

# OpenCube

Paul Hermans  
ProXML  
{paul}@proxml.be



Platform Linked Data Nederland, Amsterdam, 2 April 2015



## OpenCube background - OGD

- More than 180 Open Government Data portals around the globe provide data that “can be freely used, reused and redistributed by anyone”



## OpenCube background - Statistics

---

- A big portion of Open Government Data concerns statistics such as population figures, economic and social indicators
- For example, the majority (5867 out of 6098 datasets) of the data published on the EU Open Data Portal are of statistical nature



# Problem definition

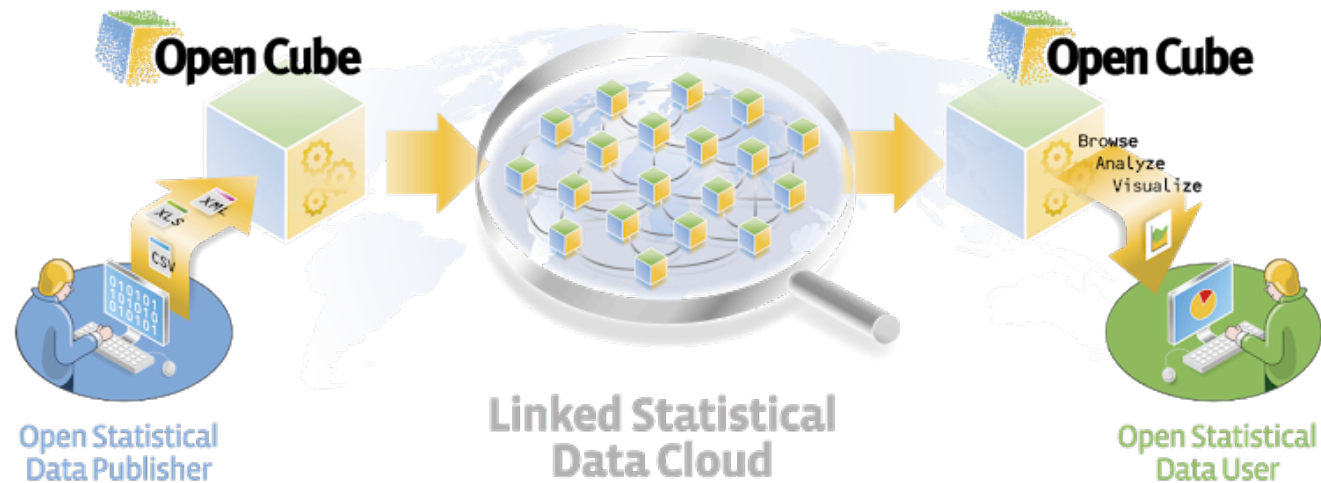
- Open Statistical Data are very **important** for the EU
- Users frequently want to **blend & combine** statistical data from multiple sources
- But, these data usually resides in files and databases (**data silos**) that are hard to combine
- **Linked Data (LD)** technology has the potential to enable **combining** and **performing analytics** on top of disparate and **previously isolated** statistical data
- However, relevant **tools** are few, scattered and un-tested under real-life conditions



