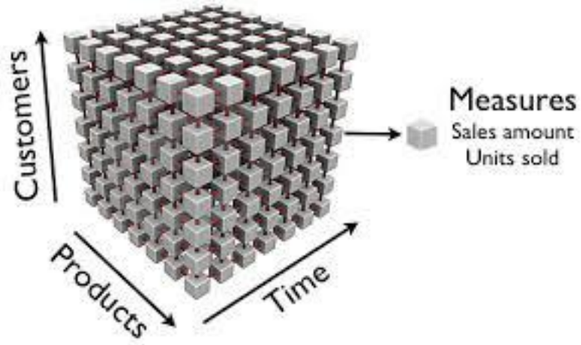
Linked Statistics OpenCube

Paul Hermans

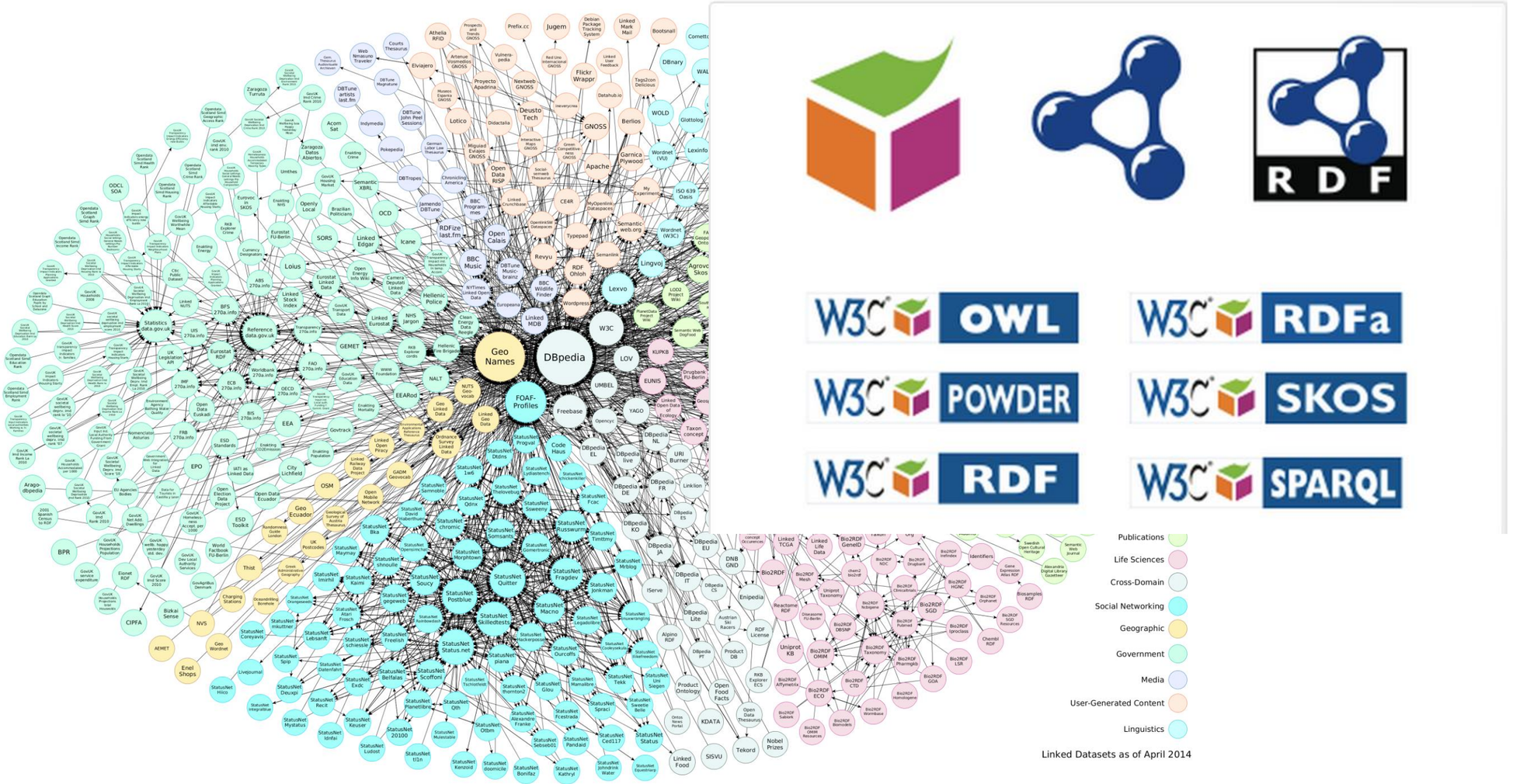
paul@proxml.be

Statistics

- Statistical data is often organised as **data cubes**, where each **cell** contains a **measure** described based on a number of **dimensions**



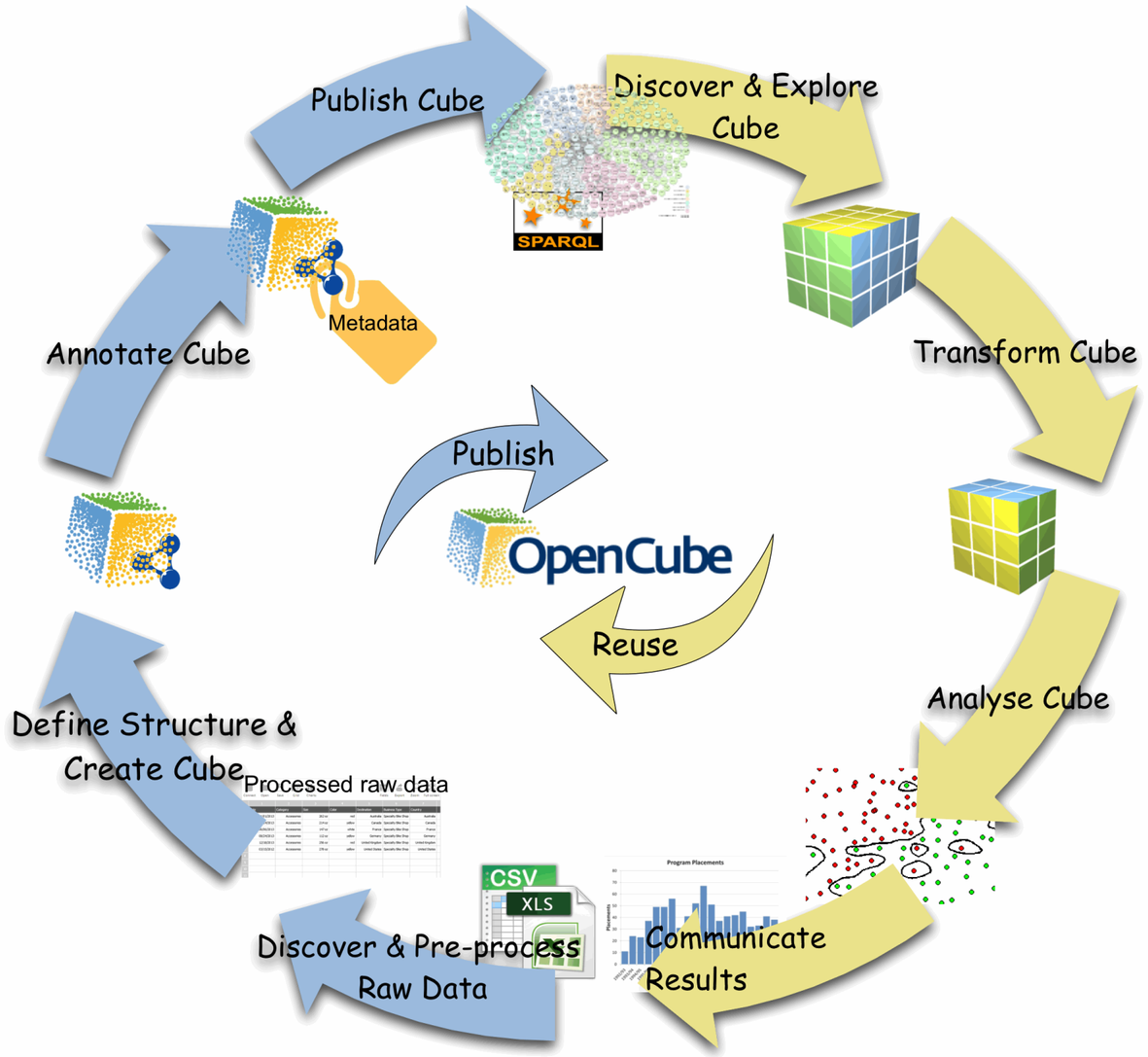
Linked Data


RDF
W3C OWL
W3C RDFa
W3C POWDER
W3C SKOS
W3C RDF
W3C SPARQL
Geo Names
DBpedia
FOAF-Profiles
Publications
Life Sciences
Cross-Domain
Social Networking
Geographic
Government
Media
User-Generated Content
Linguistics
Linked Datasets as of April 2014

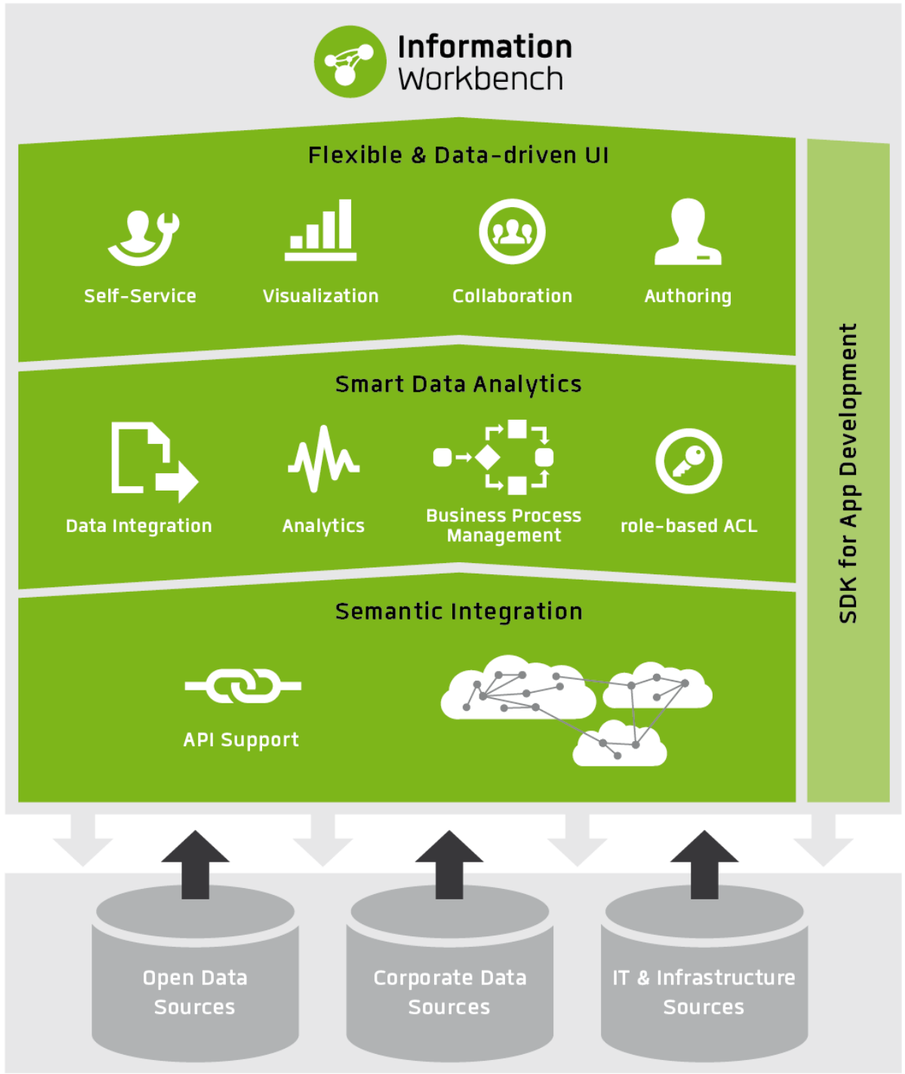
OpenCube


Open Cube
XML
XLS
CSV
Open Statistical
Data Publisher
Linked Statistical
Data Cloud
Open Cube
Browse
Analyze
Visualize
Open Statistical
Data User

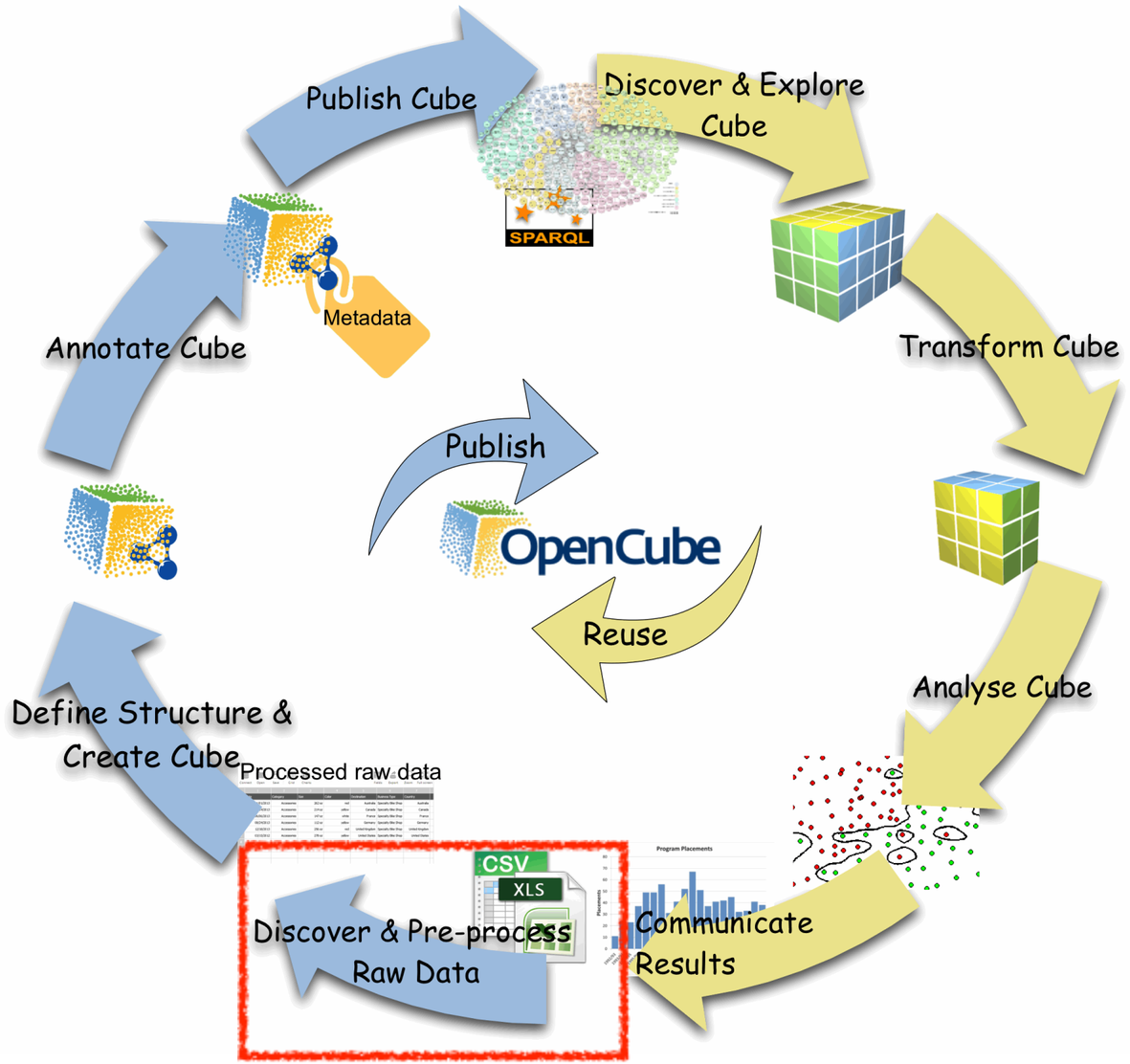
Lifecycle


Publish Cube
Discover & Explore Cube
SPARQL
Transform Cube
Analyse Cube
Communicate Results
Discover & Pre-process Raw Data
CSV
XLS
Define Structure & Create Cube
Processed raw data
Annotate Cube
Metadata
Publish
OpenCube
Reuse

fluidOps Information WorkBench



Discover and pre-process



Discover and pre-process (1)

