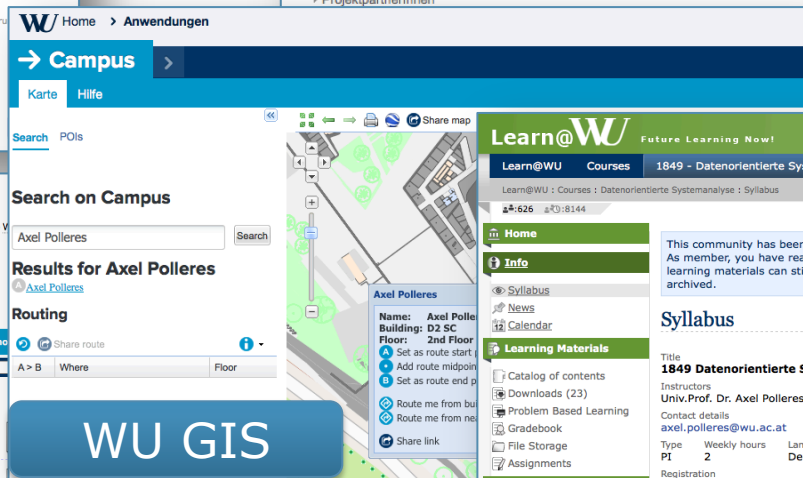


WU FIDES

WU-FIDES steht für Forschungs-Informations-Dokumentations-Evaluations-System an der Wirtschaftsuniversität Wien. Die Datenbasis dafür liefern WU-interne Systeme wie BACH und SAP.

WU-FIDES enthält forschungsrelevante Daten über:

- Departments, (Forschungs-) Institute und Abteilungen
- Wissenschaftliche MitarbeiterInnen
- Forschungsprojekte (Erfassung ab 2000)
- ProjektpartnerInnen und GeldgeberInnen
- Wissenschaftliche Ergebnisse/Publikationen (Gesamt: 54404)



WU GIS

The screenshot shows the 'Campus' map application with a search bar for 'Axel Polleres' and a 'Routing' section. A map of the WU campus is displayed with a search result for 'Axel Polleres' showing details like 'Name: Axel Polle', 'Building: D2 SC', and 'Floor: 2nd Floor'.



ePubWU Institutional Repository

A service provided by the WU Library and the WU Vienna University of Economics and Business.

Latest 5 Additions to ePubWU

1. Paetzel, Fabian and Sausgruber, Rupert and Traub, Stefan (2014) [Social Preferences and Voting on Reform: An Experimental Study](#), Department of Economics Working Paper Series, 172. WU Vienna University of Economics and Business, Vienna.
2. Antonakakis, Nikolaos and Collins, Alan (2014) [The Impact of Fiscal Austerity on Suicide: On the Empirics of a Modern Greek Tragedy](#), Social Science & Medicine. ISSN 0277-9536
3. Antonakakis, Nikolaos and Tondi, Gabriele (2014) [Does Integration and Economic Policy Coordination Promote Business Cycle Synchronization in the EU? Empirics](#), ISSN 0340-8744 Item availability restricted.
4. Novotny, Alexander and Spiekermann, Sarah (2014) [Oblivion on the Web: an Inquiry of User Needs and Technologies for European Conference on Information Systems](#), June 5-10, Tallinn, Estonia



Learn@WU Future Learning Now!

Learn@WU Courses 1849 - Datenorientierte Systemanalyse

This community has been archived. As member, you have read-only rights in this community. Please note that the contents and learning materials can still be changed by administrators, even after the community has been archived.