**We need to break down statistical data silos**

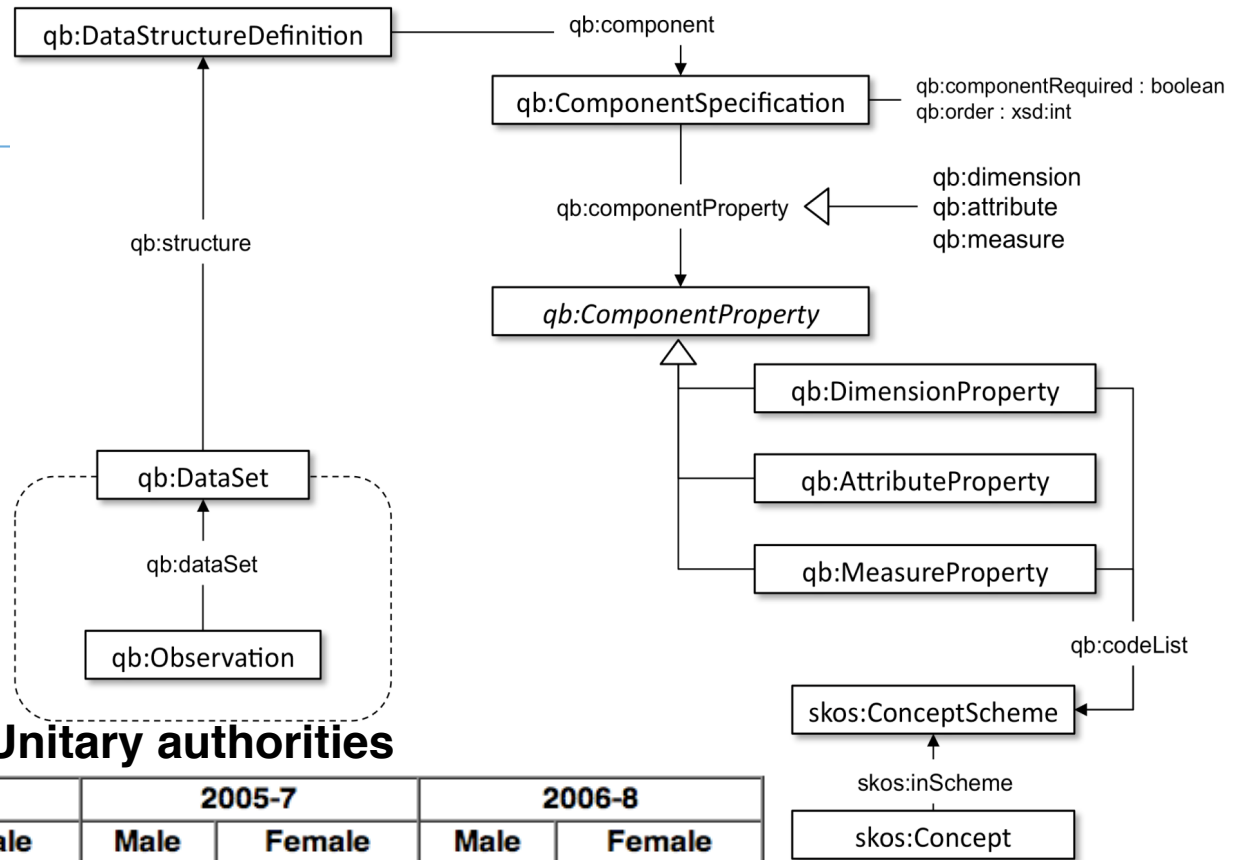
**Potential of using LD in statistical data analysis unexploited**

# The OpenCube project



- Aims:
  - Facilitate data **publishers** to create linked data cubes from legacy formats
  - Empower data **users** to browse, visualise, link, expand and analyse data cubes.
  - Enable **analysis** not possible before (**merging data cubes** at a Web scale)

# The model



## Life expectancy within Welsh Unitary authorities

	2004-6		2005-7		2006-8	
	Male	Female	Male	Female	Male	Female
<b>Newport</b>	76.7	80.7	77.1	80.9	77.0	81.5
<b>Cardiff</b>	78.7	83.3	78.6	83.7	78.7	83.4
<b>Monmouthshire</b>	76.6	81.3	76.5	81.5	76.6	81.7
<b>Merthyr Tydfil</b>	75.5	79.1	75.5	79.4	74.9	79.6