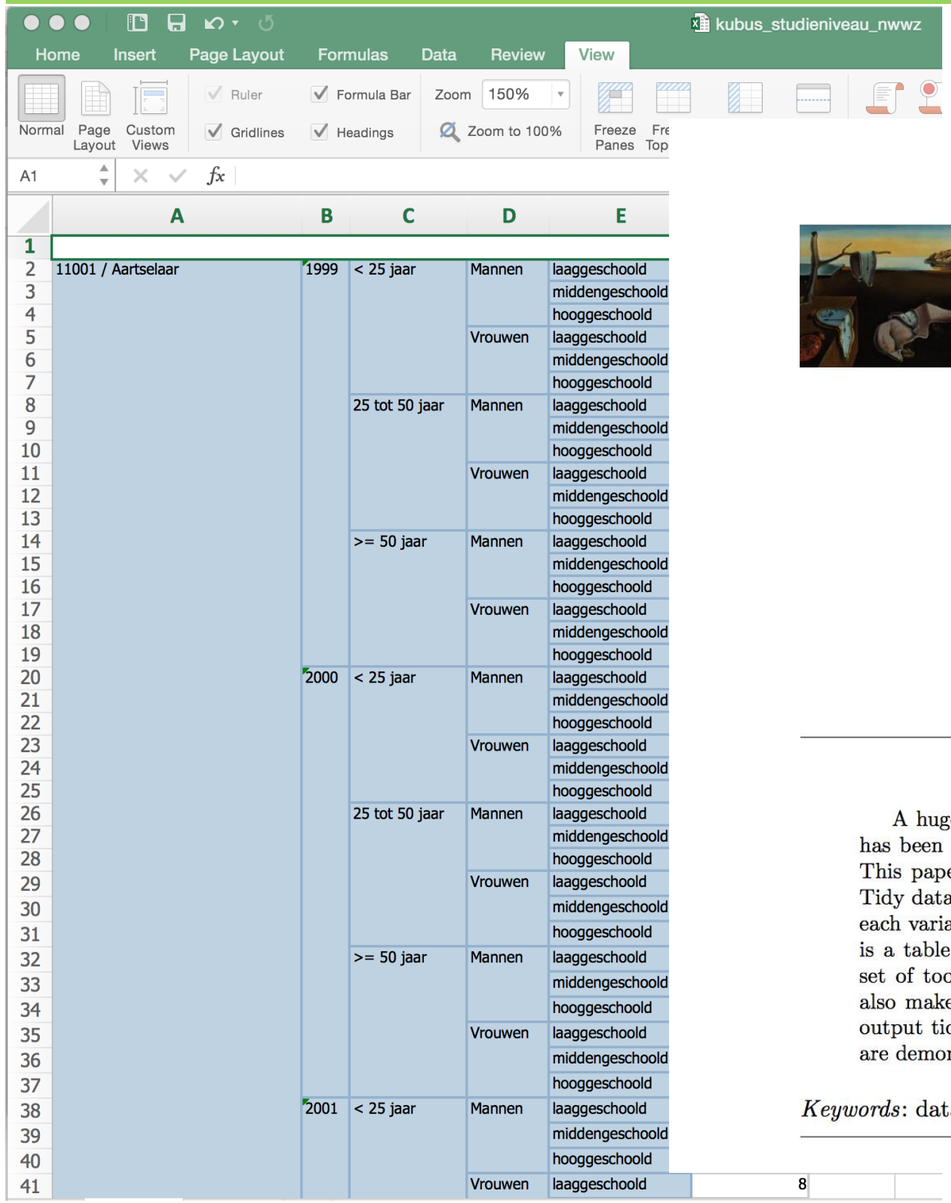
IBM Cognos Viewer - kubus_studieniveau_nwwz

eahermanpa Afmelden Info IBM.

Deze versie behouden | Dit rapport toevoegen

aantal NWWZ	1999												Leeftijdsgroep: geen filter				2000
	< 25 jaar				25 tot 50 jaar				>= 50 jaar								< 25 jaar
	laaggeschoold	middengeschoold	hooggeschoold	Studieniveau: geen filter	laaggeschoold	middengeschoold	hooggeschoold	Studieniveau: geen filter	laaggeschoold	middengeschoold	hooggeschoold	Studieniveau: geen filter	laaggeschoold	middengeschoold	hooggeschoold	Studieniveau: geen filter	laaggeschoold
24001 / Aarschot	79	82	22	**183**	296	169	40	**505**	38	9	2	**49**	**413**	**260**	**64**	**737**	61
24007 / Begijnendijk	20	19	6	**45**	84	41	13	**138**	12	3	0	**15**	**116**	**63**	**19**	**198**	14
24008 / Bekkevoort	16	15	2	**33**	53	26	8	**87**	7	2	1	**10**	**76**	**43**	**11**	**130**	11
24009 / Bertem	9	17	5	**31**	26	28	23	**77**	9	2	3	**14**	**44**	**47**	**31**	**122**	7
24011 / Bierbeek	10	9	6	**25**	52	37	22	**111**	11	3	2	**16**	**73**	**49**	**30**	**152**	14
24014 / Boortmeerbeek	17	17	4	**38**	65	45	21	**131**	15	4	0	**19**	**97**	**66**	**25**	**188**	11
24016 / Boutersem	10	5	3	**18**	39	20	13	**72**	6	1	0	**7**	**55**	**26**	**16**	**97**	10
24020 / Diest	97	77	12	**186**	314	171	45	**530**	34	8	4	**46**	**445**	**256**	**61**	**762**	63
24028 / Geetbets	13	11	1	**25**	76	43	5	**124**	4	2	0	**6**	**93**	**56**	**6**	**155**	11
24033 / Haacht	23	25	2	**50**	70	63	27	**160**	19	8	2	**29**	**112**	**96**	**31**	**239**	17
24038 / Herent	30	18	7	**55**	95	75	47	**217**	12	6	4	**22**	**137**	**99**	**58**	**294**	29
24041 / Hoegaarden	10	12	2	**24**	61	33	15	**109**	7	2	1	**10**	**78**	**47**	**18**	**143**	14
Meer																	
Ar.Leuven	**1.078**	**954**	**278**	**2.310**	**3.926**	**2.535**	**1.319**	**7.780**	**532**	**189**	**133**	**854**	**5.536**	**3.678**	**1.730**	**10.944**	**936**

Discover and pre-process (2)



	A	B	C	D	E
1					
2	11001 / Aartselaar	1999	< 25 jaar	Mannen	laaggeschoold
3					middengeschoold
4					hooggeschoold
5				Vrouwen	laaggeschoold
6					middengeschoold
7					hooggeschoold
8			25 tot 50 jaar	Mannen	laaggeschoold
9					middengeschoold
10					hooggeschoold
11				Vrouwen	laaggeschoold
12					middengeschoold
13					hooggeschoold
14			>= 50 jaar	Mannen	laaggeschoold
15					middengeschoold
16					hooggeschoold
17				Vrouwen	laaggeschoold
18					middengeschoold
19					hooggeschoold
20		2000	< 25 jaar	Mannen	laaggeschoold
21					middengeschoold
22					hooggeschoold
23				Vrouwen	laaggeschoold
24					middengeschoold
25					hooggeschoold
26			25 tot 50 jaar	Mannen	laaggeschoold
27					middengeschoold
28					hooggeschoold
29				Vrouwen	laaggeschoold
30					middengeschoold
31					hooggeschoold
32			>= 50 jaar	Mannen	laaggeschoold
33					middengeschoold
34					hooggeschoold
35				Vrouwen	laaggeschoold
36					middengeschoold
37					hooggeschoold
38		2001	< 25 jaar	Mannen	laaggeschoold
39					middengeschoold
40					hooggeschoold
41				Vrouwen	laaggeschoold 8

Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *http://www.jstatsoft.org/*

Tidy Data

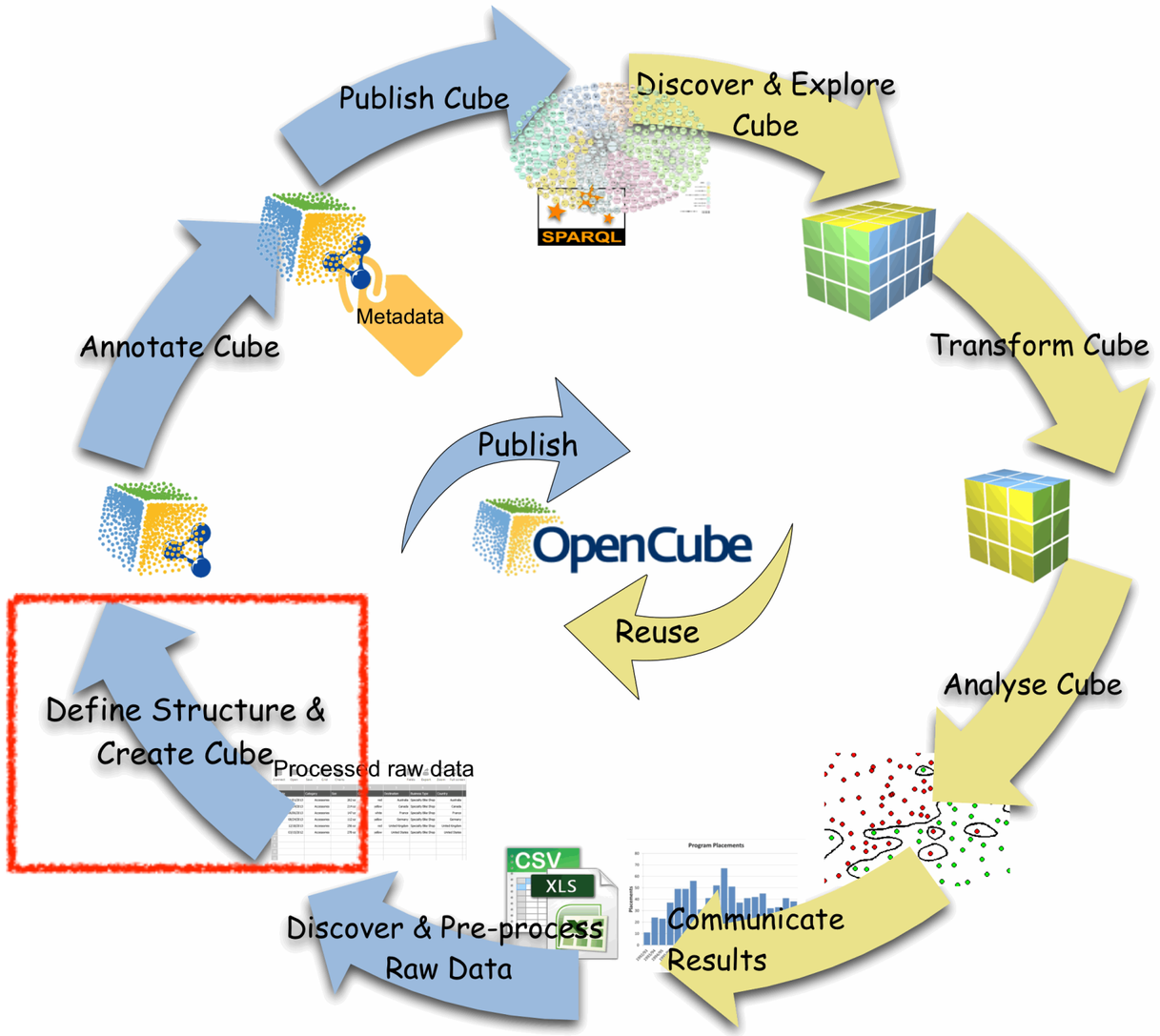
Hadley Wickham
RStudio

Abstract

A huge amount of effort is spent cleaning data to get it ready for analysis, but there has been little research on how to make data cleaning as easy and effective as possible. This paper tackles a small, but important, component of data cleaning: data tidying. Tidy datasets are easy to manipulate, model and visualise, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table. This framework makes it easy to tidy messy datasets because only a small set of tools are needed to deal with a wide range of un-tidy datasets. This structure also makes it easier to develop tidy tools for data analysis, tools that both input and output tidy datasets. The advantages of a consistent data structure and matching tools are demonstrated with a case study free from mundane data manipulation chores.

Keywords: data cleaning, data tidying, relational databases, R.

Discover and pre-process



Standards

W3C Recommendation

The RDF Data Cube Vocabulary

W3C Recommendation 16 January 2014

This version:
http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/
Latest published version:
http://www.w3.org/TR/vocab-data-cube/
Implementation report:
http://www.w3.org/2011/gld/wiki/Data_Cube_Implementations
Previous version:
http://www.w3.org/TR/2013/PR-vocab-data-cube-20131217/
Editors:
Richard Cyganiak, DERI, NUI Galway
Dave Reynolds, Epimorphics Ltd
Contributors:
Jeni Tennison

Please refer to the **errata**, a list of issues with this document discovered after publication.

