WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Open City Data Pipeline

Axel Polleres

formerly: Siemens AG Österreich, now WU Wien, Austria

City Data – Important for Infrastructure Providers & for City Decision Makers

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- City Assessment and Sustainability reports
- Tailored offerings by Infrastructure Providers


Megacities


London

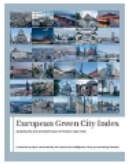

Munich


Dublin


Vienna


Trondheim

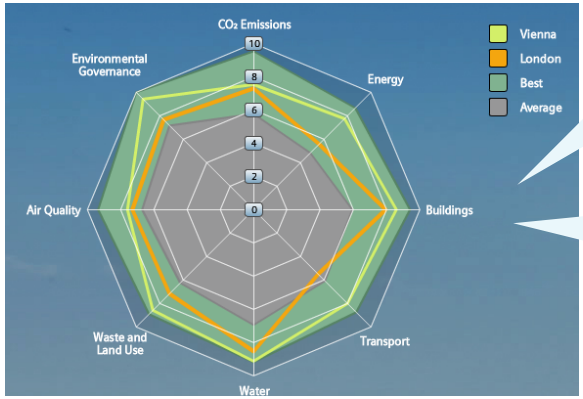

US Mayors


European Green City Index

… however, these are often **outdated** before even published!

→Needs **up-to-date City Data** and **calculates City KPIs** in a way that allows to display the current state and run scenarios of different product applications.

e.g. towards a "Dynamic" Green City Index:



Goal (short term):
▪Leverage Open Data for calculating a city' performance from public sources on the Web **automatically**

Goal (long term):
▪Define and Refine KPI models to assess specific impact of infrastructural investments and gather/check input **automatically**

EQUIS ACCREDITED

Current State of Data and Benchmarking System

- **Collecting Data for City Assessment and Benchmarking:**
- Data collection for various studies within Siemens is a manual, time-intensive process, using statistical data as well as questionnaires to city stakeholders.
- **Much of this data is available online:**
- City Open Data Initiatives publish more and more data with frequent updates

UNdata A world of information | THE WORLD BANK Open Data | DBpedia | wien.at | NYC OpenData | london.gov.uk

- City data format standards and regulations are developing:

INSPIRE - Infrastructure for Spatial Information in Europe

e.g. EU INSPIRE directive, or

European Commission eurostat Your key to European statistics

eurostat's UrbanAudit Collection of city indicators

- **Benchmarking Systems and KPIs** The Green City Index | MOST LIVEABLE CITIES
  - City Indexes benchmark cities based on top down data (example: tax income from petrol, tax/L→ Car CO_2 emissions). Bottom up approaches only if no top down data available, for approximation (example: No. cars, average distance driven per year, Emission factor → Car CO_2 emissions)
  - Approaches allowing scenario comparison and calculations of system impacts only on a city specific case study base.

Leveraging Open Data:
Openly available urban indicators frameworks (already available in Linked Data!)

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Data example: Urban Audit
- (ca. 300 indicators for European
- 330 cities, maintained by Eurostat)

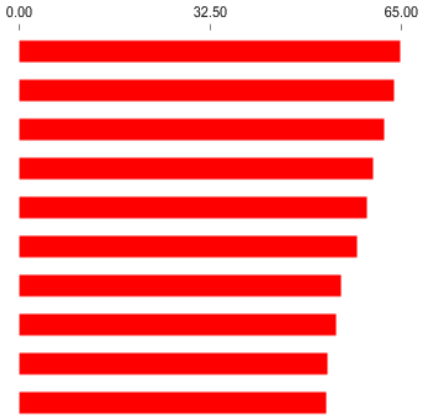
European Commission
eurostat Your key to European statistics

Cost of a monthly ticket for public transport (for 5-10 km)

Average : 46.59 High : 65 Low : 32.5

You are on page 1 of 4 (37 records)

Rank	City	Score
1	Weimar (DE)	65.00
2	Berlin (DE)	64.00
3	Frankfurt am Main (DE)	62.30
4	Augsburg (DE)	60.30
5	Wiesbaden (DE)	59.40
6	Mainz (DE)	57.60
7	Hannover (DE)	55.00
8	Bochum (DE)	54.00
9	Düsseldorf (DE)	52.60
10	Koblenz (DE)	52.30

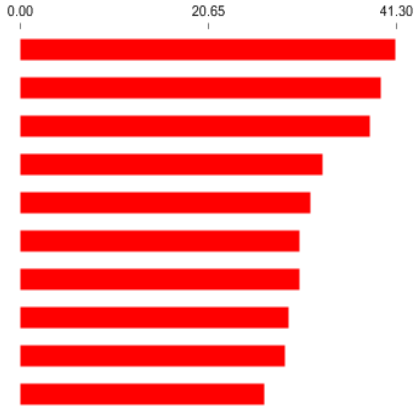


Proportion of journeys to work by public transport (rail, metro, bus, tram)

Average : 20.25 High : 41.3 Low : 4.9

You are on page 1 of 4 (40 records)

Rank	City	Score
1	München (DE)	41.30
2	Berlin (DE)	39.70
3	Frankfurt am Main (DE)	38.50
4	Hamburg (DE)	33.30
5	Stuttgart (DE)	32.00
6	Düsseldorf (DE)	30.80
7	Halle an der Saale (DE)	30.70
8	Nürnberg (DE)	29.60
9	Hannover (DE)	29.20
10	Köln - Cologne (DE)	26.90

