Future work and applications

RDF Validation tutorial

Jose Emilio Labra Gayo
WESO Research group
University of Oviedo, Spain

Eric Prud'hommeaux
World Wide Web Consortium
MIT, Cambridge, MA, USA

Harold Solbrig
Mayo Clinic, USA

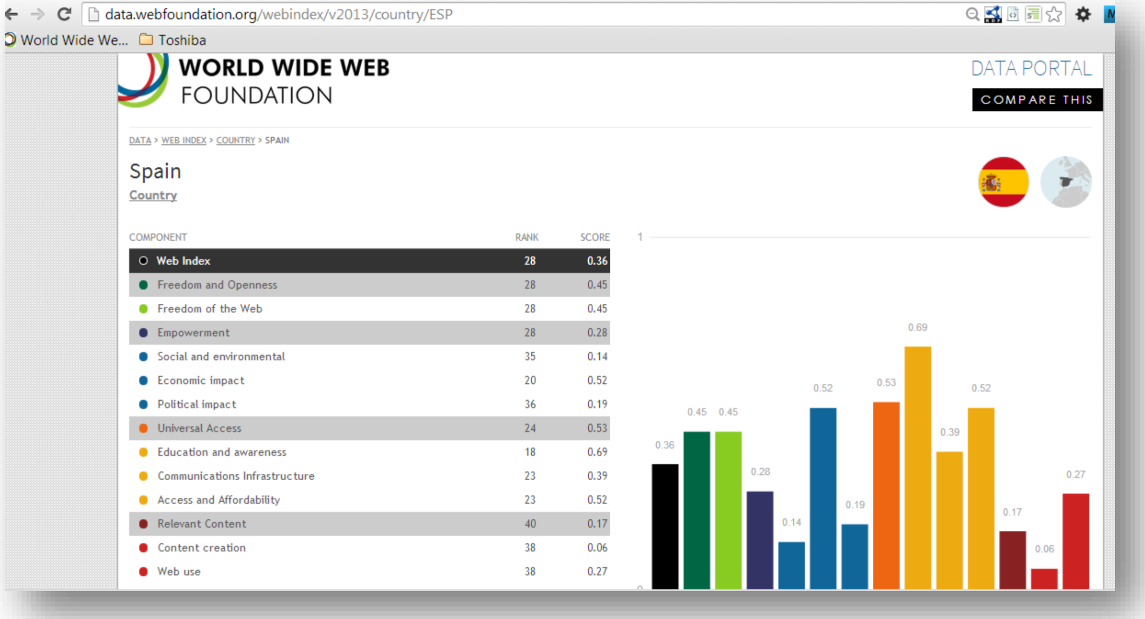
Iovka Boneva
LINKS, INRIA & CNRS
University of Lille, France

Contents

- Some applications
  - ShEx
    - WebIndex: A linked data portal using ShEx
    - FHIR
  - SHACL
- Future work

Web Index

- Measure WWW's contribution to development and human rights by country
  - Developed by the Web Foundation
  - 81 countries, 116 indicators, 5 years (2007-12)
- Linked data portal
  - http://data.webfoundation.org/webindex/2013