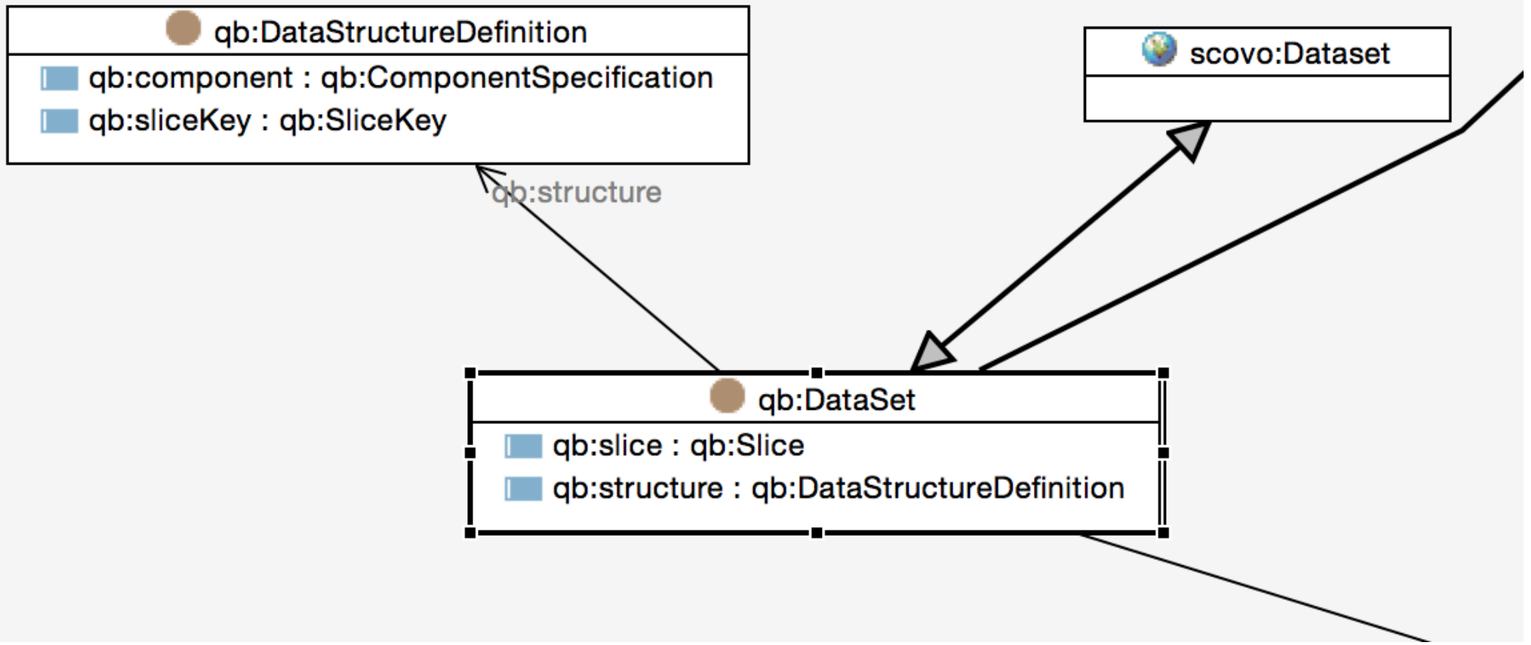
This document is also available in this non-normative format: diff to previous version

The English version of this specification is the only normative version. Non-normative translations may also be available.

Classes

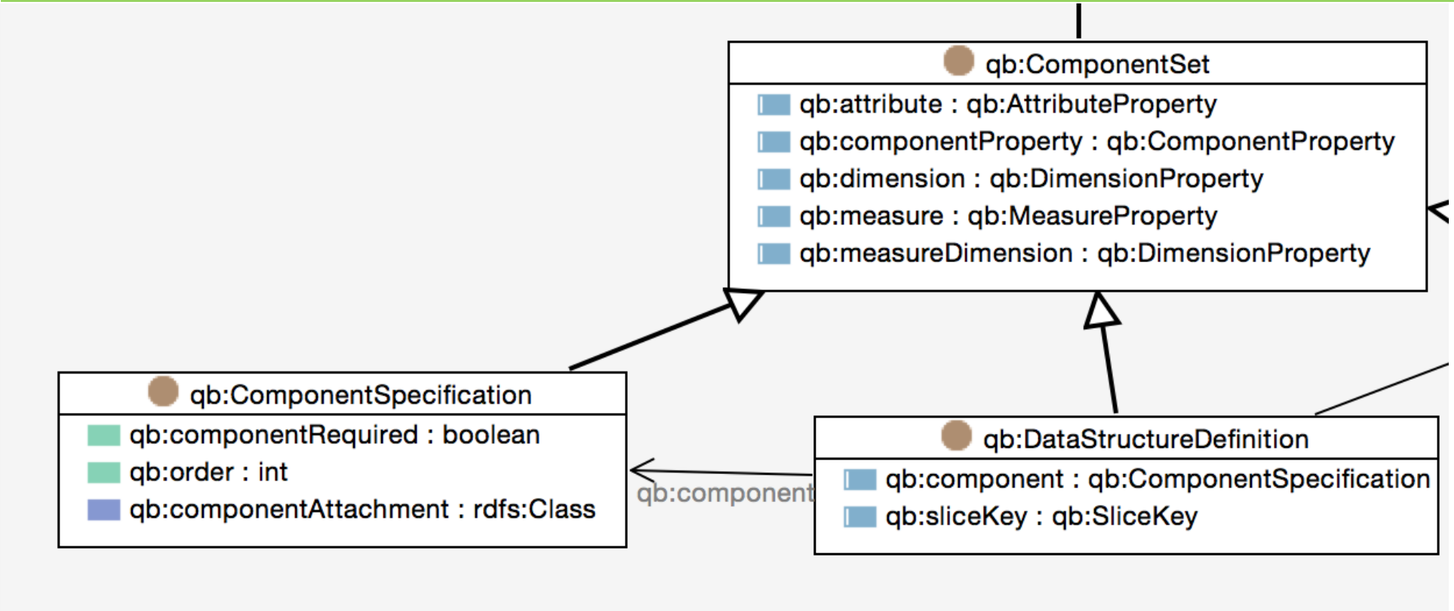
- DataSet
- DataStructureDefinition
  - ComponentSpecification
- Observation

qb:DataSet




```
<http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz#id>
        a               qb:DataSet ;
        rdfs:label      "Kubus studieniveau nwwz"@nl ;
        qb:structure  <http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz/dsd#id> .
```

qb:DataStructureDefinition




```
<http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz/dsd#id>
        a               qb:DataStructureDefinition .
```

Components

12.4 Dimensions, Attributes, Measures

See Section Dimensions, attributes and measures.

Class: `qb:Attachable`
Abstract superclass for everything that can have attributes and dimensions.

Class: `qb:ComponentProperty` ***Sub class of:*** `rdf:Property`
Abstract super-class of all properties rep

Class: `qb:DimensionProperty` ***Sub class of:*** `q`
The class of component properties which

Class: `qb:AttributeProperty` ***Sub class of:*** `q`
The class of component properties which

Class: `qb:MeasureProperty` ***Sub class of:*** `qb:`
The class of component properties which

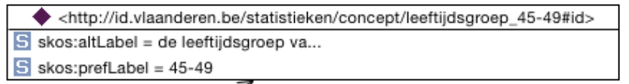
Class: `qb:CodedProperty` ***Sub class of:*** `qb:Co`
Superclass of all coded component prop

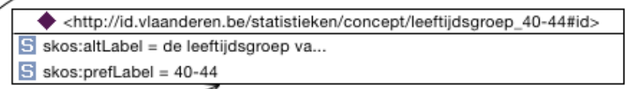
```
<http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz/dsd#id>
      a            qb:DataStructureDefinition ;
      qb:component  [ a           qb:ComponentSpec
                      qb:dimension  statsvl:refArea
                    ] ;
      qb:component  [ a           qb:ComponentSpecification ;
                      qb:dimension  statsvl:timePeriod
                    ] ;
      qb:component  [ a           qb:ComponentSpecification ;
                      qb:dimension  sdmx-dimension:sex
                    ] ;
      qb:component  [ a           qb:ComponentSpecification ;
                      qb:dimension  statsvl:ageGroup
                    ] ;
      qb:component  [ a           qb:ComponentSpecification ;
                      qb:dimension  statsvl:educationLev
                    ] ;
      qb:component  [ a          qb:ComponentSpecification ;
                      qb:measure  statsvl:jobSeekers
                    ] .
```

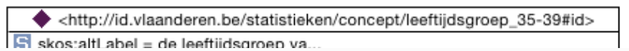
Pre-existing

```
sdmx-dimension:sex  a     qb:DimensionProperty , qb:CodedProperty , rdf:Property ;
        rdfs:comment      "The state of being male or female."@en ;
        rdfs:isDefinedBy  <http://sdmx.org/wp-content/uploads/2009/01/01_sdmx_cog_annex_1_cdc_2009.pdf> ;
        rdfs:label        "Geslacht"@nl , "Sex"@en ;
        rdfs:range        sdmx-code:Sex , rdfs:Resource , skos:Concept ;
        rdfs:seeAlso      <http://sdmx.org/wp-content/uploads/2009/01/01_sdmx_cog_annex_1_cdc_2009.pdf> ;
        qb:codeList       sdmx-code:sex ;
        qb:concept        sdmx-concept:sex .
```

SKOS



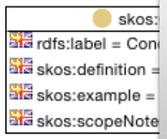


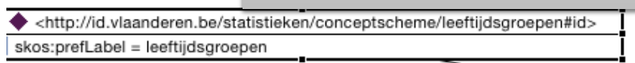


skos:narrower

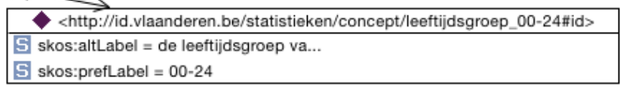
skos:narrower

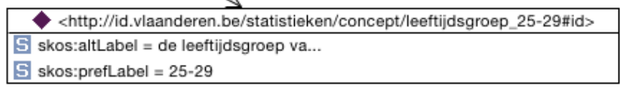
```
sdmx-code:sex  a            sdmx:CodeList , skos:ConceptScheme ;
       rdfs:label           "Code list for Sex (SEX) - codelist scheme"@en ;
       rdfs:seeAlso         sdmx-code:Sex ;
       skos:definition      <http://sdmx.org/wp-content/uploads/2009/01/02_sdmx_cog_annex_2_cl_2009.pdf> ;
       skos:hasTopConcept   sdmx-code:sex-U , sdmx-code:sex-M , sdmx-code:sex-F , sdmx-code:sex-N , sdmx-code:sex-T ;
       skos:notation        "CL_SEX" ;
       skos:note            "This code list provides the gender."@en ;
       skos:prefLabel       "Code list for Sex (SEX) - codelist scheme"@en .
```





skos:hasTopConcept





XKOS

Unofficial Draft

XKOS

An SKOS extension for representing statistical classifications

Unofficial Draft 28 May 2014

Editor:
Franck Cotton, INSEE
Authors:
Richard Cyganiak, DERI
Daniel Gillman, Bureau of Labor Statistics
R.T.A.M. Grim, Tilburg University
Yves Jaques, Food and Agriculture Organization
Wendy Thomas, Minnesota Population Center

XKOS: why?

- skos:broader/narrower ->
  - generic (generic-specific)
  - partitive (whole-part).
- statistical classifications are structured according to levels
- skos:related
  - causal,
  - sequential,
  - temporal
- mapping
  - e.g. NACE 2003 ⇔ NACE 2008

Hierarchical relations

```
<http://id.fedstats.be/nis/11022#id>
        a               skos:Concept ;
        xkos:isPartOf   <http://id.fedstats.be/nis/11000#id> ;
        void:inDataset  <http://id.fedstats.be/dataset/nis#id> ;
        skos:inScheme   <http://id.fedstats.be/conceptscheme/nis#id> ;
        skos:prefLabel  "Kalmthout"@de , "Kalmthout"@fr , "Kalmthout"@en , "Kalmthout"@nl .
```

Semantics Alert

Levels

```
<http://id.fedstats.be/classificationlevel/province#id>
        a               xkos:ClassificationLevel ;
        xkos:depth      "2"^^xsd:positiveInteger ;
        void:inDataset  <http://id.fedstats.be/dataset/nis#id> ;
        skos:member     <http://id.fedstats.be/nis/80000#id> , <http://id.fedstats.be/nis/10000#id> , <http://id.fedstats.be/nis/40000#id> , <http://id.fedstats.be/nis/50000#id> , <http://id.fed
        skos:prefLabel  "Provincie"@nl , "Province"@en , "Province"@fr , "Provinz"@de .
```

NIS

Codelists by Statistics Belgium

HTML JSON-LD TURTLE N-TRIPLES XML

Eigenbrakel

Pref Label	Eigenbrakel
In Scheme	NIS codelijst
In Dataset	NIS dataset
Notation	25014
Type	http://www.w3.org/2004/02/skos/core#...

http://www.w3.org/2004/02/skos/core#Concept

Broader	Arrondissement Nijvel (25000)

References from other resources

via 'Member' from	Gemeente

Status

```
<http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz/dsd#id>
        a               qb:DataStructureDefinition ;
        qb:component  [ a             qb:ComponentSpec
                        qb:dimension  statsvl:refArea
                      ] ;
        qb:component  [ a             qb:ComponentSpecification ;
                        qb:dimension  statsvl:timePeriod
                      ] ;
        qb:component  [ a             qb:ComponentSpecification ;
                        qb:dimension  sdmx-dimension:sex
                      ] ;
        qb:component  [ a             qb:ComponentSpecification ;
                        qb:dimension  statsvl:ageGroup
                      ] ;
        qb:component  [ a             qb:ComponentSpecification ;
                        qb:dimension  statsvl:educationLev
                      ] ;
        qb:component  [ a           qb:ComponentSpecification ;
                        qb:measure  statsvl:jobSeekers
                      ] .
```