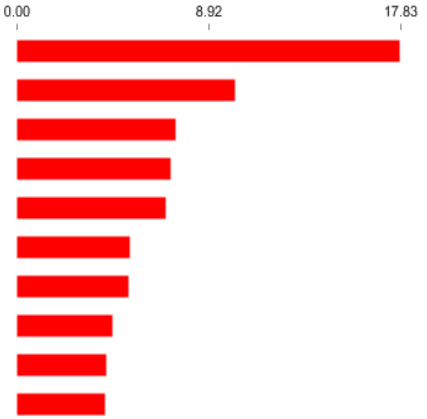


Number of stops of public transport per km2

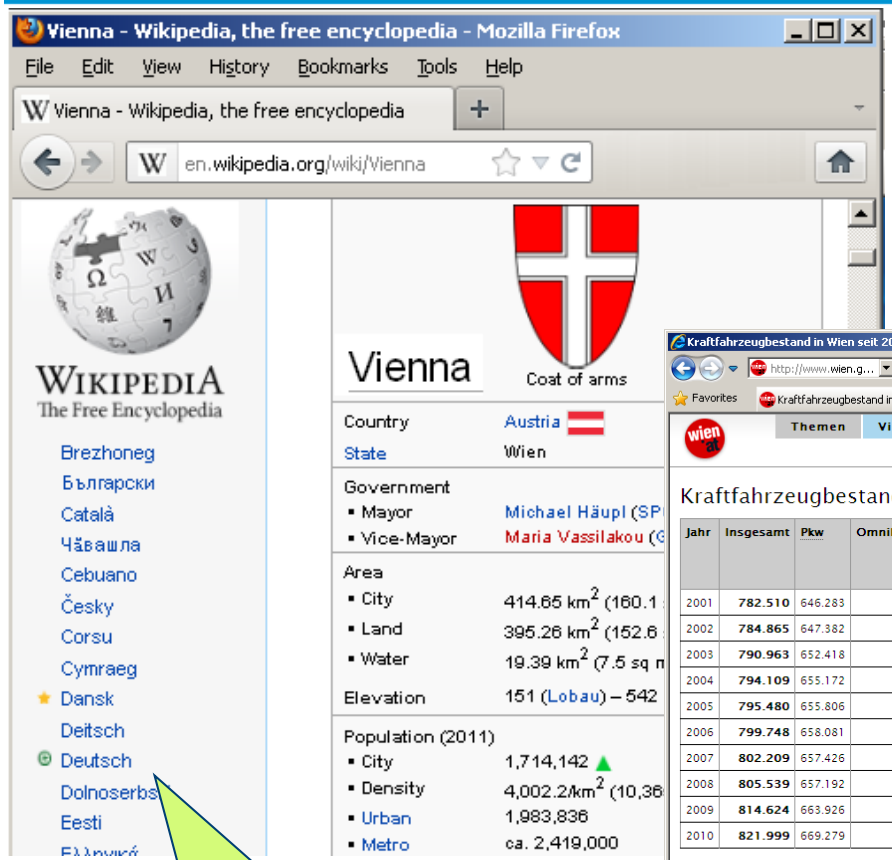
Average : 3.88 High : 17.83 Low : 1.4

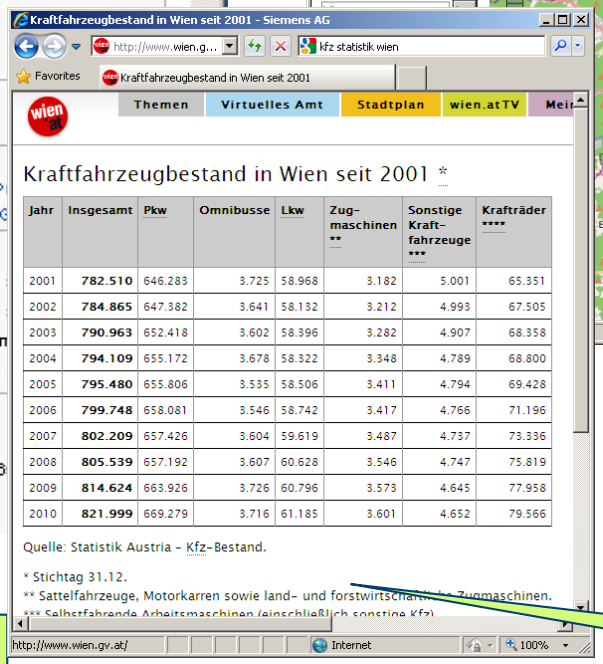
You are on page 1 of 4 (38 records)