data.webfoundation.org/webindex/v2013/country/ESP

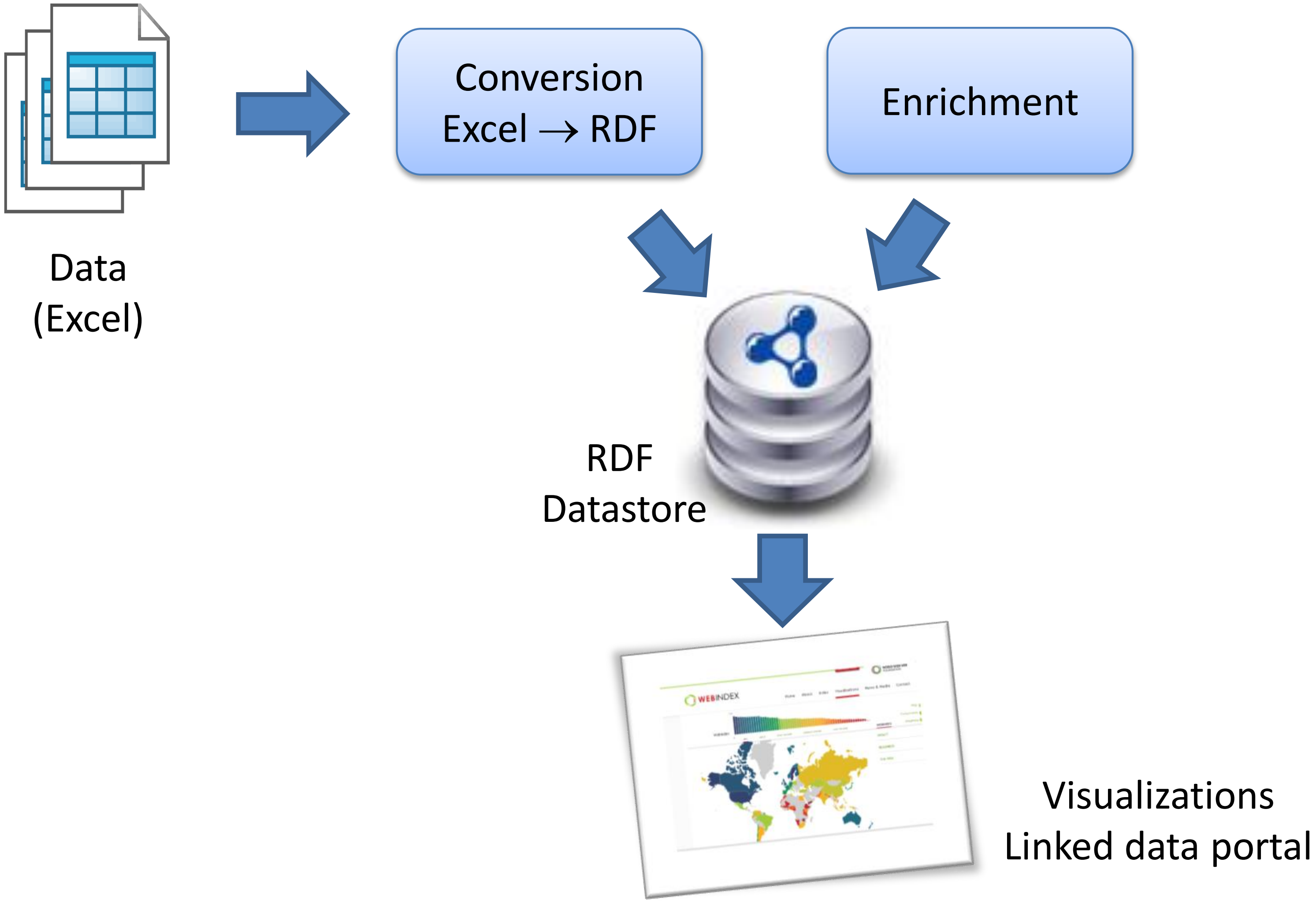
DATA > WEB INDEX > COUNTRY > SPAIN

Spain

Country

COMPONENT	RANK	SCORE
○ Web Index	28	0.36
● Freedom and Openness	28	0.45
● Freedom of the Web	28	0.45
● Empowerment	28	0.28
● Social and environmental	35	0.14
● Economic impact	20	0.52
● Political impact	36	0.19
● Universal Access	24	0.53
● Education and awareness	18	0.69
● Communications Infrastructure	23	0.39
● Access and Affordability	23	0.52
● Relevant Content	40	0.17
● Content creation	38	0.06
● Web use	38	0.27

Webindex workflow



Data
(Excel)