# Standards used



## Data Documentation Initiative

What is DDI? DDI Alliance DDI At Work Resources Specification RDF Vocabularies

W3C Recommendation

W3C

The

W3C

This v

Latest

Implem

Previo

Editors

Contrib

Please

This do

The En

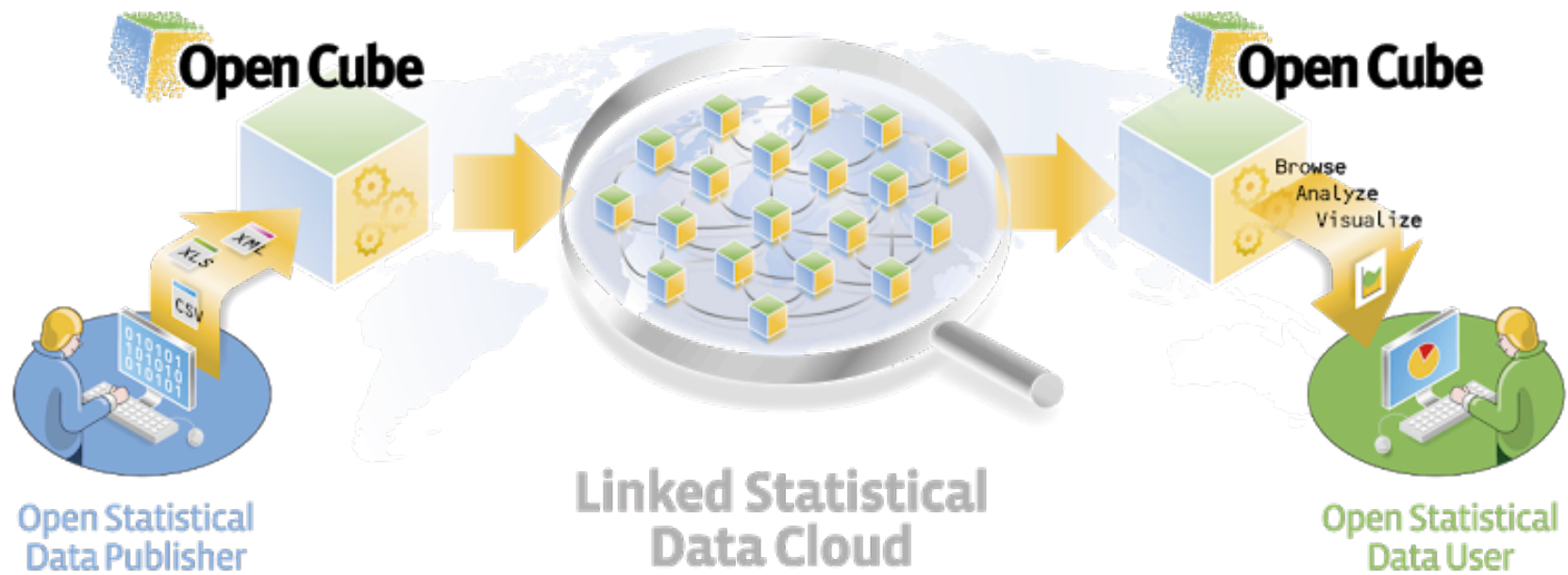
463

```
439 .
440 <http://id.vlaanderen.be/statistieken/dq/kubus-gemiddelde-prijs/observatie/10037#id>
441   rdf:type qb:Observation ;
442   statsvl:aantal "89431"^^xsd:int ;
443   statsvl:goedtype <http://id.fedstats.be/concept/goedtype_woonhuis#id> ;
444   statsvl:refArea <http://id.fedstats.be/nis/23101#id> ;
445   statsvl:timePeriod <http://id.vlaanderen.be/statistieken/concept/jaar_1994#id> ;
446   qb:dataSet <http://id.vlaanderen.be/statistieken/dq/kubus-gemiddelde-prijs#id> ;
447 .
448 <http://id.vlaanderen.be/statistieken/dq/kubus-gemiddelde-prijs/observatie/10038#id>
449   rdf:type qb:Observation ;
450   statsvl:aantal "285479"^^xsd:int ;
451   statsvl:goedtype <http://id.fedstats.be/concept/goedtype_villa#id> ;
452   statsvl:refArea <http://id.fedstats.be/nis/23101#id> ;
453   statsvl:timePeriod <http://id.vlaanderen.be/statistieken/concept/jaar_1994#id> ;
454   qb:dataSet <http://id.vlaanderen.be/statistieken/dq/kubus-gemiddelde-prijs#id> ;
455 .
456 <http://id.vlaanderen.be/statistieken/dq/kubus-gemiddelde-prijs/observatie/10039#id>
457   rdf:type qb:Observation ;
458   statsvl:aantal "64671"^^xsd:int ;
459   statsvl:goedtype <http://id.fedstats.be/concept/goedtype_appartement#id> ;
460   statsvl:refArea <http://id.fedstats.be/nis/23101#id> ;
461   statsvl:timePeriod <http://id.vlaanderen.be/statistieken/concept/jaar_1994#id> ;
462   qb:dataSet <http://id.vlaanderen.be/statistieken/dq/kubus-gemiddelde-prijs#id> ;
463 .
```