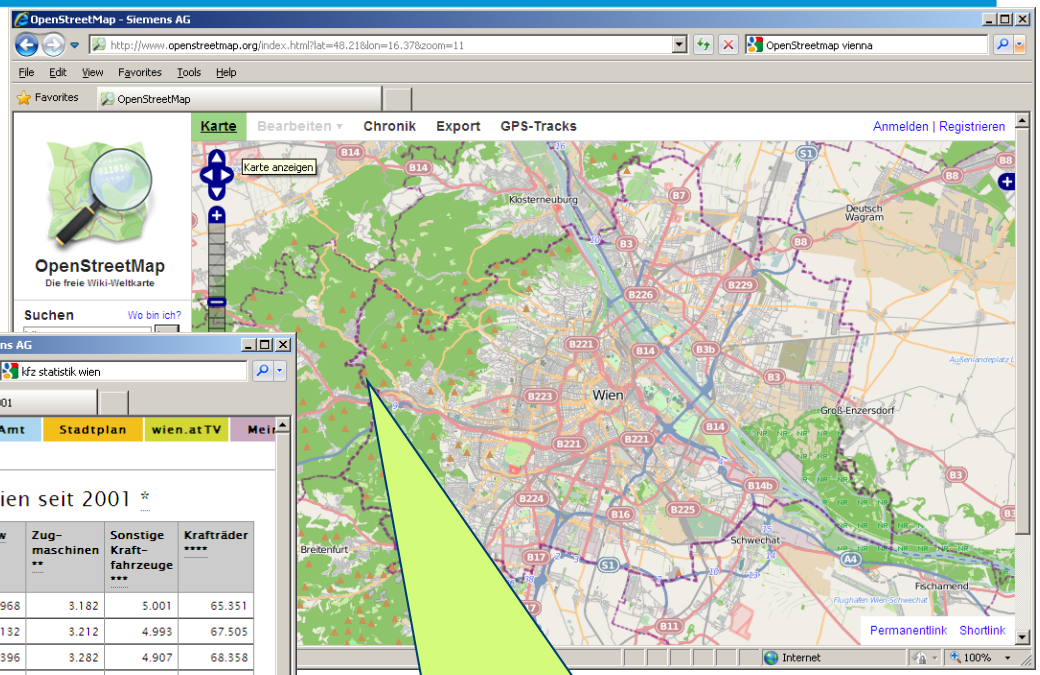
Rank	City	Score
1	Stuttgart (DE)	17.83
2	Wuppertal (DE)	10.17
3	Berlin (DE)	7.38
4	Düsseldorf (DE)	7.13
5	Kiel (DE)	6.93
6	Halle an der Saale (DE)	5.25
7	Augsburg (DE)	5.22
8	Dresden (DE)	4.47
9	Mainz (DE)	4.18
10	Göttingen (DE)	4.09



Leveraging Open Data: Other Open Data Sources





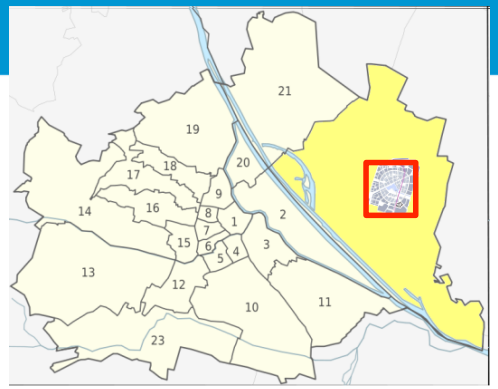


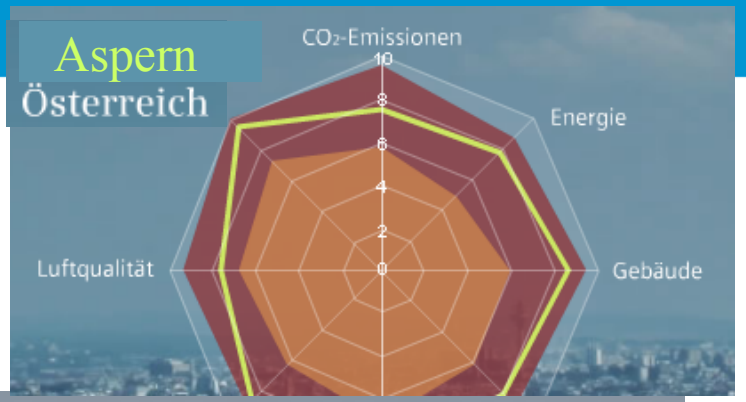
Structured information on most cities in the world (base indicators: population, economy, climate, ...)

Free GIS data for most cities in the world (base information: area, land-use, administrative districts, ...)

Open Government Data

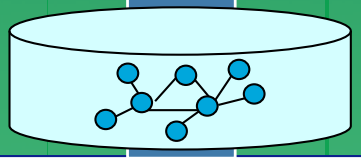
City Data Pipeline: Overview





Dynamic Calculation of KPIs at variable Granularity (City, District, Neighbourhood, Building)

3. Analysis/Statistical Correlation/Aggregation:
Statistical Methods, Semantic Technologies, Constraints



Extensible CityData Model

2. Semantic Integration: Unified Data Model, Data Consolidation

1. Periodic Data Gathering of registered sources ("Focused Crawler"):
Various Formats (CSV, HTML, XML …) & Granularity (monthly, annual, daily)