Conversion
Excel → RDF

Enrichment

RDF
Datastore

Visualizations
Linked data portal

WebIndex data model

Model based on RDF Data Cube

Main entity = Observation

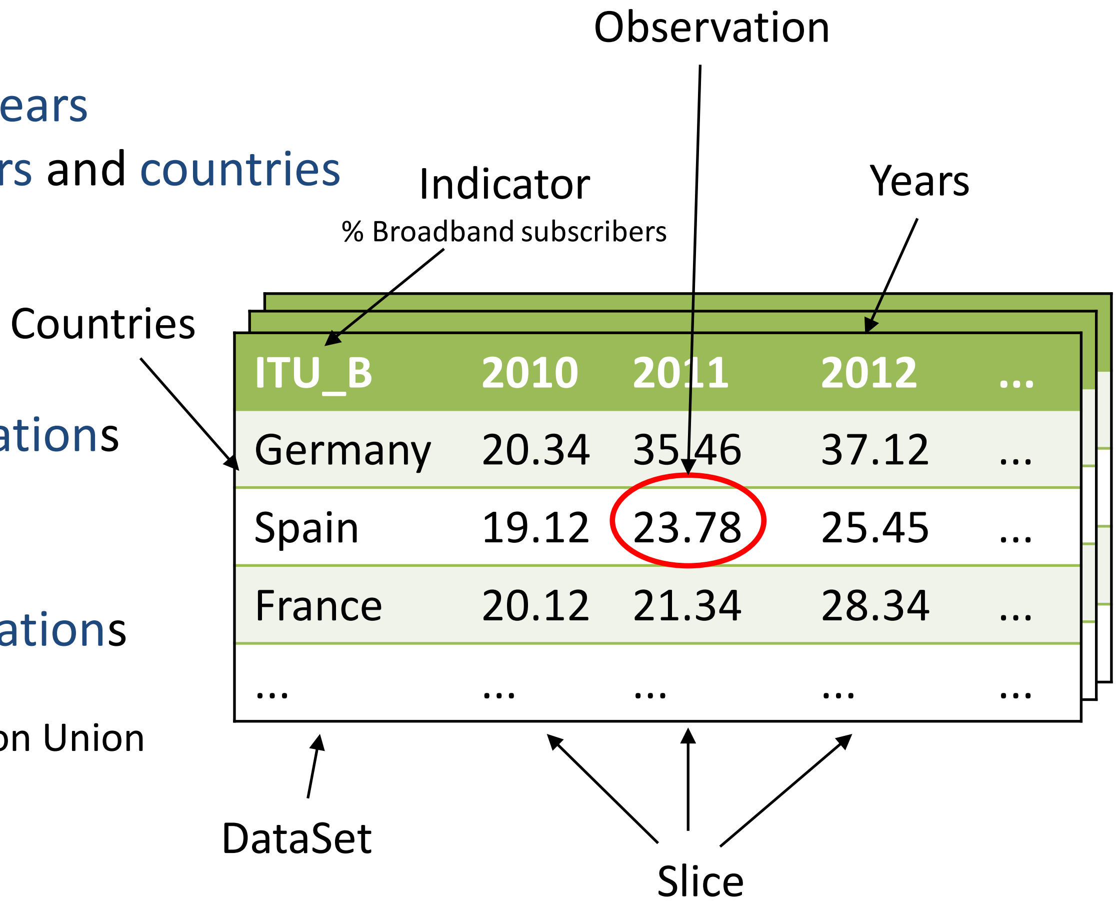
- Observations have values by years
- Observations refer to indicators and countries