Eigen dimensie

```
statsvl:refArea  a          qb:DimensionProperty , rdf:Property ;
        rdfs:comment        "The country or geographic area to which the measured statistical phenomenon relates."@en ;
        rdfs:label          "Referentie Area"@nl ;
        rdfs:range          skos:Concept ;
        rdfs:subPropertyOf  sdmx-dimension:refArea ;
        qb:codeList         <http://id.fedstats.be/conceptscheme/nis#id> ;
        qb:concept          sdmx-concept:refArea .
```

DSD

```
<http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz#id>
  rdf:type qb:DataSet ;
  qb:structure <http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz/dsd#id> ;
  rdfs:label "Kubus studieniveau nwwz"@nl ;
.
<http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz/dsd#id>
  rdf:type qb:DataStructureDefinition ;
  qb:component [
      rdf:type qb:ComponentSpecification ;
      qb:dimension statsvl:ageGroup ;
    ] ;
  qb:component [
      rdf:type qb:ComponentSpecification ;
      qb:dimension statsvl:educationLev ;
    ] ;
  qb:component [
      rdf:type qb:ComponentSpecification ;
      qb:dimension statsvl:refArea ;
    ] ;
  qb:component [
      rdf:type qb:ComponentSpecification ;
      qb:dimension statsvl:timePeriod ;
    ] ;
  qb:component [
      rdf:type qb:ComponentSpecification ;
      qb:dimension sdmx-dimension:sex ;
    ] ;
  qb:component [
      rdf:type qb:ComponentSpecification ;
      qb:measure statsvl:jobSeekers ;
    ] ;
.
```

DSD

```
statsvl:ageGroup
  rdf:type qb:DimensionProperty ;
  rdf:type rdf:Property ;
  qb:codeList <http://id.vlaanderen.be/statistieken/conceptscheme/leeftijdsgroepen#id> ;
  rdfs:comment "The length of time that a person has lived in age groups."@en ;
  rdfs:label "Leeftijdsgroep"@nl ;
  rdfs:range skos:Concept ;
.
statsvl:educationLev
  rdf:type qb:DimensionProperty ;
  rdf:type rdf:Property ;
  qb:codeList <http://id.vlaanderen.be/statistieken/conceptscheme/scholingsgraden#id> ;
  qb:concept sdmx-concept:educationLev ;
  rdfs:comment "The highest level of an educational programme the person has successfully completed."@en ;
  rdfs:label "Opleidingsniveau"@nl ;
  rdfs:range skos:Concept ;
  rdfs:subPropertyOf sdmx-concept:educationLev ;
.
statsvl:jobSeekers
  rdf:type qb:MeasureProperty ;
  rdf:type rdf:Property ;
  rdfs:comment "Total amount"@en ;
  rdfs:label "aantal niet-werkende werkzoekenden"@nl ;
  rdfs:range xsd:integer ;
  rdfs:subPropertyOf sdmx-measure:obsValue ;
.
statsvl:refArea
  rdf:type qb:DimensionProperty ;
  rdf:type rdf:Property ;
  qb:codeList <http://id.fedstats.be/conceptscheme/nis#id> ;
  qb:concept sdmx-concept:refArea ;
  rdfs:comment "The country or geographic area to which the measured statistical phenomenon relates."@en ;
  rdfs:label "Referentie Area"@nl ;
  rdfs:range skos:Concept ;
  rdfs:subPropertyOf sdmx-dimension:refArea ;
.
statsvl:timePeriod
  rdf:type qb:DimensionProperty ;
  rdf:type rdf:Property ;
  qb:codeList <http://id.vlaanderen.be/statistieken/conceptscheme/jaren#id> ;
  qb:concept sdmx-concept:timePeriod ;
  rdfs:comment "The period of time or point in time to which the measured observation refers."@en ;
  rdfs:label "Periode in de tijd"@nl ;
  rdfs:range skos:Concept ;
  rdfs:subPropertyOf sdmx-dimension:timePeriod ;
.
```

qb:Observation

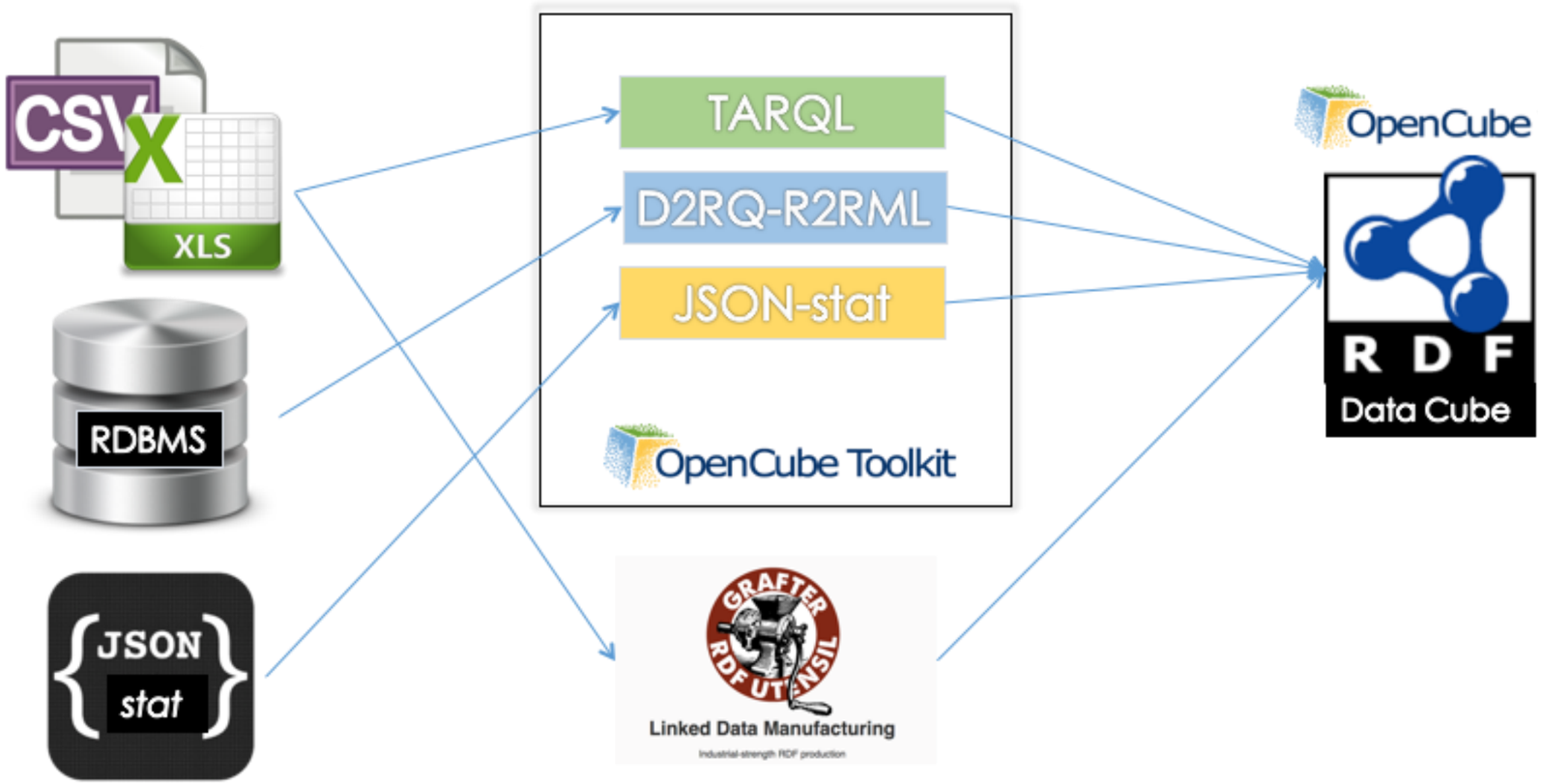
```
<http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz/observatie/0#id>
  rdf:type qb:Observation ;
  statsvl:ageGroup <http://id.vlaanderen.be/statistieken/concept/leeftijdsgroep_00-24#id> ;
  statsvl:educationLev <http://id.vlaanderen.be/statistieken/concept/scholingsgraad_laag#id> ;
  statsvl:jobSeekers "10"^^xsd:int ;
  statsvl:refArea <http://id.fedstats.be/nis/11001#id> ;
  statsvl:timePeriod <http://id.vlaanderen.be/statistieken/concept/jaar_1999#id> ;
  qb:dataSet <http://id.vlaanderen.be/statistieken/dq/kubus-studieniveau-nwwz#id> ;
  sdmx-dimension:sex sdmx-code:sex-M ;
.
```

Conversion

project.clj × properties-cube-3.clj ×

```
{
; This is a property file setting all required configurations to produce an RDF data cube w.

; Set the number of sheets to be loaded from a workbook and noisy headers irrelevant to obs
; Value: a vector of integers
; Example: To load the first 2 sheets and remove the first row from each of them, please wr.
 :noisy-headers [1 1]

; Set the column names for a loaded dataset.
; Value: a vector of strings
; Example: ["a" "b" "c"]
 :column-names ["city-id-name" "time-period" "age" "sex" "education-level" "number"]

; Set the columns that requires to be filled down with values when a value spans multiple r
; Value: a vector of column names contained in the aforementioned column name definition.
; Example: ["a" "b"]
 :columns-to-fill-down ["city-id-name" "time-period" "age" "sex"]

; Set the columns that requires to be filled with given values when there are empty cells i
; Value: a map with column names as keys and corresponding values to be filled in the colum
; Example: {"a" 0 "b" "zero"}
 :columns-to-fill-value {}

; Set the columns that MUST BE non-blank. All the rows with empty cells at the given column
; Value: a vector of column names
; Example: ["c"]
 :non-blank-columns ["number"]

; Set the columns that require type conversion. For example, in order to manipulate a time
; Value: a vector of maps of column names and corresponding type conversion functions to be
; Example: [{:col-name "a" :func "str"} {:col-name "b" :func "int"} {:col-name "c" :func "f
 :columns-type-cast
 [
  {:col-name "time-period"
   :func "str"}
  {:col-name "number"
   :func "int"}
 ]
```

OpenCube Publishing



sdmx
Statistical Data and Metadata eXchange

http://csarven.ca/linked-sdmx-data

TARQL

contact.csv

4 columns, **500+** rows

11.95 MB, Comma-Separated, UTF-8

EntityNumber	EntityContact	ContactType	Value
0200.362.111	ENT	TEL	067 28 01 11
0200.362.111	ENT	EMAIL	info@iecbw.be
0200.362.111	ENT	WEB	www.iecbw.be/
0200.362.408	ENT	TEL	081 61 42 48
0200.362.408	ENT	EMAIL	secretariat@isbw.be
0201.310.929	ENT	TEL	089/32.39.50
0201.310.929	ENT	EMAIL	info@iglimburg.be
0201.543.234	ENT	TEL	071 44 00 40
0201.543.234	ENT	EMAIL	info@icdi.be
0201.645.281	ENT	TEL	071 20 28 11
0201.645.281	ENT	EMAIL	info@igretec.com
0202.395.052	ENT	TEL	04/367 84 11
0202.395.052	ENT	EMAIL	info@cile.be
0203.201.340	ENT	TEL	02 221 21 11

contact.sparql

contact.sparql

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX rov: <http://www.w3.org/ns/regorg#>
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema: <http://schema.org/>
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>

CONSTRUCT {
?URI ?prop ?Value.
}
FROM <file:contact.csv>
WHERE {
    BIND (REPLACE(ENCODE_FOR_URI(?EntityNumber),'\\.','_') as ?kort)
    BIND (URI(concat('http://data.kbodata.be/organisation/',?kort,'#id')) as ?URI)
    BIND (IF(?ContactType='TEL',vcard:hasTelephone,IF(?ContactType='EMAIL',vcard:hasEmail,vcard:hasURL)) as ?prop)
}
OFFSET 1
```

Constraint checking

IC-1. Unique DataSet

Every `qb:Observation` has exactly one associated `qb:DataSet`.

```
ASK {
  {
    # Check observation has a data set
    ?obs a qb:Observation .
    FILTER NOT EXISTS { ?obs qb:dataSet ?dataset1 . }
  } UNION {
    # Check has just one data set
    ?obs a qb:Observation ;
       qb:dataSet ?dataset1, ?dataset2 .
    FILTER (?dataset1 != ?dataset2)
  }
}
```

IC-2. Unique DSD

Every `qb:DataSet` has exactly one associated `qb:DataStructureDefinition`.

```
ASK {
  {
    # Check dataset has a dsd
    ?dataset a qb:DataSet .
    FILTER NOT EXISTS { ?dataset qb:structure ?dsd . }
  } UNION {
    # Check has just one dsd
    ?dataset a qb:DataSet ;
       qb:structure ?dsd1, ?dsd2 .
    FILTER (?dsd1 != ?dsd2)
  }
}
```

IC-3. DSD includes measure

Every `qb:DataStructureDefinition` must include at least one declared measur