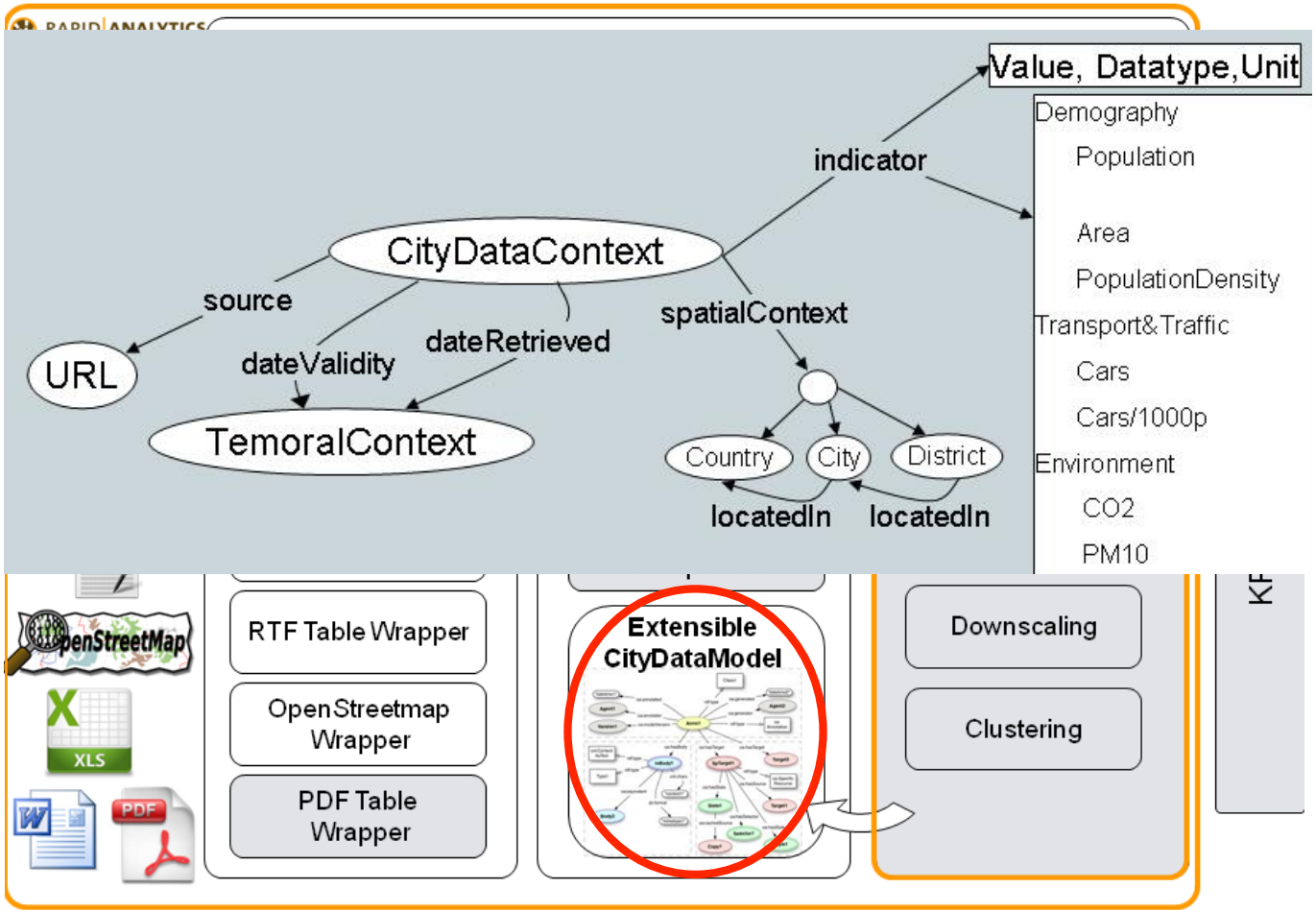
Cities:
Berlin, Vienna, London, …

+

Open Data:
DBpedia Open Data
UNdata

City Data Pipeline: Architecture

We have developed a data pipeline to
(1) (semi-)automatically collect and integrate various Open Data Sources in different formats
(2) compose and calculate complex city KPIs from the collected data



Semantic Integration Technologies** build the core of **our Data-Pipeline (RDF Data Store, Ontology-based)

Deploying Semantic Web Standards:

City Data Pipeline: Current Data - Summary

- Ca. **475** different indicators
  - *Categories: Demography, Geography, Social Aspects, Economy, Environment, etc.*

- from **32** sources (html, CSV, RDF, …)
  - *Wikipedia, urbanaudit.org, Statistics from City homepages, country Statistics, iea.org*

- Covering **350+** cities in **28** European countries
  - District Data for selected cities (Vienna, Berlin)
  - Mostly snapshots, Partially covering timelines
  - On average ca. **285** facts per city.