DataSets are published by Organizations

- Datasets contain several slices
- Slices group observations

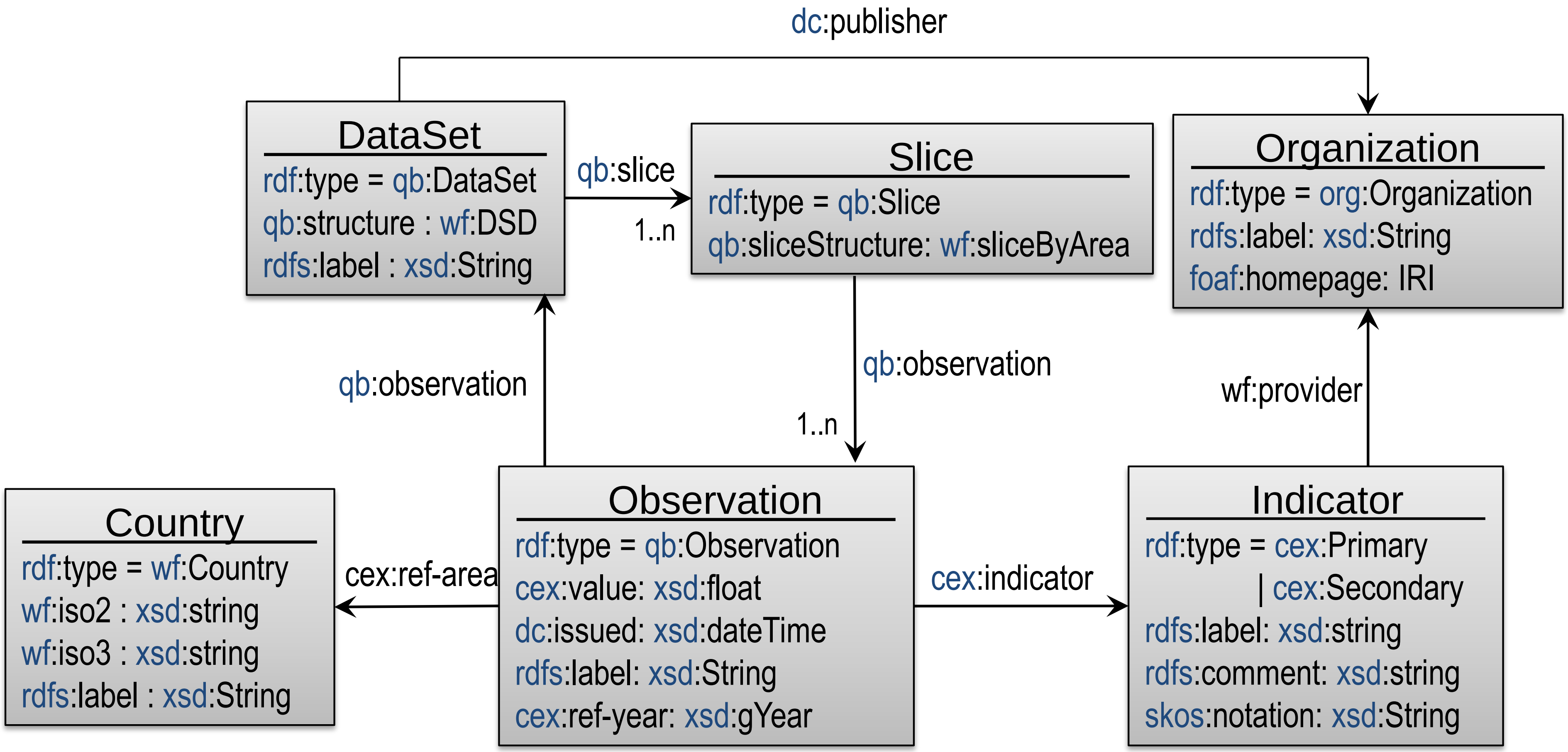
Indicators are provided by Organizations

- Examples
- ITU = International Telecommunication Union
- UN = United Nations
- WB = World bank
- ...



ITU_B	2010	2011	2012	...
Germany	20.34	35.46	37.12	...
Spain	19.12	23.78	25.45	...
France	20.12	21.34	28.34	...
...	...	...	...	...

Main webIndex data model*



DataSet
rdf:type = qb:DataSet
qb:structure : wf:DSD
rdfs:label : xsd:String

Slice
rdf:type = qb:Slice
qb:sliceStructure: wf:sliceByArea

Organization
rdf:type = org:Organization
rdfs:label: xsd:String
foaf:homepage: IRI