Syllabus

1849 Datenorientierte Systemanalyse

Instructors: Univ.Prof. Dr. Axel Polleres

Contact details: axel.polleres@wu.ac.at

Type: Weekly hours Language of instruction: PI 2 Deutsch

Registration: 09/12/13 to 12/02/13

Registration via LPIS

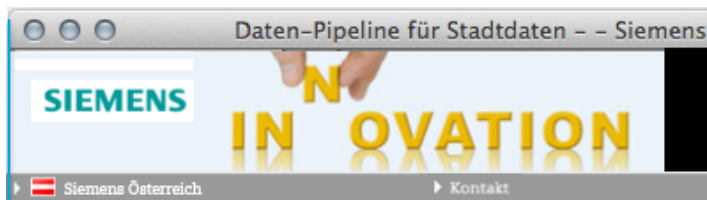
Notes to the course

Subject(s): Bachelor Programs

Specialisation in Business Administration

WU learn

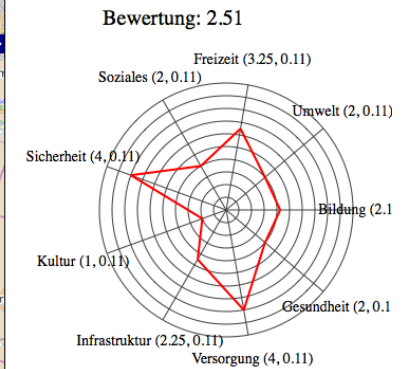
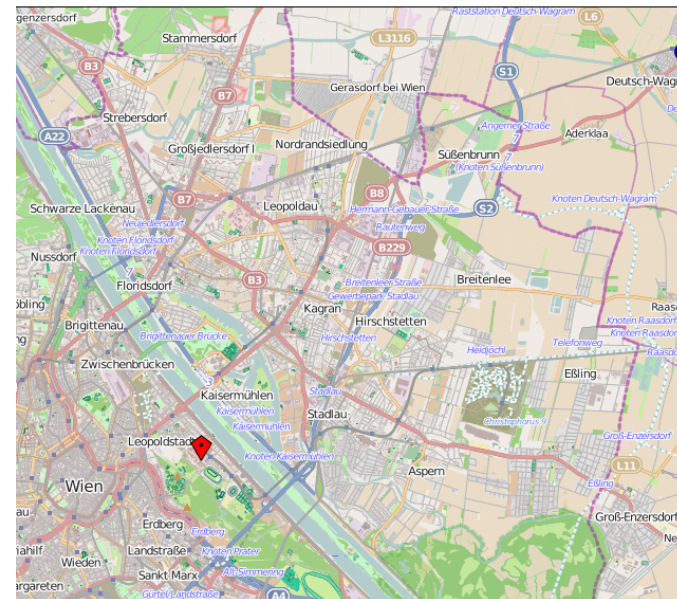
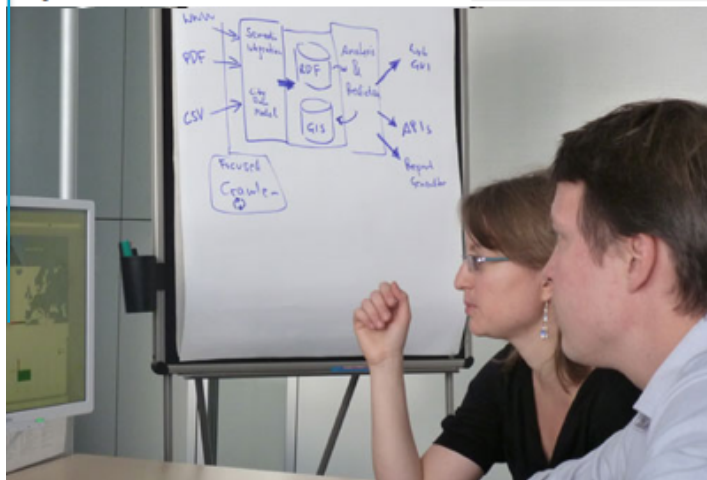
Topic2: Integrate Open City Data for Sustainability Assessment:



Nachhaltigere Städte durch Offene Daten

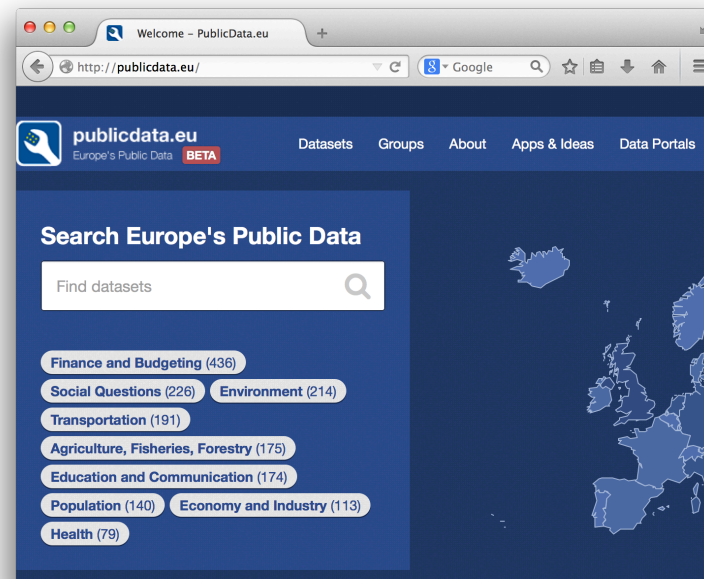
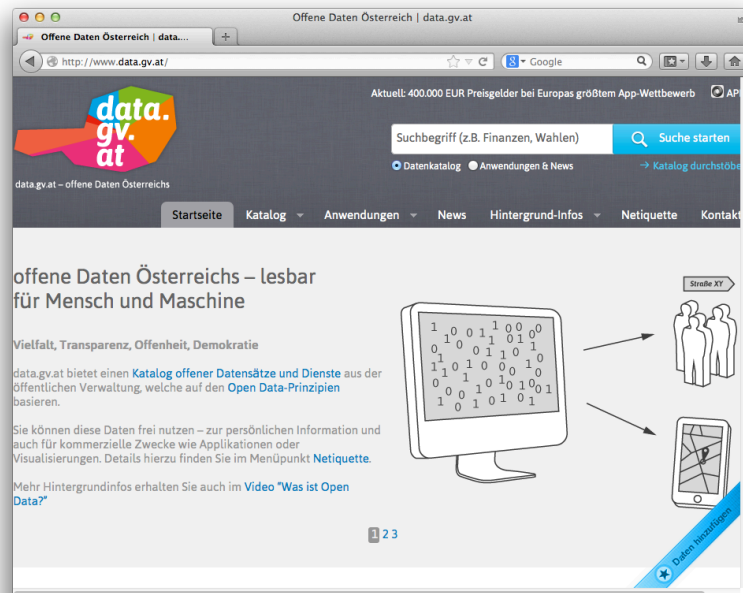
Siemens baut eine Daten-Pipeline für Stadtdaten. Welche Faktoren bestimmen die Nachhaltigkeit von Städten? Wie verändern sich diese im Laufe der Zeit? Will man Herausforderungen wie Klimawandel, demographischen Veränderungen oder Urbanisierung gewachsen sein, braucht man Antworten auf diese Fragen.

Ähnlich einer Web-Suchmaschine Pipeline öffentliche Stadtdaten vor Wikipedia und Webportalen. Ca. 2 mehr als 300 Städten sind derzeit laufend aktualisiert und erweitert.



Possible collaboration with Prof. Gunther Maier, Institute for the Environment and Regional Development

Topic 3: Open Data Quality



- **Analyse & Quantify Data Problems in Open Data:**
 - **Use of standards?** Different Data formats/encodings, etc.
 - **Incomplete** Data: how much data is missing for particular domains
 - **Incomparable** Data: **Heterogeneity** across Open Government Data efforts
 - **Different Licenses** of Open Data: e.g. CC-BY, country specific licences, etc.