```
ASK {
  ?dsd a qb:DataStructureDefinition .
  FILTER NOT EXISTS { ?dsd qb:component [qb:componentProperty [a qb
}
```

branch: **master** ▾ **NoSPA-RDF-Data-Cube-Validator** / +

Deleted jar files from repository, uploaded them to the release page

yyz1989 authored on Jan 27 — latest commit `b48b3ae981`

src/main	Fixed a small sign bug, re-evaluated the performance	3 months ago
LICENSE	Update LICENSE	6 months ago
README.md	Deleted jar files from repository, uploaded them to the release page	3 months ago
largeTest.ttl	A new release with packaged jar files	3 months ago
pom.xml	A new release with packaged jar files	3 months ago
test.ttl	Created property file to deal with file paths	6 months ago

README.md

NoSPA RDF Data Cube Validator

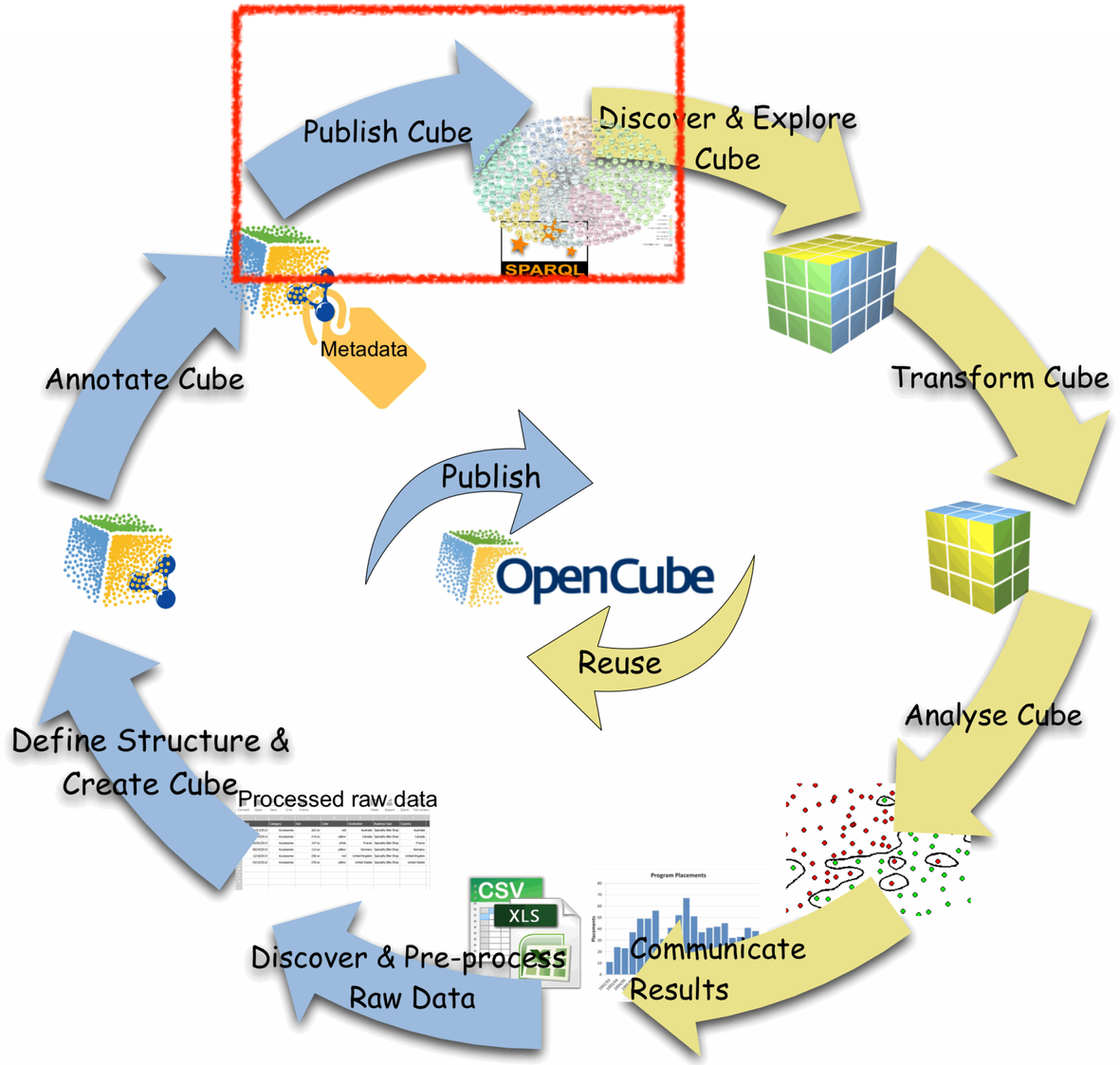
Introduction

This is an RDF Data Cube Validator. Its significant difference from other existing validators is that it is not based on SPARQL queries, as its name “NoSPA”. Jena library is used to manipulate RDF models. The official SPARQL queries for constraint checks are interpreted and parsed by this validator to search functions with nested statement listing functions provided by Jena and filters for different conditions. It has an outstanding performance because the entire process is executed in memory. I believe that it is valuable to sacrifice some memory for saving time.

Here are some references and knowledge background for this tool:

- The official RDF data cube spec: The RDF Data Cube Vocabulary
- Jena API: Apache Jena
- The official SPARQL spec: SPARQL 1.1 Query Language

Publish


Publish Cube
Discover & Explore Cube
SPARQL
Metadata
Annotate Cube
Transform Cube
Publish
OpenCube
Reuse
Define Structure & Create Cube
Analyse Cube
Processed raw data
CSV
XLS
Discover & Pre-process Raw Data
Communicate Results

RShiny - Cubes

Werkzoekenden (NWWZ) volgens studieniveau

niveau referentiegebied

Gemeente

gebied(en)

Leuven Mechelen

geslacht

Vrouwelijk

leeftijdsgroep

00–24

studieniveau

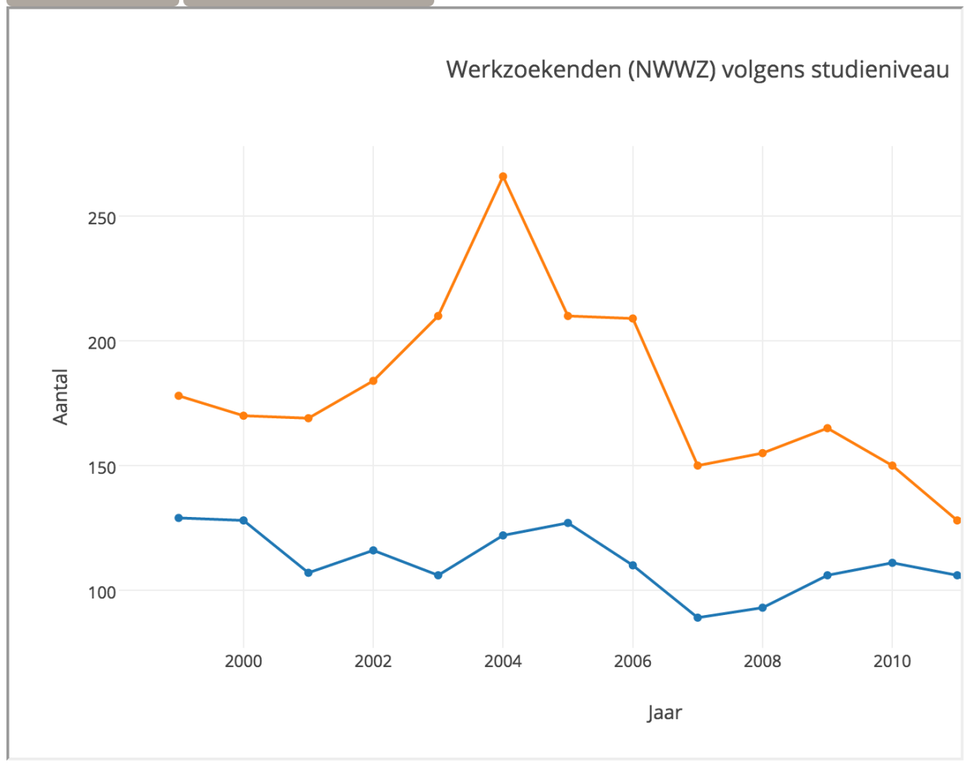
Laaggeschoold

Query

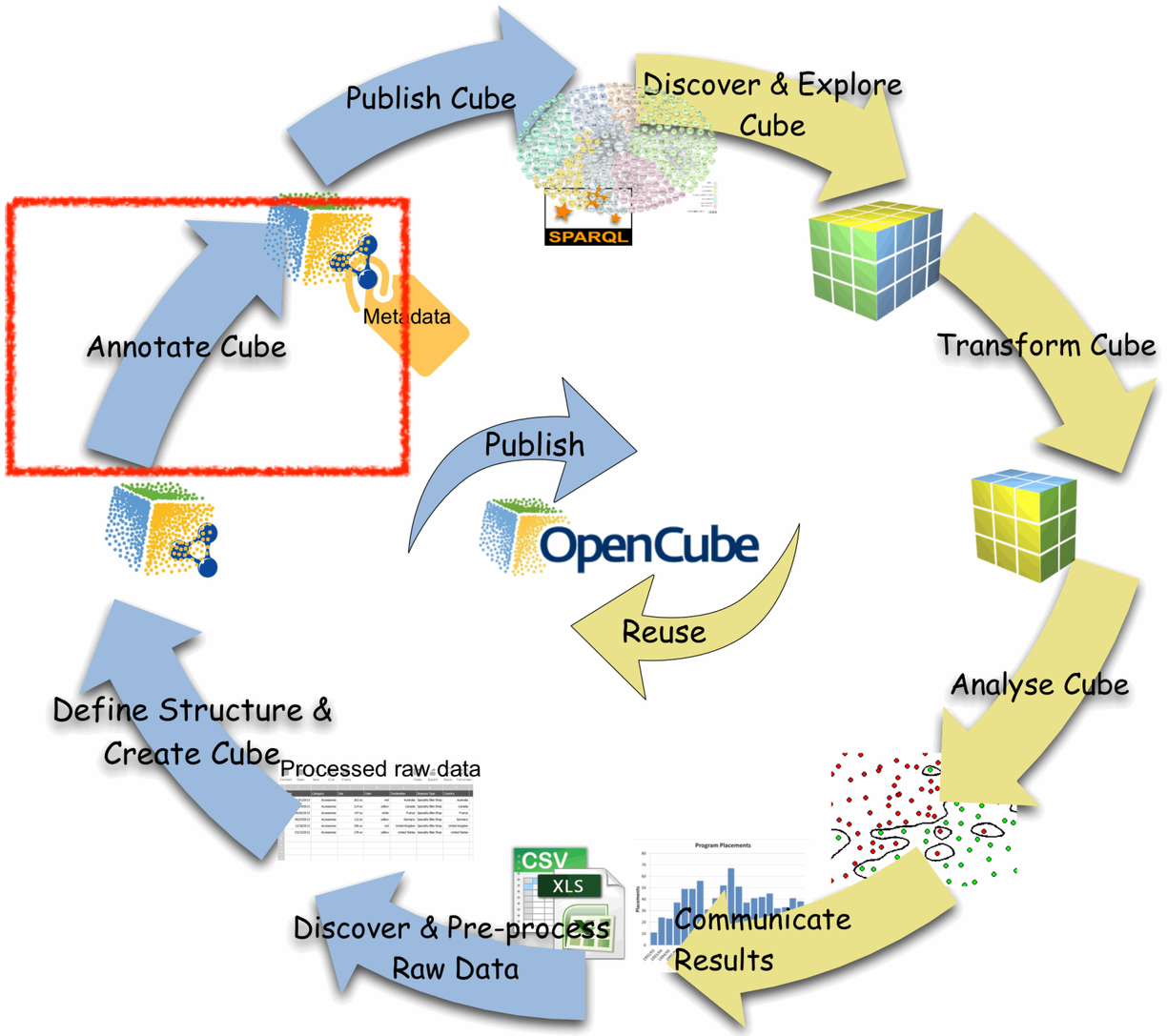
selecteer type grafiek

lijngrafiek (aantal)

Toon ruwe data | Toon SPARQL resultaten



Annotate

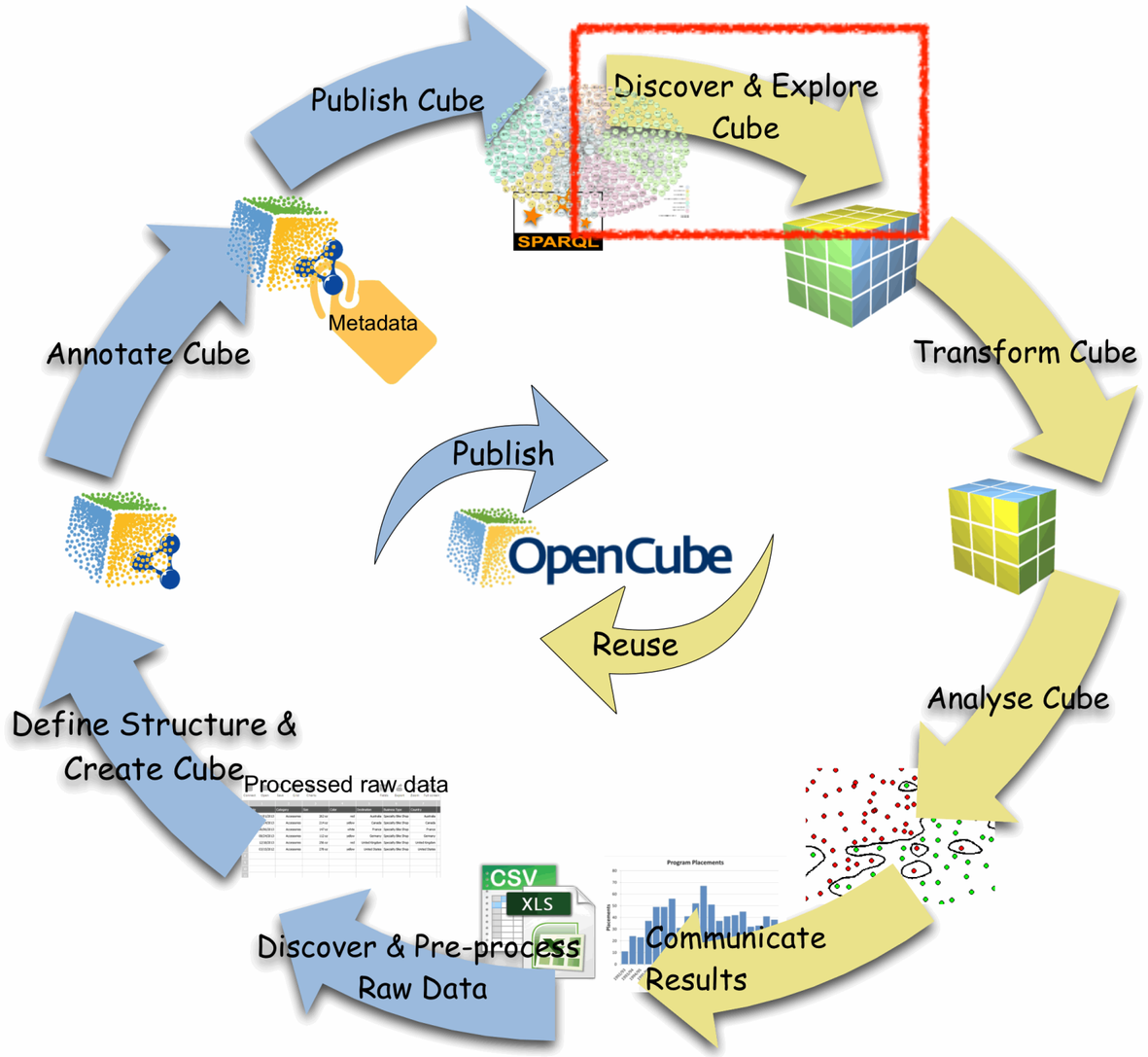

Publish Cube
Discover & Explore Cube
SPARQL
Transform Cube
Analyse Cube
Communicate Results
Discover & Pre-process Raw Data
CSV
XLS
Define Structure & Create Cube
Processed raw data
Annotate Cube
Metadata
Publish
OpenCube
Reuse

OpenCube Compatibility Explorer

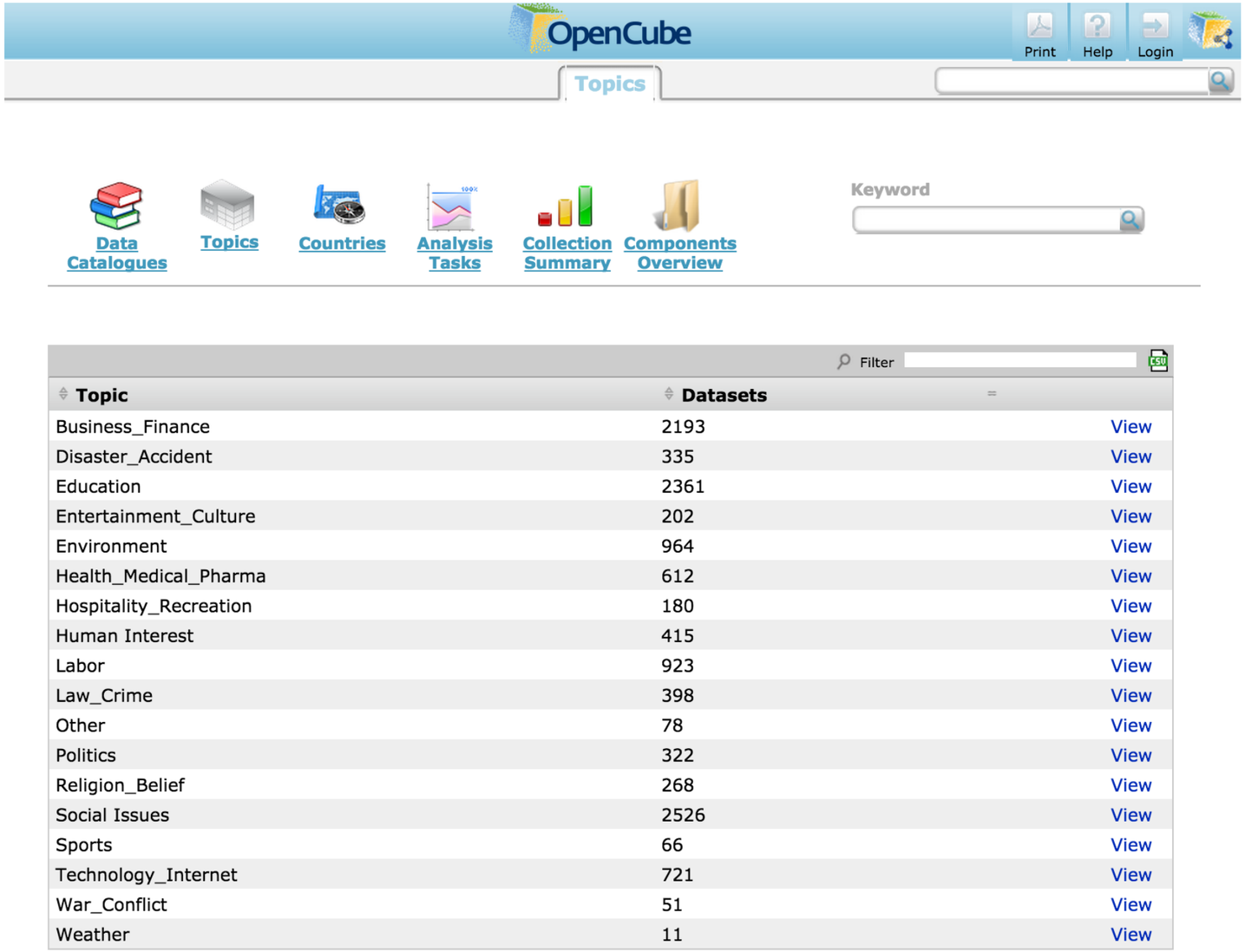
Add measure.
An expansion cube is capable to add a new measure to an original cube if: i) both cubes have the same dimensions, ii) the expansion cube has at least the same values at each dimension of the original cube (it may contain more values than the original cube) and iii) the expansion cube has at least one measure that does not exist at the original cube.
Add value to dimension.