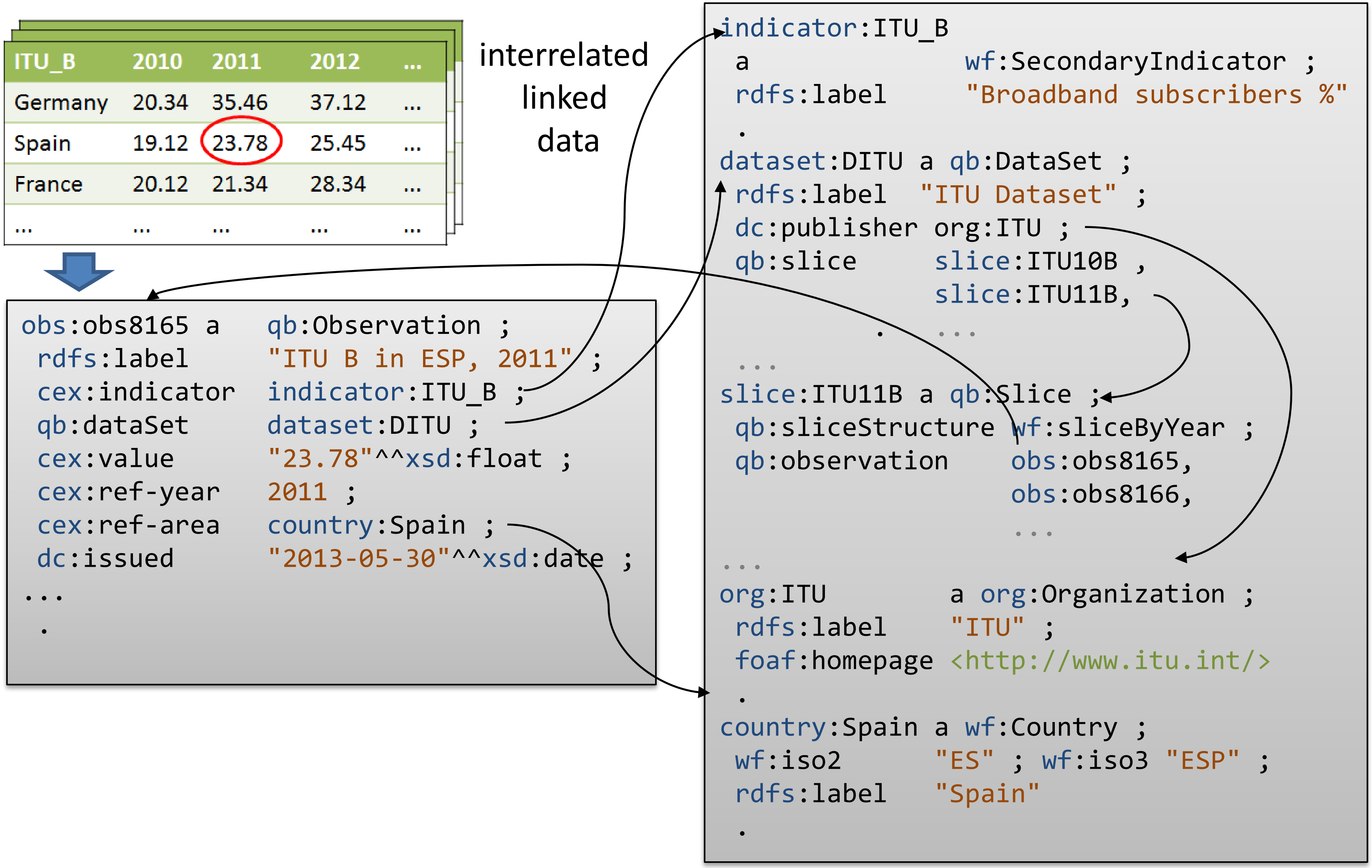
Country
rdf:type = wf:Country
wf:iso2 : xsd:string
wf:iso3 : xsd:string
rdfs:label : xsd:String

Observation
rdf:type = qb:Observation
cex:value: xsd:float
dc:issued: xsd:dateTime
rdfs:label: xsd:String
cex:ref-year: xsd:gYear

Indicator
rdf:type = cex:Primary
| cex:Secondary
rdfs:label: xsd:string
rdfs:comment: xsd:string
skos:notation: xsd:String

*Simplified

Excel → RDF (Turtle)



ITU_B	2010	2011	2012	...
Germany	20.34	35.46	37.12	...
Spain	19.12	23.78	25.45	...
France	20.12	21.34	28.34	...
...	...	...	...	...

interrelated
linked
data

```
obs:obs8165 a   qb:Observation ;
 rdfs:label     "ITU B in ESP, 2011" ;
 cex:indicator  indicator:ITU_B ;
 qb:dataSet     dataset:DITU ;
 cex:value      "23.78"^^xsd:float ;
 cex:ref-year   2011 ;
 cex:ref-area   country:Spain ;
 dc:issued      "2013-05-30"^^xsd:date ;
...
 .
```

```
indicator:ITU_B
 a              wf:SecondaryIndicator ;
 rdfs:label     "Broadband subscribers %"
 .
dataset:DITU a qb:DataSet ;
 rdfs:label  "ITU Dataset" ;
 dc:publisher org:ITU ;
 qb:slice      slice:ITU10B ,
               slice:ITU11B,
         .     ...
 ...
slice:ITU11B a qb:Slice ;
 qb:sliceStructure wf:sliceByYear ;
 qb:observation    obs:obs8165,
                   obs:obs8166,
                   ...
...
org:ITU         a org:Organization ;
 rdfs:label     "ITU" ;
 foaf:homepage <http://www.itu.int/>
 .
country:Spain a wf:Country ;
 wf:iso2       "ES" ; wf:iso3 "ESP" ;
 rdfs:label    "Spain"
 .
```

Description and Validation

Lots of constraints

- *Observations must be linked to some country*
- *Observations have a float value*
- *Observations are related with an indicator, a country and a year*
- *Dataset contains several slices and slices contain several observations*
- ....etc.