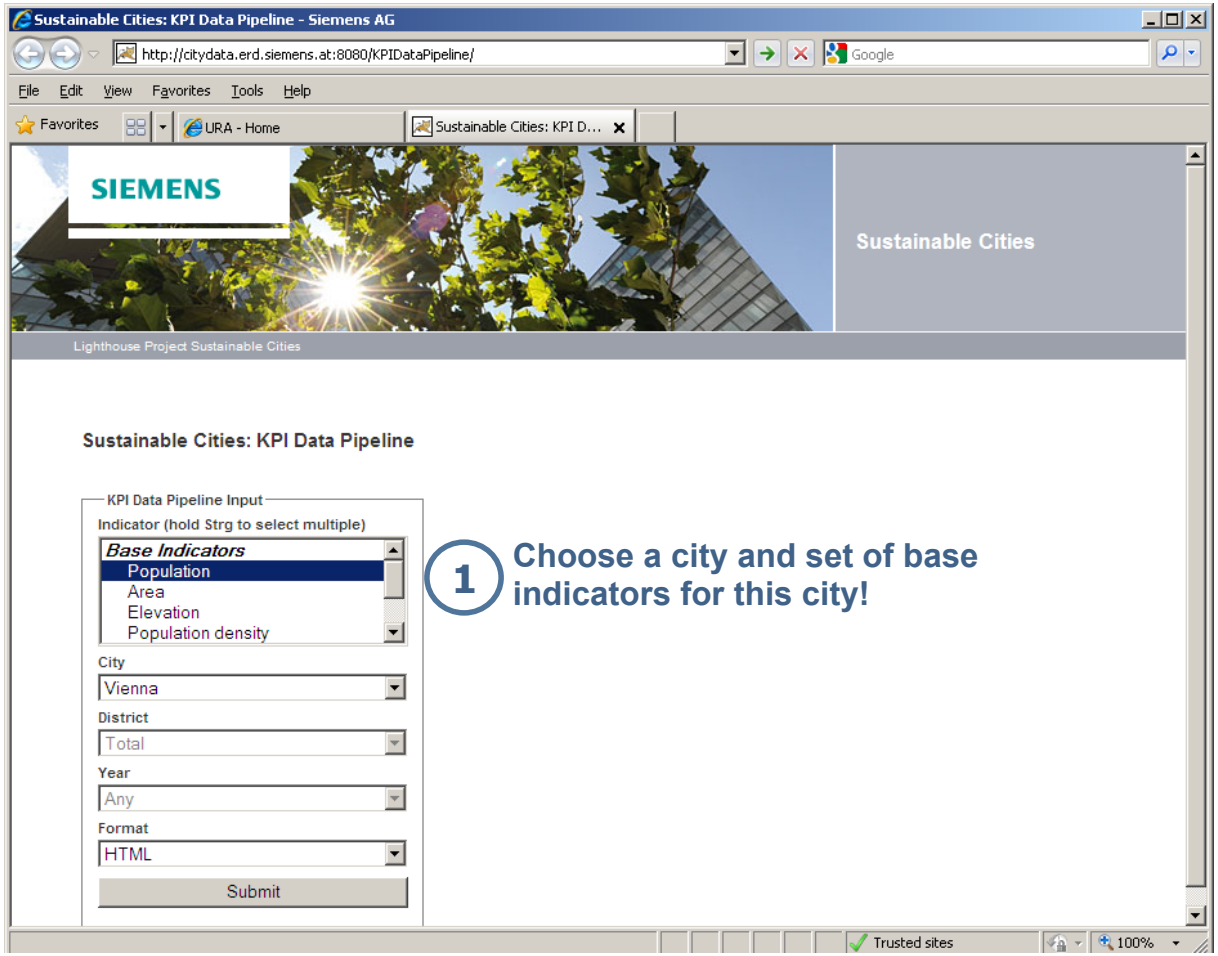
- Examples of sources:
  - UrbanAudit (from http://eurostat.linked-statistics.org)
  - http://geonames.org (population, georeference, elevation)

Further data sources on target to integrate include:

- http://data.worldbank.org/ WorldBank (mostly data at country level)
- www.eea.europa.eu/ European Environmental Agency (weather/climate data)

City Data Pipeline: Web Interface

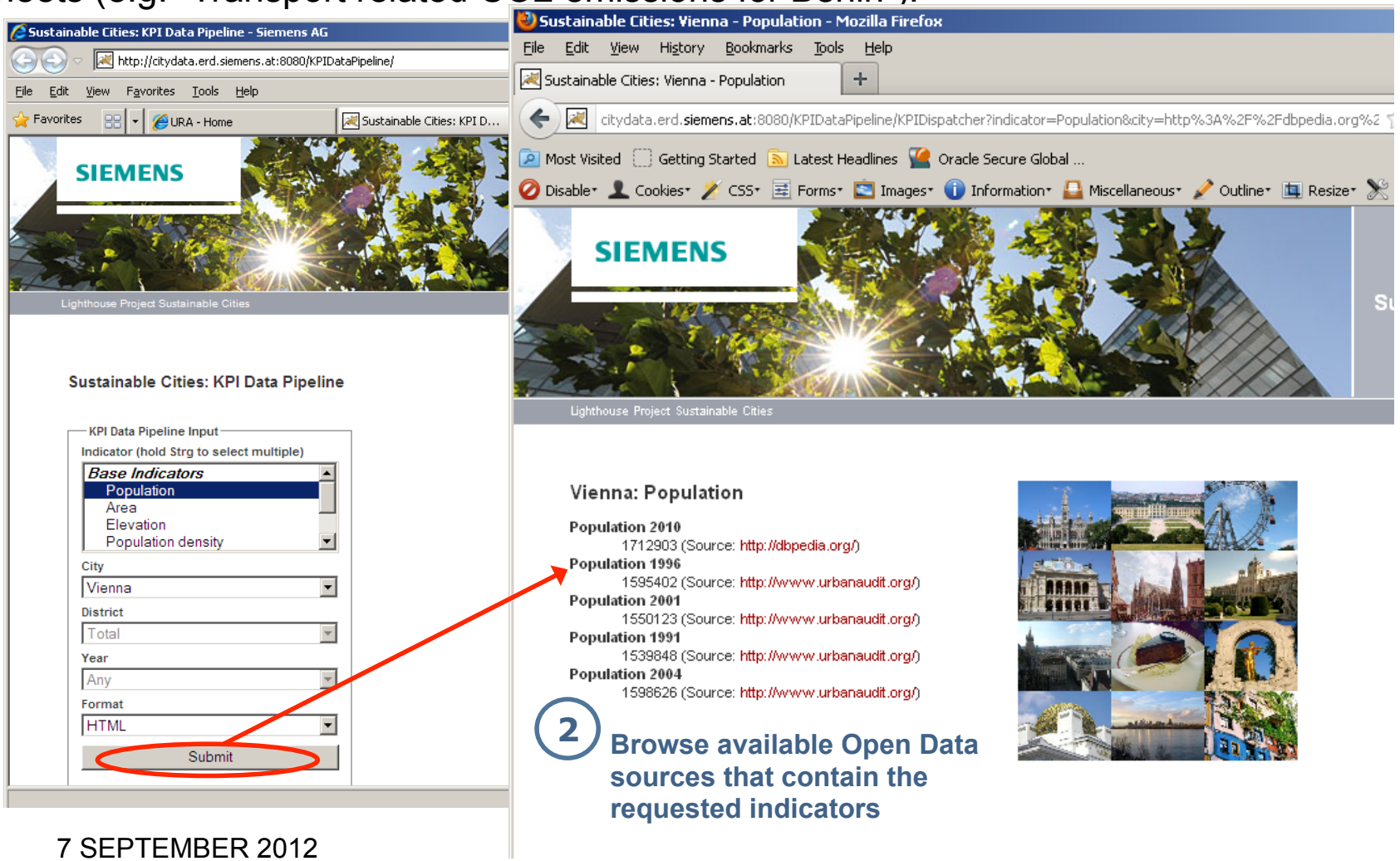
Our Web interface allows to browse data and download complex composed KPIs as Excel sheets (e.g. “Transport related CO2 emissions for Berlin”):