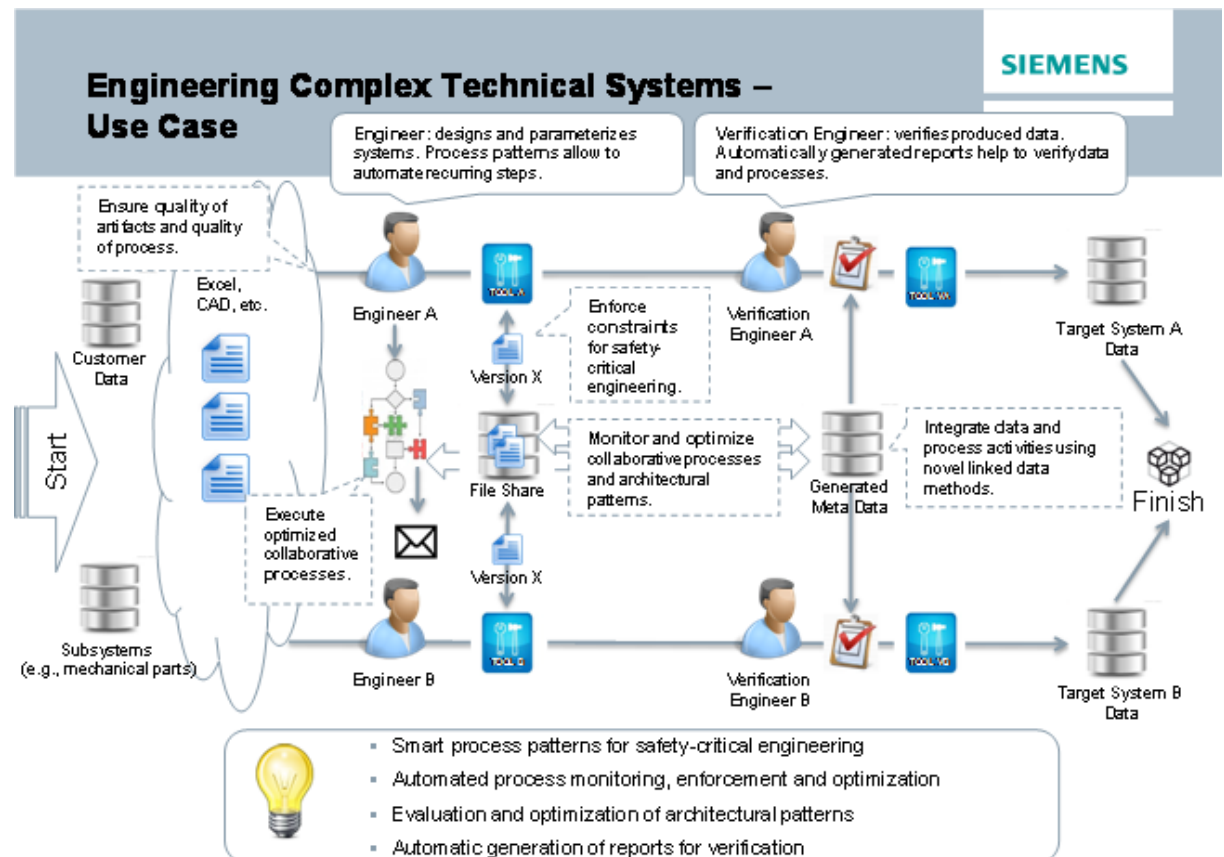
Topic 4: Data Analytics for Process Monitoring & Optimization (rather MSc)

- FFG funded research project, start October 2014:

■ **SHAPE**

(*Safety-critical Human- & dAta-centric Process management in Engineering projects*)

Together with Prof. Mending & Siemens



- 16.6.: Final Project presentations ← Q&A now!

- 18.6.: The logo for SAP HANA, featuring the word "SAP" in white on a blue background and "HANA" in yellow on a white background.

- Last "assignment" (voluntary):
 - 3 things you liked about the lecture
 - 3 things you didn't like/where you see possible improvement (I am happy to take harsh criticism, but please formulate it constructively ;-))
- will keep it Open until 19.6.

- Please don't hesitate to give feedback per email or through learn@WU !!!