Q: How can we express those constraints easily?

- Our proposal: Shape expressions

Country

A <Country> has *at least* the following properties:

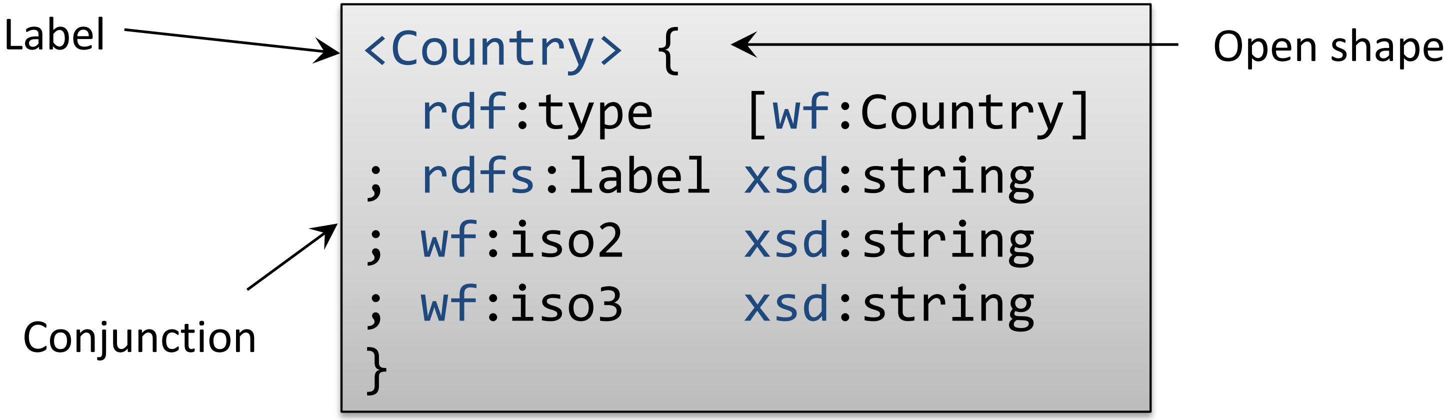
rdf:type with value wf:Country

rdfs:label with value of type xsd:string

wf:iso2 with value of type xsd:string

wf:iso3 with value of type xsd:string

Using shape Expressions:

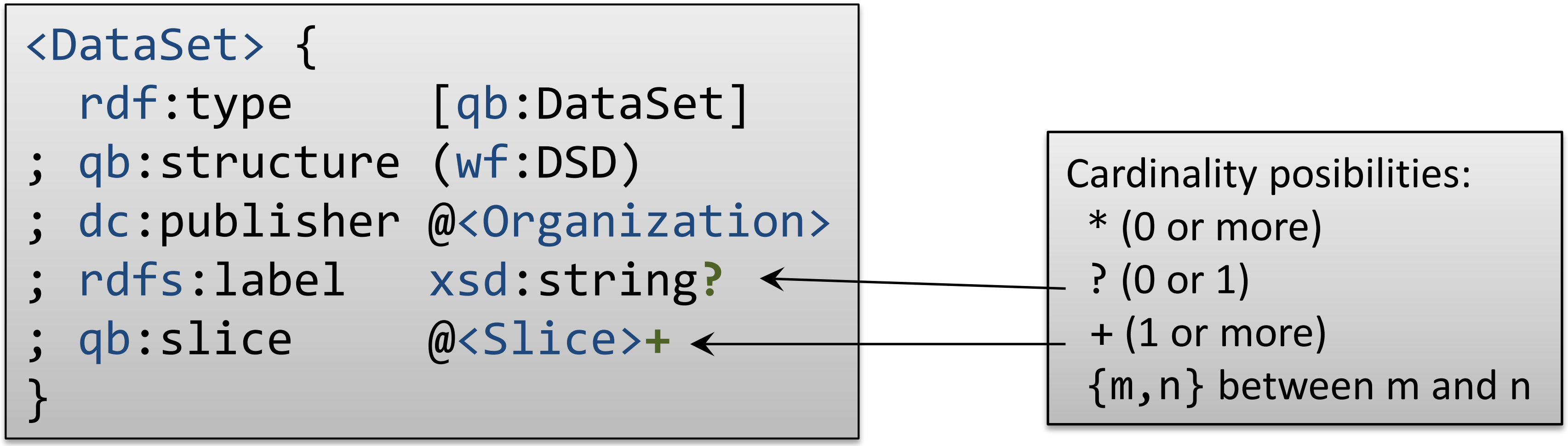


```
<Country> {
  rdf:type   [wf:Country]
; rdfs:label xsd:string
; wf:iso2    xsd:string
; wf:iso3    xsd:string
}
```