An expansion cubes is compatible to add a new value to a dimension of an original cube if: i) both cubes have the same dimensions, ii) both cubes have the same measures and iii) the expansion cube has at least one value more than the original cube at the expansion dimension and has the same values with the original cube at all the rest dimensions.

OpenCube Consuming

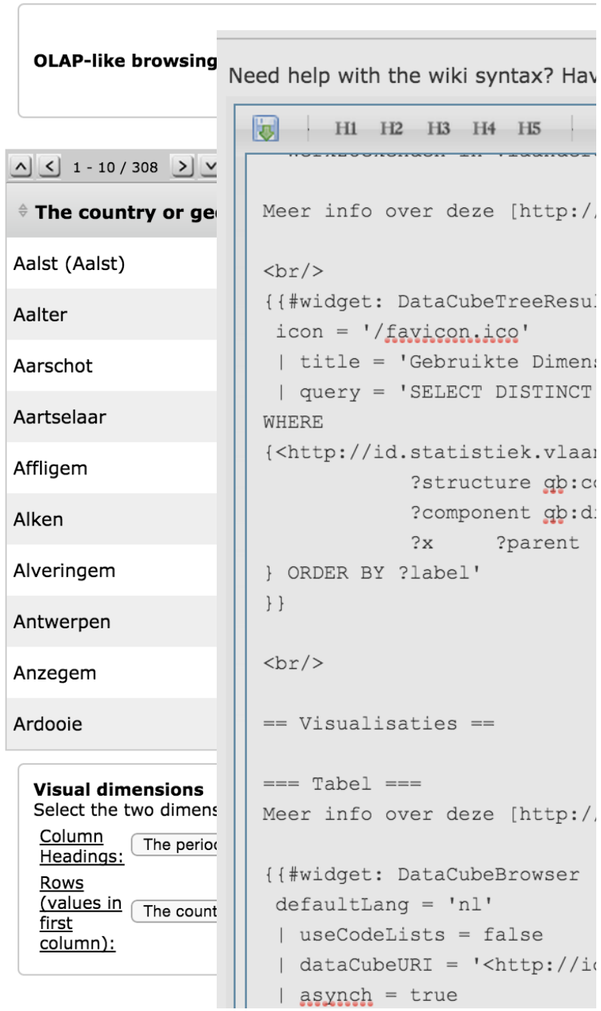
Discover and explore



Data Catalogue management solution



OpenCube Browser



http://json-stat.org/samples/galicia.json

connect

Population by province of residence, place of birth, age, gender, and year in Galicia

Rows & Columns

Age group ↔ Place of birth

Filters

Total Gender

2001 Year

Ourense Province of residence

Constants

Population (persons) Concept

	Total	County of residence	Another county in the same province	Another province of Galicia	In another autonomous community	Abroad
Total	338,446	202,971	78,981	17,219	16,110	23,165
0–4	9,364	7,339	1,118	242	286	379
5–9	11,161	8,039	1,512	324	415	871
10–14	13,668	9,239	2,023	390	597	1,419
15–19	17,447	11,697	2,614	452	849	1,835
20–24	22,348	14,033	3,400	815	1,460	2,640
25–29	22,813	12,656	3,837	934	1,573	3,813
30–34	22,117	11,169	4,956	1,099	1,580	3,313
35–39	21,536	10,068	5,995	1,351	1,590	2,532
40–44	22,300	10,666	6,918	1,578	1,475	1,663
45–49	20,194	10,175	6,227	1,603	1,208	981
50–	20,358	10,901	6,242	1,529	1,056	630

OpenCube OLAP Browser new gen.

Please select a cube to visualize:
Kubus studieniveau nwwz

Language
Select the language of the visualized data:
nl

Dimensions
- [ ] The country or geographic area to which the measured statistical phenomenon relates.
- [x] The period of time or point in time to which the measured observation refers.
- [ ] The length of time that a person has lived or a thing has existed.
- [x] The state of being male or female.
- [ ] The highest level of an educational programme the person has successfully completed.

Measures
- [x] Total amount

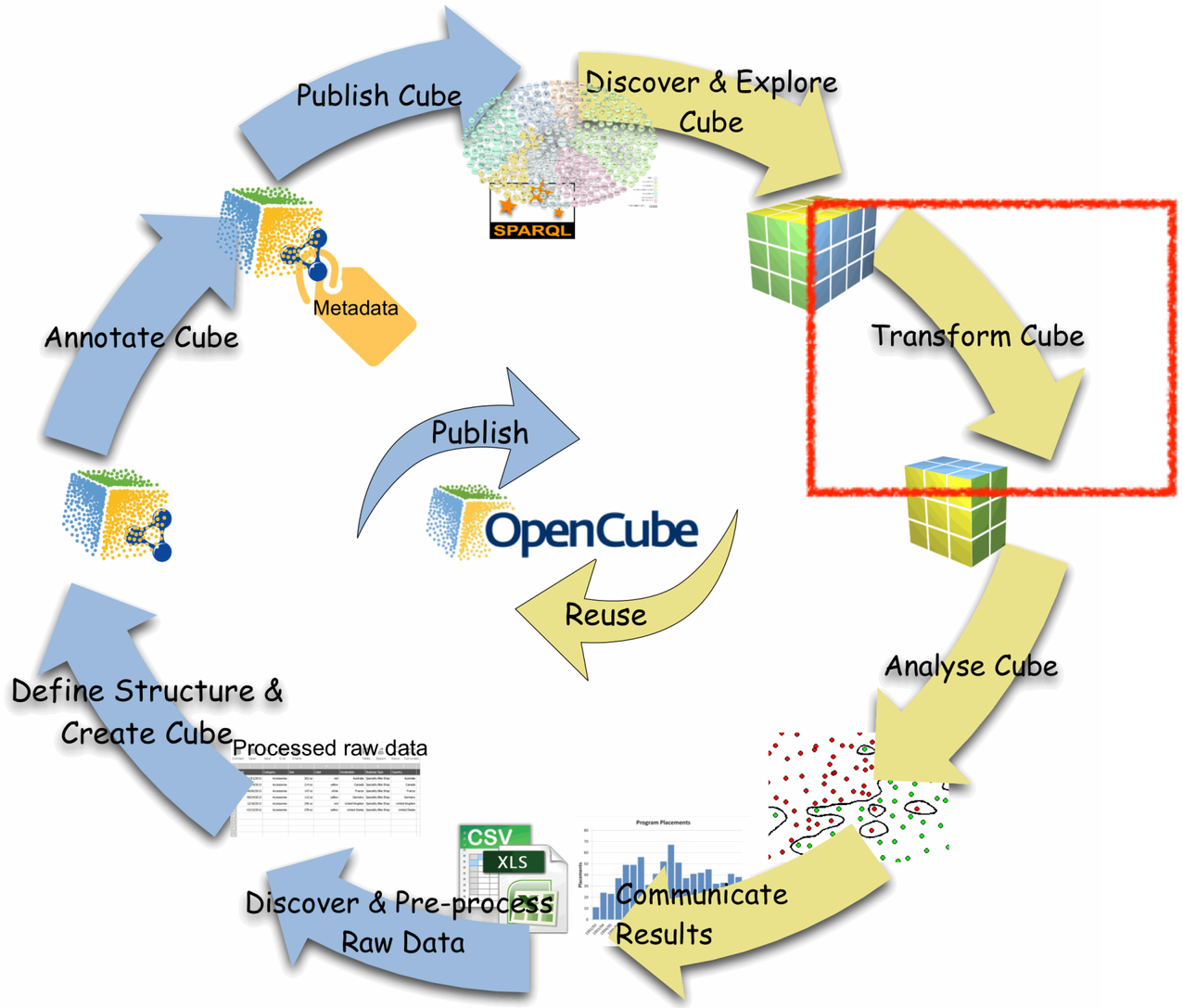
Columns: The period of time or point in time…
Rows: The country or geographic area to w…

Filter:
- The length of time that a person has lived or a thing has existed.: 50+
- The state of being male or female.: Mannelijk
- The highest level of an educational programme the person has successfully completed.: hooggeschoold

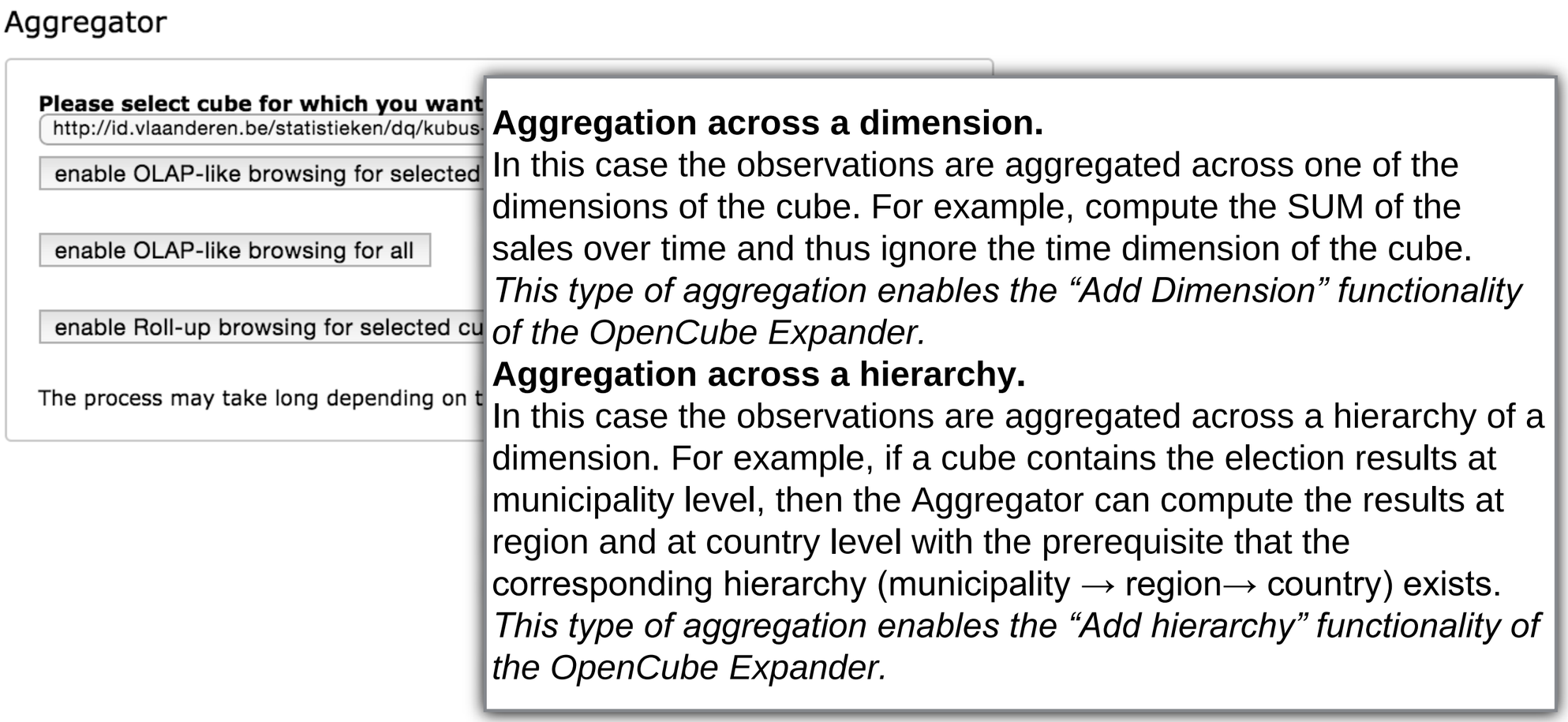
1 - 10 / 308 Show 10 rows (max. 1000) Filter

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
23	36	46	41	51	41	50	55	55	51	57	59
7	2	6	5	7	6	8	11	11	16	12	16
4	5	5	5	11	12	14	16	17	12	14	23
7	4	7	10	9	12	15	10	11	12	13	14
2	2	8	7	4	2	1	4	1	8	5	7
2	0	0	2	1	2	2	3	3	1	5	7
0	0	0	0	1	2	2	5	4	3	4	5
168	236	315	365	362	333	436	462	486	470	529	593
1	2	3	3	5	4	5	4	5	7	3	4
1	2	2	2	0	0	2	2	2	1	1	2

Transform


Publish Cube
Discover & Explore Cube
SPARQL
Metadata
Annotate Cube
Transform Cube
Publish
OpenCube
Reuse
Define Structure & Create Cube
Processed raw data
Analyse Cube
CSV
XLS
Discover & Pre-process Raw Data
Communicate Results

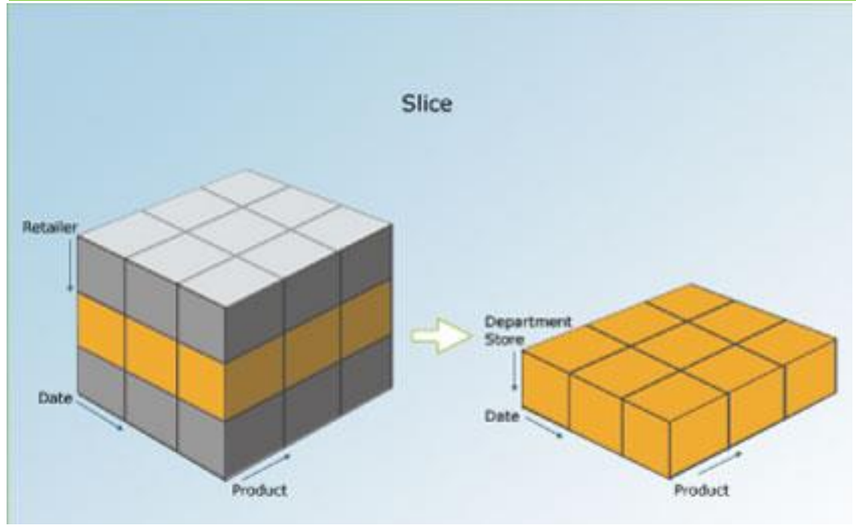
OpenCube Aggregator



Aggregation across a dimension.