City Data Pipeline: Web Interface

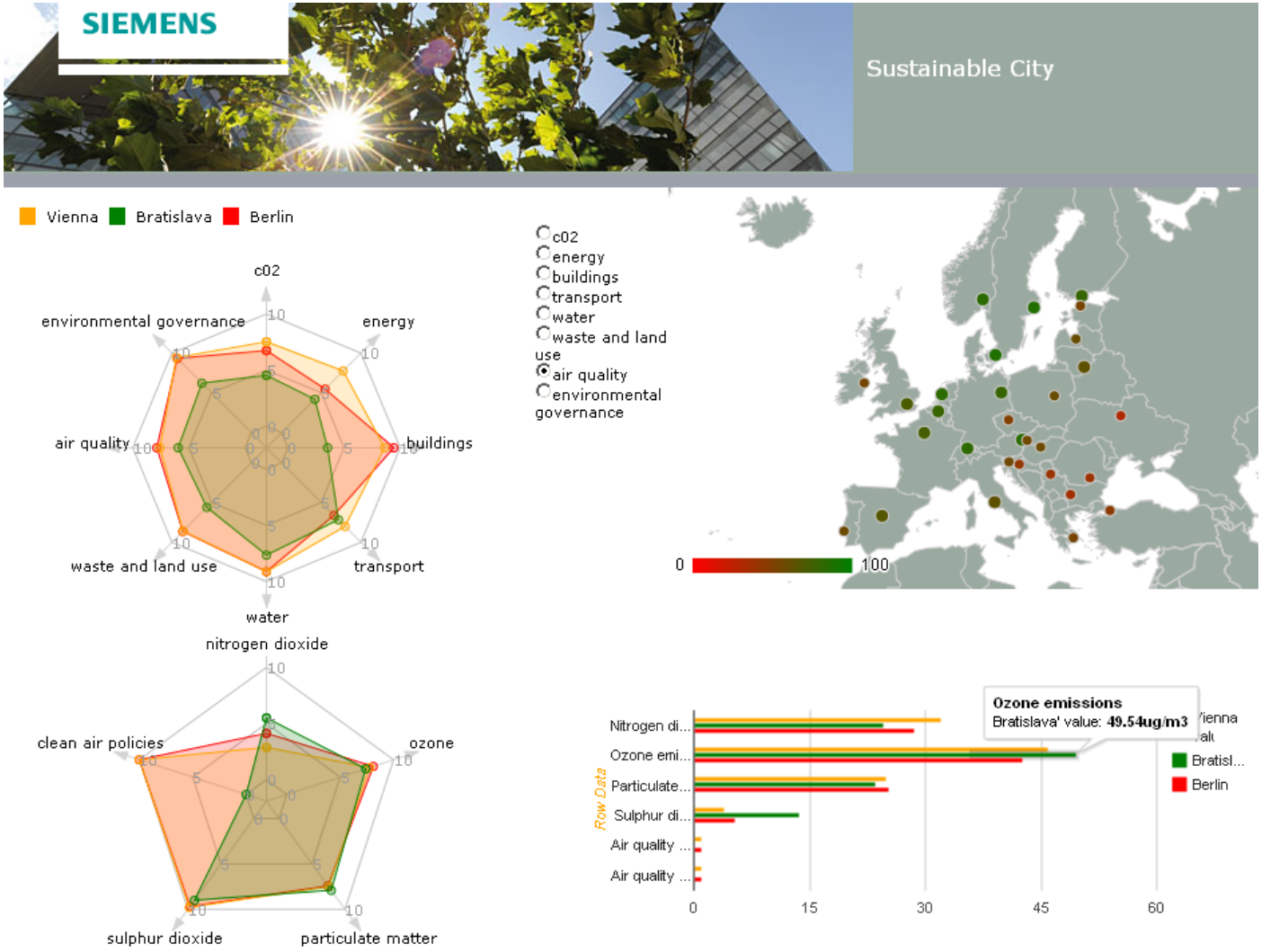
WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Our Web interface allows to browse data and download complex composed KPIs as Excel sheets (e.g. “Transport related CO2 emissions for Berlin”):



City Data Pipeline: Web Interface

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS



3

Also available: Graphical user interface to visually compare base indicators for different cities.