Copyright © 2012–2014 W3C (MIT, ERCIM, Keio, Beijing), All Rights Reserved. W3C liability, trademark and document use rules apply.

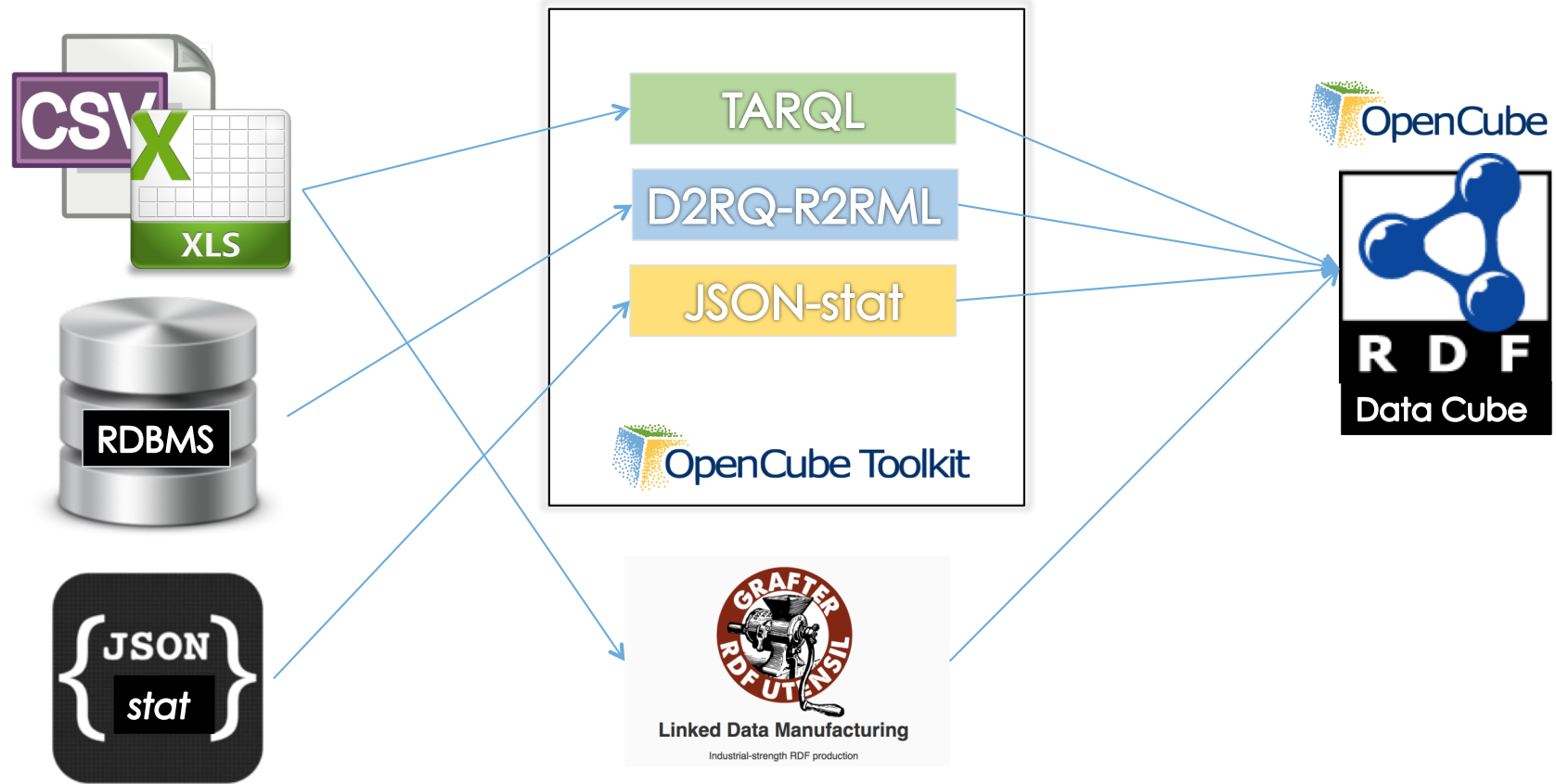


# The OpenCube tools





# Publishing Components



# Consume: OpenCube browser

- It enables the **exploration** of an RDF data cube by presenting a **two-dimensional slice** of the cube as a **table**.
- The slice is created by setting a **fixed values for each dimension** that is not presented in the table.

Summarize observations across a dimension (dimension reduction)

Change the language

The screenshot shows the OpenCube browser interface. At the top, there's a header "OpenCube Browser" and a sub-header "The OpenCube browser enables the exploration of an RDF Data Cube by presenting each time a two-dimensional slice". Below this, there are several control panels:

- Dimensions:** Summarize observations by adding/removing dimensions. It includes checkboxes for Age class, Sex, Country of citizenship, Geopolitical entity (reporting), and timePeriod, all of which are checked.
- Language:** Select the language of the visualized data. A dropdown menu shows "en".
- Visual dimensions:** Select the two dimensions that define the table of the browser. It includes "Column Headings:" with a dropdown set to "Age class" and "Rows (values in first column):" with a dropdown set to "Geopolitical entity (reporting)".
- Fixed dimensions:** Change the values of the fixed dimensions. It includes "Sex:" with a dropdown set to "Females", "Country of citizenship:" with a dropdown set to "Foreign country", and "timePeriod:" with a dropdown set to "1991-01-01".