In this case the observations are aggregated across one of the dimensions of the cube. For example, compute the SUM of the sales over time and thus ignore the time dimension of the cube.
This type of aggregation enables the "Add Dimension" functionality of the OpenCube Expander.

Aggregation across a hierarchy.
In this case the observations are aggregated across a hierarchy of a dimension. For example, if a cube contains the election results at municipality level, then the Aggregator can compute the results at region and at country level with the prerequisite that the corresponding hierarchy (municipality → region→ country) exists.
This type of aggregation enables the "Add hierarchy" functionality of the OpenCube Expander.

OpenCube Slicing component



OpenCube Slicer

Please select the dimensions to fix for the slice:

- [ ] The highest level of an educational programme the person has successfully completed.
- [ ] The state of being male or female.
- [x] The length of time that a person has lived or a thing has existed.
- [x] The country or geographic area to which the measured statistical phenomenon relates.
- [ ] The period of time or point in time to which the measured observation refers.

Please select the values of the fixed dimensions for the slice:

The country or geographic area to which the measured statistical phenomenon relates.

Leuven

The length of time that a person has lived or a thing has existed.

50+

createSlice

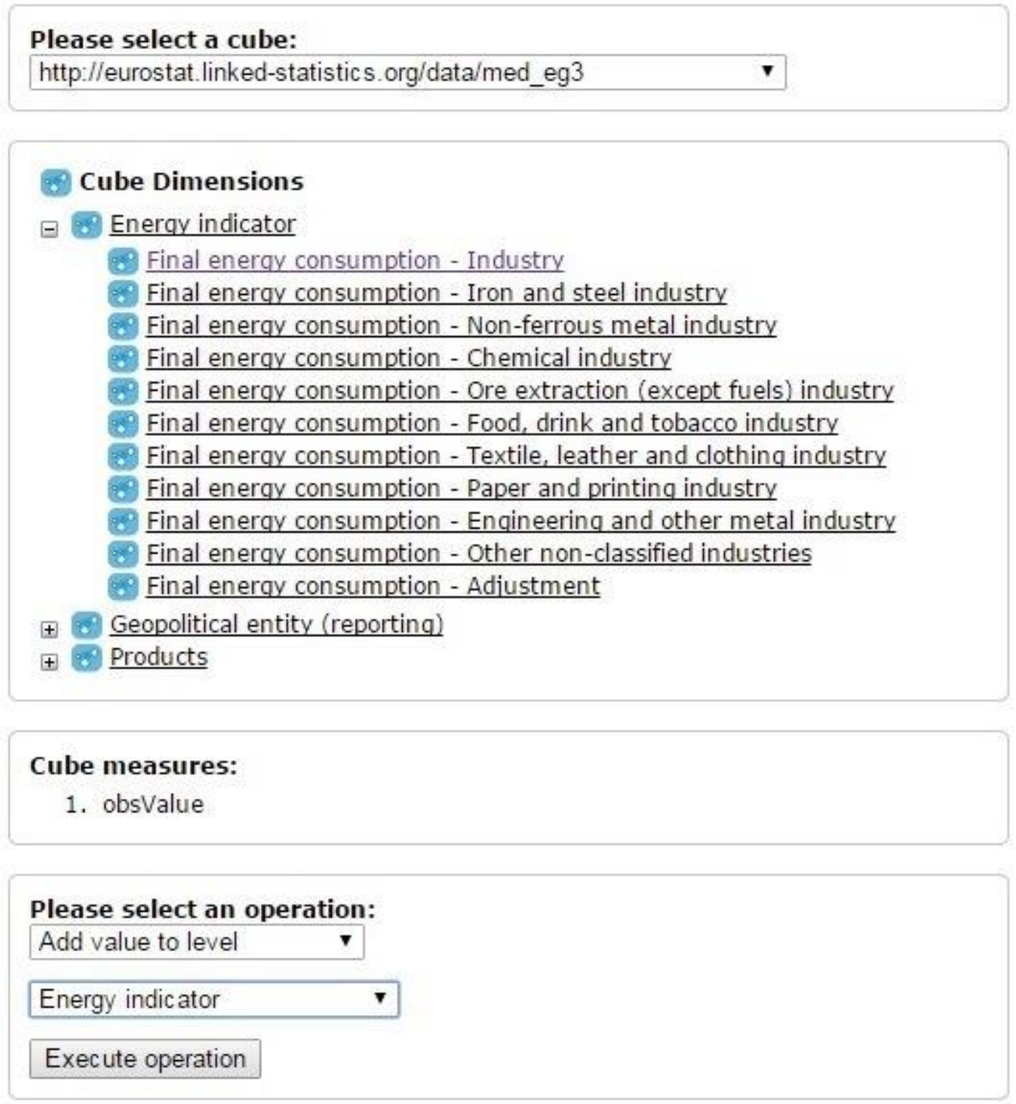
OpenCube Expander

Example: Start with an initial cube

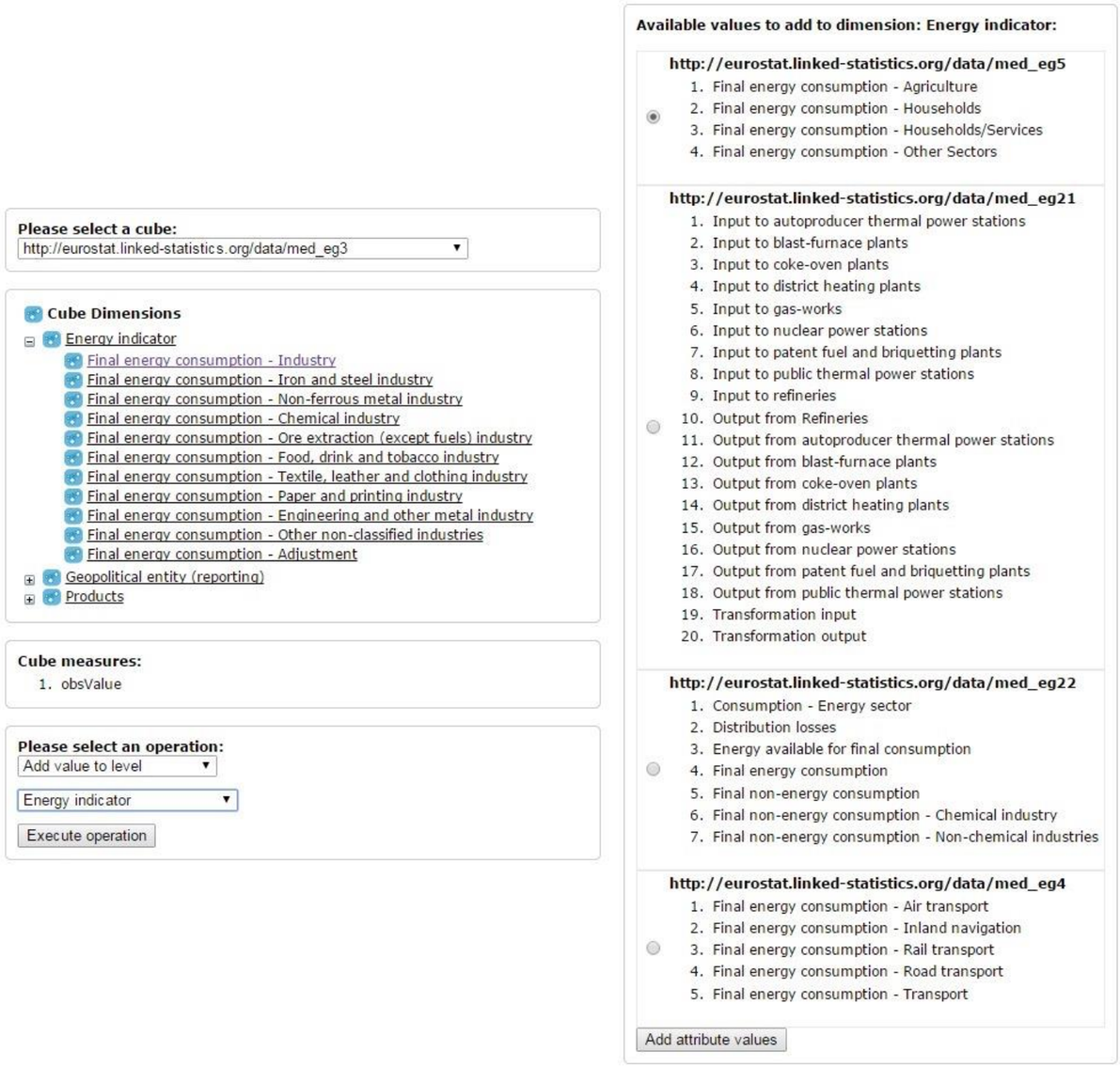
Dimensions

Measures

Operations



Example: Discover & Select compatible cubes



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

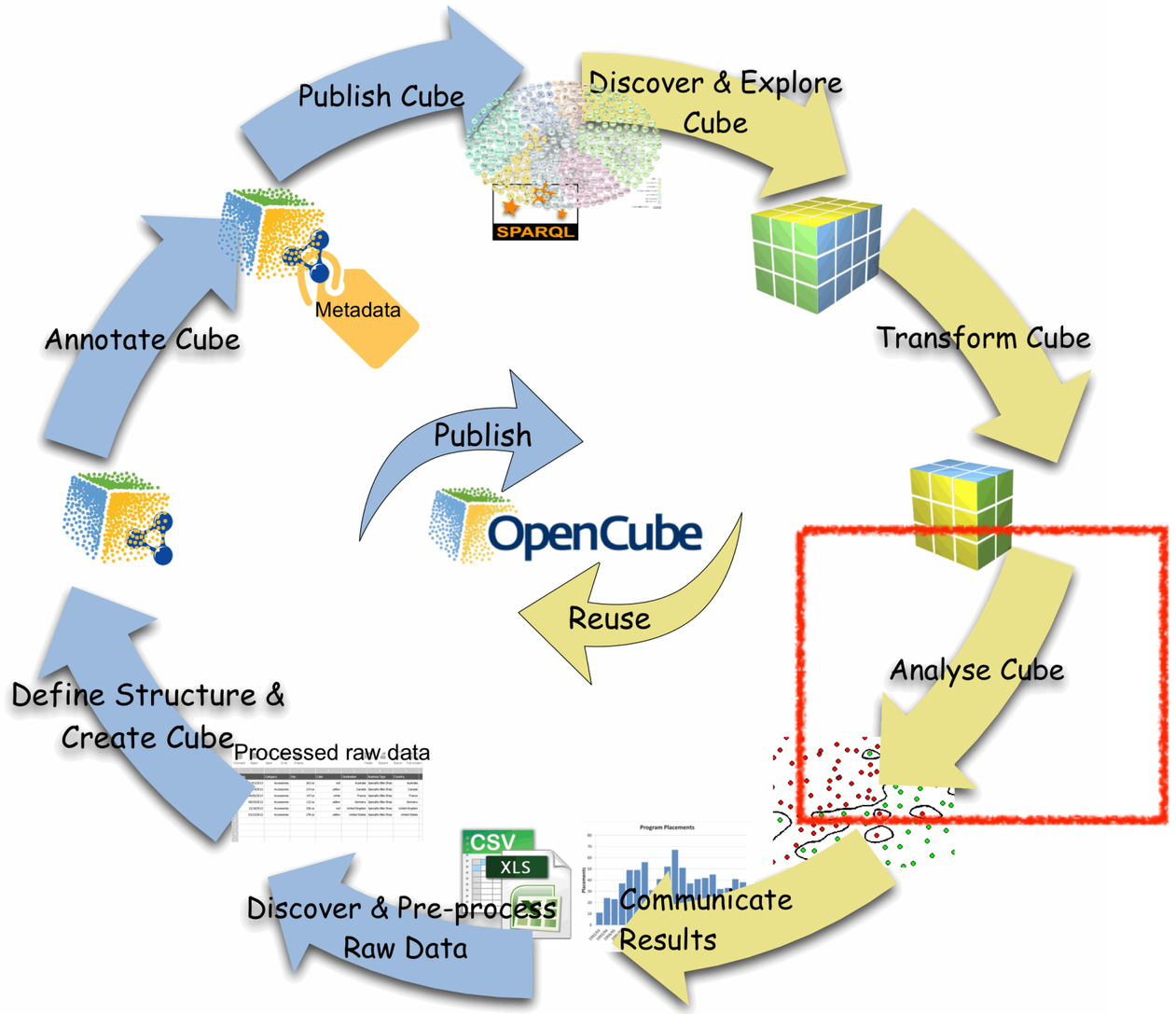
Measures
Select the measures to visualize:
- ☑ obsValue

Language
Select the language of the visualized data:
en ▾

Filter

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

Analyse


Publish Cube
Discover & Explore Cube
SPARQL
Metadata
Annotate Cube
Transform Cube
Publish
OpenCube
Reuse
Analyse Cube
Define Structure & Create Cube
Processed raw data
CSV
XLS
Discover & Pre-process Raw Data
Communicate Results

R statistical analysis support

Input parameters

Here you can add input data to be used in your R script. Each input parameter can be either a table (data frame object in R) or a single value. To pass a table, you can select one of the time ser

Note: for each input parameter only one of the two settings (*Query* or *Value*) should be provided.

Data Frame (columns: *date*, *value*)

R Variable	hicp	
Query	fr_hicp	
Value		

Delete

Data Frame (columns: *date*, *value*)

R Variable	gdp	
Query	frgdp	
Value		

Delete

Add input data | Create new time series

Script

Here you can specify the R script which will perform your analysis. In your script, you can:

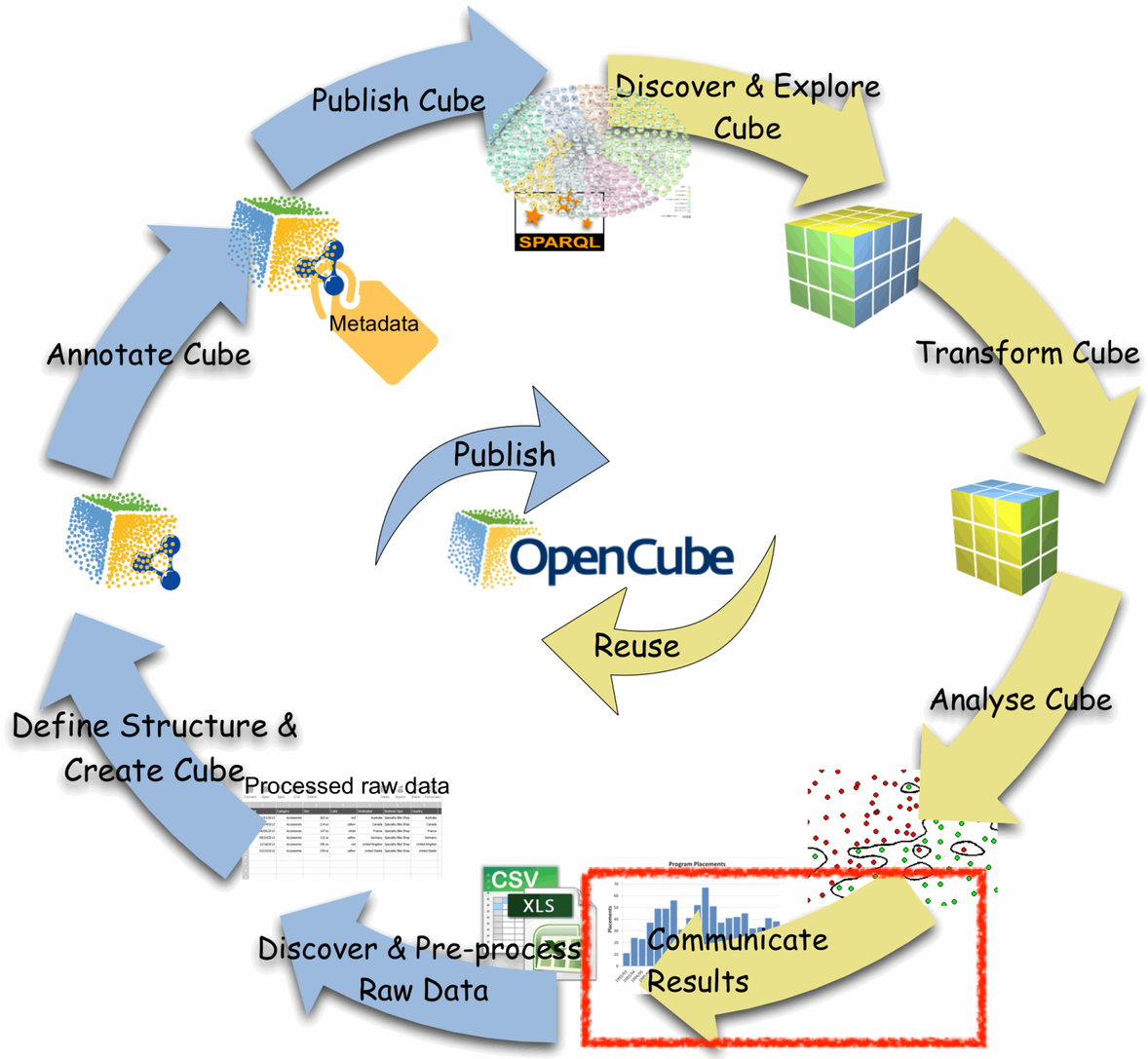
- Draw an R chart and visualize it (by pressing **Show as a chart**)
- Generate a data frame and show it in a table (by pressing **Show in a table**).

Script