EFMD EQUIS ACCREDITED

Applications & Challenges

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

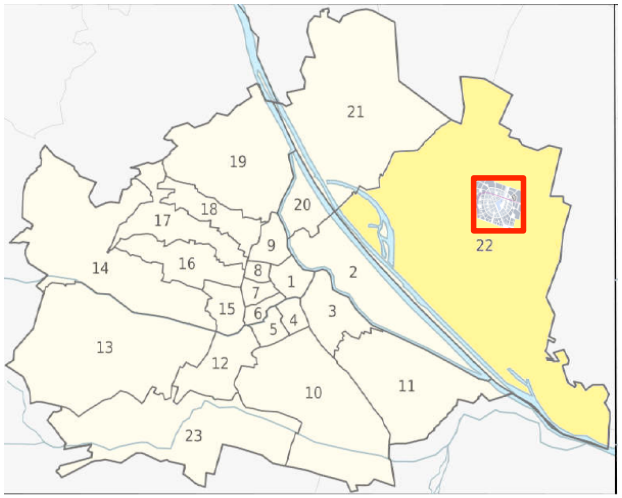
1. **Application: trend prediction - Example “Seestadt Aspern”**
2. Integration & enrichment of “Green City Index” Data
3. Challenges with Open Data Experienced

EFMD EQUIS ACCREDITED

Data Prediction/Quality: Statistical methods

- Showcase: Estimate how Seestadt Aspern will be developing in comparison to the rest of Vienna

  - Semantic integration of open data in our data pipeline
  - Prediction of indicators for Aspern and Vienna
  - Graphical representation of the results

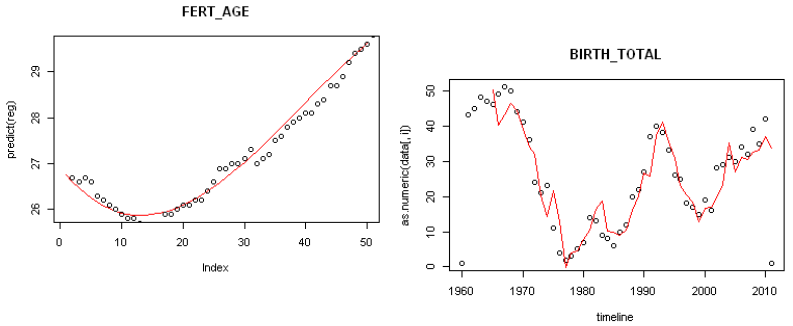


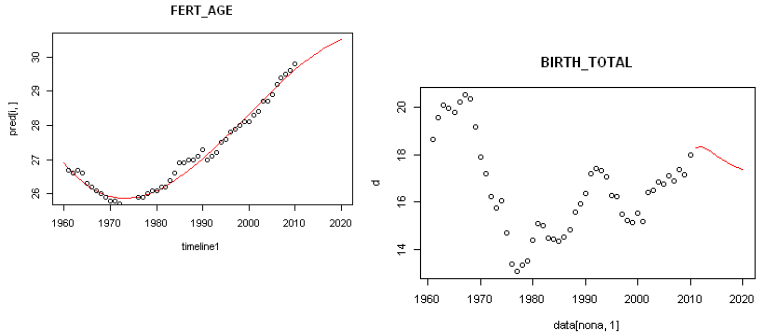
- **Questions:**

- *How will Seestadt Aspern perform in comparison to the other districts of Vienna?*

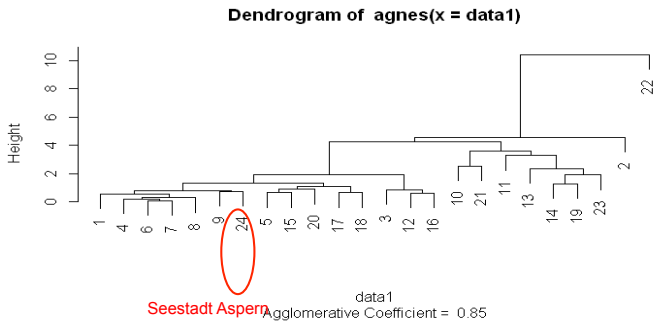
- *How do the goal indicators from the Aspern Masterplan compare with the "typical district behavior" of Vienna?*

Statistical Methods

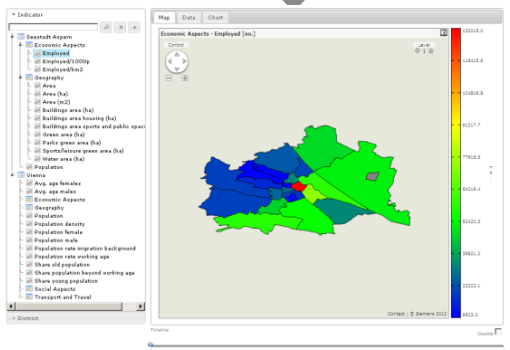




Completion & Prediction by statistical methods



finding patterns in the data



Vienna: districts and Seestadt Aspern

Berlin

Budapest

Vienna

Aim: Monitor development and compare to other cities/districts in order to take most effective infrastructural measures.