The main part of the interface is a table with the following data:

Geopolitical entity (reporting)	65 years or over	80 years or over	From 10 to 14 years	From 15 to 19 years	From 15
Austria	6822	3126	14662	18094	164910
Belgium	19951	6819	34385	33789	283622
Bulgaria	-	-	-	-	-
Cyprus	-	-	-	-	-
Czech Republic	-	-	-	-	-
Denmark	1537	390	5187	5592	52668
Estonia	-	-	-	-	-
Finland	511	319	687	533	7939
France	63976	43705	139108	125757	1126495
Germany (until 1990 former territory of the FRG)	-	-	-	-	-
Greece	3764	1263	4150	-	-
Hungary	-	-	-	-	-
Iceland	-	-	-	-	-
Ireland	-	-	-	-	-
Italy	6072	2852	6693	8005	136021

Change the axes of the table

Change the fixed values

# Consume: OpenCube MapView

- Visualization of RDF data cubes on a map.
- It supports:
  - Markers
  - Bubble
  - Choropleth maps

opencube-toolkit.eu

### OpenCube MapView

The OpenCube Map View enables the visualization of RDF data cubes on a map based on their geospatial dimension.

**Type of map**  
In order to view the map please select one of the following map types:

Choropleth map

**Dimensions**  
Summarize observations by adding/removing dimensions:

freq

Reason

timePeriod

Enterprise

Classification of economic activities - NACE Rev.2

**Language**  
Select the language of the visualized data:

en

±

**Fixed dimensions**  
Change the values of the fixed dimensions:

freq: freq-A

Reason: Bank branch known for good client relationships

timePeriod: 2007-01-01

Enterprise: Other enterprises

Classification of economic activities - NACE Rev.2: Construction

This data set contains unrounded figures, rounded figures are available in Table 253, available for download as an Excel spreadsheet.