```
hicp$date <- as.character(hicp$date);

gdp$date <- as.character(gdp$date);

# converting the HICP input values to numeric
hicp$value <- as.numeric(hicp$value);

# converting the GDP input values to numeric
gdp$value <- as.numeric(gdp$value);

# Since hicp values are provided monthly, we need to preserve only the rows for January each year.
hicp <- hicp[ substr(hicp$date, nchar(hicp$date)-5, nchar(hicp$date)) == '-01-01', ];

# Calculate inflation rate and add to the HICP table
hicp$inflation <- c(NA, diff(log(hicp$value)));

# Calculate GDP growth rate and add to the GDP table
gdp$growth <- c(NA, diff(log(gdp$value)));

# Merge 2 tables by common date
merged <- merge(hicp, gdp, by = "date");

# Calculate the sum of inflation and growth rate
merged$adjustedRate <- merged$growth + merged$inflation;

# Form resulting table
result <- merged[, c("date", "adjustedRate")];

result;
```

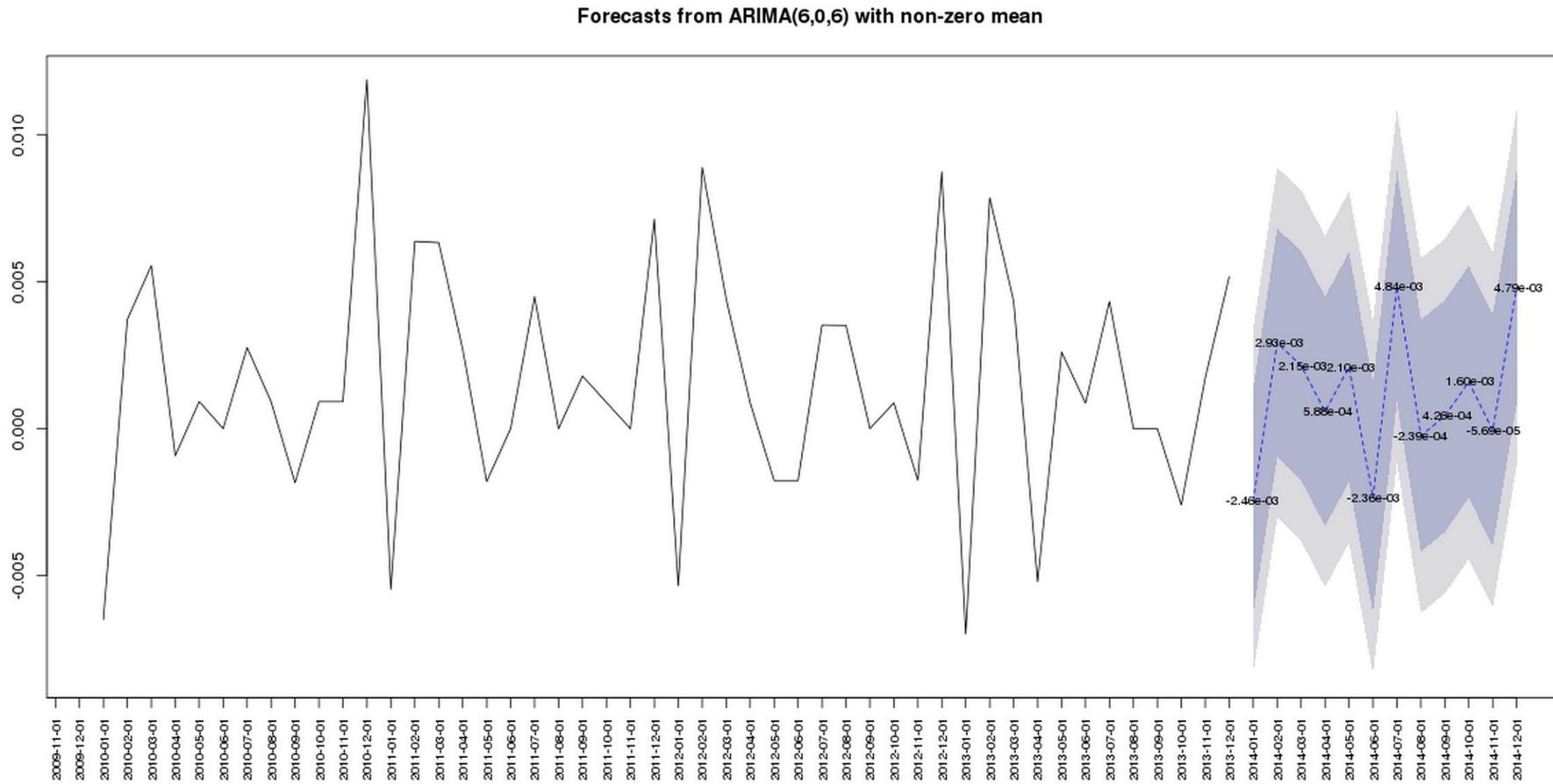
Show as a chart | Show in a table | Delete analysis

Communicate results



R statistical analysis support

RChartView



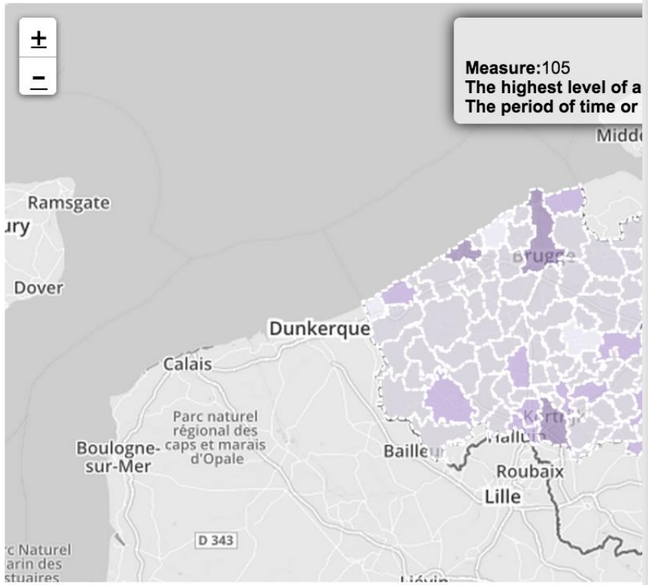
OpenCube Map View

Kaart

Meer info over deze MapView component.

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

OLAP-like browsing



Need help with the wiki syntax? Have a look here.

```
            ?x      ?parent   ?child.
} ORDER BY ?label'
}}

<br/>

== Visualisaties ==

=== Tabel ===
Meer info over deze

{{#widget: DataCube
 defaultLang = 'nl'
 | useCodeLists = f
 | dataCubeURI = '<
 | asynch = true
}}
<br/>
=== Grafiek ===

Deze grafiek toont
De gegevens (de pro

<br/>

{{#widget: DataCube
 title = 'Test'
 | dataCubeConfigu
   {{ datasetURI =
    | datasetViewConfig =
      {{ xDimensionURI = 'sdmx-dimension:timePeriod'
```

Add widget

Widget *	DataCubeBrowser
Data Cube URI *	<http://id.statistiek.vlaanderen.be/da
Use Code Lists	false
Sparql Service	
Default Lang	nl
Ignore Lang	
Asynchronously	true
Width	
Height	

fields with a * are required

Submit

Fixed dimensions
Change the values of the fixed dimensions:

The highest level of an educational programme the person has successfully completed- hooggeschoold

The period of time or point in time to which the measured observation refers- 2013-01-01

PublishMyData

Swirrl

PublishMyData
Linked Data Platform

Case Studies
Our work

About Us
Who we are

Our Blog
News and thoughts

PUBLISH MY DATA

PublishMyData is our **Linked Data** publishing platform. It lets you serve your **5-star** Open Data on the web in a format that's easy to understand, but it's also machine readable so data experts can exploit it.

With PublishMyData your organisation's data can be used to far greater effect: linking with other, related data and reaching a wider audience who, in turn, can manipulate it for themselves.

VIEW PRICING AND BUY

Q&A