Applications & Challenges

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

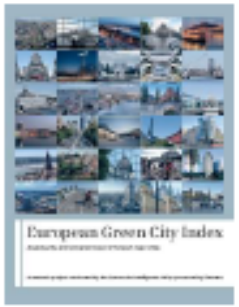
1. Application: trend prediction - Example “Seestadt Aspern”
2. **Integration & enrichment of “Green City Index” Data**
3. Challenges with Open Data Experienced

EFMD EQUIS ACCREDITED

Collected Data vs. Green City Index Data: Overlaps

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Together with colleagues from CC, we identified 20 quantitative raw data indicators that are overlapping between the Siemens' "*Green City Index*" and our current Data sources. The picture below visualizes the availability of data for these indicators for the cities of the European GCI:



European Green City Index

City / Raw Indicator	Gr	Bu	Po	Are	La	GD	GE	GD	Oz	NO	PM	Ho	Jo	re	Jo	Le	Jo	Jo	rec	Le	Le	Le	Le	Av
Amsterdam	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x
Antwerp	x	x			x	x	x	x	x	x	x		x	x		x			x				x	
Athens	x	x			x				x	x	x					x								x
Belgrade		x	x	x																				
Berlin	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Bratislava	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x			
Bremen	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Brussels	x	x	x	x	x	x	x	x	x	x	x		x	x		x			x	x	x	x	x	
Bucharest	x	x	x	x	x	x		x	x	x	x					x								
Budapest	x	x		x	x	x	x	x	x	x	x			x		x	x		x	x	x	x	x	x
Cologne	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Copenhagen	x	x		x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x		x	x	x
Dublin		x	x	x		x		x	x	x	x		x	x	x									
Essen	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Frankfurt		x		x																				x
Gothenburg	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			x		
Hamburg	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Hanover	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Helsinki	x	x			x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x
Istanbul		x																						
Kiev		x	x	x																				
Leipzig	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Lisbon		x	x	x	x	x	x	x	x	x	x		x			x	x	x		x	x		x	
Ljubljana	x	x	x	x	x				x	x	x		x	x		x				x	x	x	x	
London		x	x	x	x	x	x	x	x	x	x	x		x		x			x					
Luxembourg_(city)	x	x			x	x	x	x	x	x	x		x	x	x	x			x		x	x	x	
Madrid	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x			x	x	x
Malm%C3%B6	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Munich	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Nuremberg	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Oslo		x		x																				x
Paris		x		x	x	x	x	x	x	x	x			x										
Prague	x	x		x	x	x	x	x	x	x			x	x	x	x								
Riga	x	x		x	x	x	x	x	x	x	x	x		x		x			x	x	x	x		
Rome	x	x		x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x		x
Rotterdam	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x			
Sofia	x	x	x	x	x	x	x	x	x	x				x		x								x
Stockholm	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Stuttgart	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
Tallinn	x	x	x	x	x	x	x	x	x	x	x		x	x		x	x	x	x	x	x	x		
The_Hague	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	
Vienna	x	x	x	x	x	x	x	x	x	x			x	x	x	x			x			x		
Vilnius	x	x	x	x	x	x		x	x	x	x			x		x								
Warsaw	x	x	x	x	x	x	x	x	x	x	x			x		x				x	x	x	x	x
Zagreb		x		x																				
Zurich		x		x																				

>65% of raw date could be covered by publically available data that we have collected automatically

Data quality?

- Not all indicators are 100% comparable (different scales, units, etc., sources of different quality)
- for some indicators (e.g. Population) already less than 2% median error.
- The more data we collect, the better the quality!

EFMD EQUIS ACCREDITED

Applications & Challenges

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

1. Application: trend prediction - Example “Seestadt Aspern”
2. Integration & enrichment of “Green City Index” Data
3. **Challenges with Open Data Experienced**

EFMD EQUIS ACCREDITED

Challenges & Lessons Learnt – Is Open Data fit for industry?

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Base assumption (for our use case):

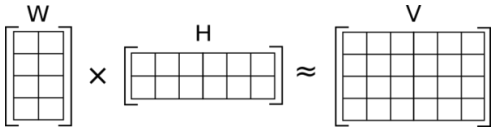
Added value comes from **comparable** Open datasets being **combined**

EFMD EQUIS ACCREDITED

Challenges & Lessons Learnt – Is Open Data fit for industry?

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- **Incomplete** Data: can be partially overcome
  - By ontological reasoning (RDF & OWL), by aggregation, or by rules & equations, e.g.
    - *:populationDensity = :population / :area* , cf. **[ESWC2013]**
  - *by statistical methods or Multi-dimensional Matrix Decomposition:*



unfortunately only partially successful, because these algorithms assume normally-distributed data.

- **Incomparable** Data:

dbpedia:populationTotal

dbpedia:populationCensus

- **Heterogeneity** across Open Government Data efforts:
  - Different ***Indicators***, Different Temporal and Spatial ***Granularity***
  - Different ***Licenses*** of Open Data: e.g. CC-BY, country specific licences, etc.
  - Heterogeneous ***Formats*** (CSV != CSV) ... Maybe the W3C CSV on the Web WG will solve this issue)

→Open Data needs stronger standards to be useful

[ESWC2013] Stefan Bischof and Axel Polleres. RDFS with attribute equations via SPARQL rewriting. In *Proc. Of the 10th ESWC*, vol. 7882 of *Lecture Notes in Computer Science (LNCS)*, p. 335-350, May 2013. Springer.

EFMD EQUIS ACCREDITED