#### Mapper

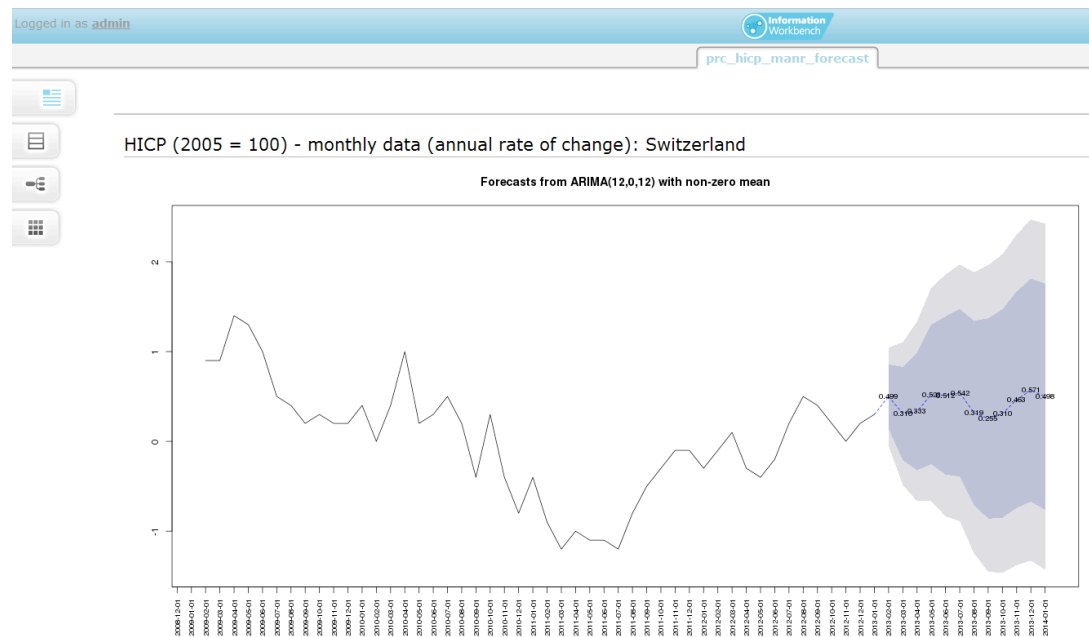
This dataset contains multidimensional data (a data cube) which can be displayed as a grid to compare two dimensions at a time. Use the drop-down menus below the grid to choose which dimensions to show as rows and columns (and, optionally, to filter the other dimensions by value).

#### Spreadsheet view

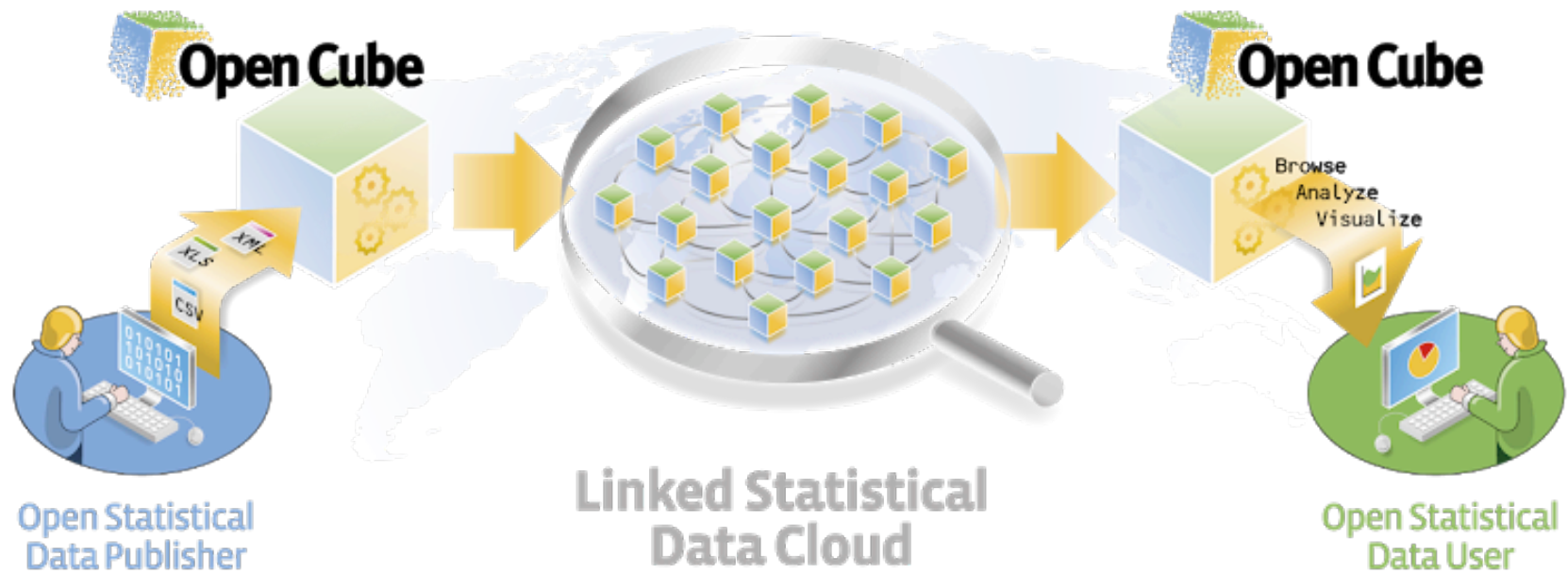
Reference area	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014
E06000001 Hartlepool	230	150	190	170	170
E06000002 Middlesbrough	320			220	
E06000003 Redcar and Cleveland	210	250	260	230	270
E06000004...					

# Consume: Integration with R

- Visualisation of analysis results (charts & tables)
- Reuse of analysis results: preserving R output as linked data



# The OpenCube tools



## Linking Statistical Data

---

- Enables Performing analytics on top of combined data cubes
- Steps:
  1. Select a data cube
  2. Discover cubes on the Web of Linked Data having **compatible structure**; i.e. cubes with dimensions, measures etc. that can expand the initial cube
  3. Create expanded views of the initial cube
  4. Consume the new cube(s)

# Example: Start with an initial cube

Dimensions

Measures

Operations

Please select a cube:

**Cube Dimensions**

Energy indicator

- [Final energy consumption - Industry](#)
- [Final energy consumption - Iron and steel industry](#)
- [Final energy consumption - Non-ferrous metal industry](#)
- [Final energy consumption - Chemical industry](#)
- [Final energy consumption - Ore extraction \(except fuels\) industry](#)
- [Final energy consumption - Food, drink and tobacco industry](#)
- [Final energy consumption - Textile, leather and clothing industry](#)
- [Final energy consumption - Paper and printing industry](#)
- [Final energy consumption - Engineering and other metal industry](#)
- [Final energy consumption - Other non-classified industries](#)
- [Final energy consumption - Adjustment](#)

Geopolitical entity (reporting)

Products

**Cube measures:**

- obsValue

Please select an operation:

# Example: Discover & Select compatible cubes

Please select a cube:  
[http://eurostat.linked-statistics.org/data/med\\_eg3](http://eurostat.linked-statistics.org/data/med_eg3)

**Cube Dimensions**

- Energy indicator
  - Final energy consumption - Industry
  - Final energy consumption - Iron and steel industry
  - Final energy consumption - Non-ferrous metal industry
  - Final energy consumption - Chemical industry
  - Final energy consumption - Ore extraction (except fuels) industry
  - Final energy consumption - Food, drink and tobacco industry
  - Final energy consumption - Textile, leather and clothing industry
  - Final energy consumption - Paper and printing industry
  - Final energy consumption - Engineering and other metal industry
  - Final energy consumption - Other non-classified industries
  - Final energy consumption - Adjustment
- Geopolitical entity (reporting)
- Products

**Cube measures:**

- obsValue

Please select an operation:  
Add value to level

Energy indicator

Execute operation

**Available values to add to dimension: Energy indicator:**

[http://eurostat.linked-statistics.org/data/med\\_eg5](http://eurostat.linked-statistics.org/data/med_eg5)

- Final energy consumption - Agriculture
- Final energy consumption - Households
- Final energy consumption - Households/Services
- Final energy consumption - Other Sectors

[http://eurostat.linked-statistics.org/data/med\\_eg21](http://eurostat.linked-statistics.org/data/med_eg21)

- Input to autoproducer thermal power stations
- Input to blast-furnace plants
- Input to coke-oven plants
- Input to district heating plants
- Input to gas-works
- Input to nuclear power stations
- Input to patent fuel and briquetting plants
- Input to public thermal power stations
- Input to refineries
- Output from Refineries
- Output from autoproducer thermal power stations
- Output from blast-furnace plants
- Output from coke-oven plants
- Output from district heating plants
- Output from gas-works
- Output from nuclear power stations
- Output from patent fuel and briquetting plants
- Output from public thermal power stations
- Transformation input
- Transformation output

[http://eurostat.linked-statistics.org/data/med\\_eg22](http://eurostat.linked-statistics.org/data/med_eg22)

- Consumption - Energy sector
- Distribution losses
- Energy available for final consumption
- Final energy consumption
- Final non-energy consumption
- Final non-energy consumption - Chemical industry
- Final non-energy consumption - Non-chemical industries

[http://eurostat.linked-statistics.org/data/med\\_eg4](http://eurostat.linked-statistics.org/data/med_eg4)

- Final energy consumption - Air transport
- Final energy consumption - Inland navigation
- Final energy consumption - Rail transport
- Final energy consumption - Road transport
- Final energy consumption - Transport

Add attribute values

Compatible cubes to add new value(s) to a level (dimension)



# Example: Browse an expanded view of the initial cube

**Dimensions**  
Summarize observations by adding/removing dimensions:

- Energy indicator
- timePeriod
- Geopolitical entity (reporting)
- Products

**Measures**  
Select the measures to visualize:

- obsValue

**Language**  
Select the language of the visualized data:

en ▼

Energy indicator	All petroleum products	All products	BKB/PB	Biomass and renewable wastes	Coke	Crude oil and NGL	Derived gases
Final energy consumption - Adjustment	1513.55	2082.37	0		50.16	0	0
Final energy consumption - Agriculture	1398.48	1643.75	0		0	0	0
Final energy consumption - Chemical industry	0	0	0		0	0	0
Final energy consumption - Engineering and other metal industry	0	20.85	0		0	0	0
Final energy consumption - Food, drink and tobacco industry	255.52	426.48	0		0	0	0
Final energy consumption - Households	1362.54	2757.07	0		469.92	0	0
Final energy consumption - Households/Services	2801.63	4559.63	0		499.62	0	0
Final energy consumption - Industry	1769.07	2554.78	0		50.16	0	0
Final energy consumption - Iron and steel industry	0	0	0		0	0	0
Final energy consumption - Non-ferrous metal industry	0	0	0		0	0	0
Final energy consumption - Ore extraction (except fuels) industry	0	0	0		0	0	0
Final energy consumption - Other Sectors	0	0	0		0	0	0
Final energy consumption - Other non-classified industries	0	0	0		0	0	0
Final energy consumption - Paper and printing industry	0	0	0		0	0	0
Final energy consumption - Textile, leather and clothing industry	0	25.08	0		0	0	0

# Gratis Linked Statistics workshop

---

- Amersfoort, 8 mei
- Inhoud
  - RDF Data Cube vocabulary
  - SKOS en XKOS
  - Specificeren van de transformatie
  - Bevraag de datasets
  - Visualiseer de datasets
  - Link de datasets

# Questions?

- For more information

- <http://opencube-project.eu>
- <http://opencube-toolkit.eu>

- OpenCube consortium