DataSets

A <DataSet> has the shape:

- rdf:type with value qb:Dataset
- qb:structure with value wf:DSD
- *Optional* rdfs:label with value of type xsd:string
- One or more qb:slice with shape <Slice>



```
<DataSet> {
  rdf:type      [qb:DataSet]
; qb:structure (wf:DSD)
; dc:publisher @<Organization>
; rdfs:label    xsd:string?
; qb:slice      @<Slice>+
}
```

Cardinality posibilities:
- * (0 or more)
- ? (0 or 1)
- + (1 or more)
- {m,n} between m and n

Slices

```
<Slice> {
   rdf:type           [qb:Slice]
 ; qb:sliceStructure (wf:sliceByYear)
 ; qb:observation    @<Observation>+
 ; cex:indicator     @<Indicator>
 }
```

<Slice> has the properties:

rdf:type with value qb:Slice

qb:SliceStructure with value wf:sliceByYear

Several qb:observation with shape <Observation>

cex:indicator with shape <Indicator>

Observations

```
<Observation> {
  rdf:type      [qb:Observation]
; cex:value     xsd:float ?
; dc:issued     xsd:dateTime
; rdfs:label    xsd:string ?
; qb:dataSet    @<DataSet>
; cex:ref-area  @<Country>
; cex:indicator @<Indicator>
; cex:ref-year  xsd:gYear
}
```

...and more

Indicators

```
<Indicator> {
  rdf:type      [wf:PrimaryIndicator wf:SecondaryIndicator]
; rdfs:label    xsd:string
; rdfs:comment  xsd:string ?
; skos:notation xsd:string ?
}
```

Organizations

```
<Organization> {
  rdf:type [org:Organization]
; rdfs:label xsd:string
; foaf:homepage IRI
; org:hasSubOrganization @<Organization>
}
```

Use of shape expressions in WebIndex

1. Documentation of linked data portal

   Human-readable

   Machine processable

   http://weso.github.io/wiDoc

2. Team communication

   Communicate the developers which shapes they had to generate

3. Validation

   For example: check if a value of type `qb:Observation` has shape <Observation>

WebIndex as a benchmarking

We have created a tool to generate synthetic RDF data that conforms (or not) to the WebIndex data model

The tool can be used to benchmark ShEx and SHACL

See: http://labra.github.io/wiGen/

HL7 FHIR

ShEx is currently being used to develop FHIR/RDF

- validate examples (in documentation)
- exchange site-specific restrictions
- enable consumer and producer validation

See: https://www.w3.org/2016/FHIR-tutorial/Constellations

SHACL applications

TopBraid Composer includes support for SHACL

See: http://www.topquadrant.com/technology/shacl/tutorial/

RDFUnit is also planning to include SHACL support

See: https://github.com/AKSW/RDFUnit

OpenPublicData: prototype to list, filter and present open data

See: http://www.openpublicdata.com/

Schema.org converted to SHACL

See:http://datashapes.org/schema

Future work

SHACL Recommendation

- Data Shapes WG chartered until Jul 2017

Other features

- Property paths
- Named graphs
- …

ShEx vs SHACL

- Translate ShEx to SHACL (looks difficult, impossible?)
- Translate SHACL to ShEx (work in progress, see Shaclex)

Future work

SHACL: Data Shapes Working Group:

Mailing list, list of issues,...

https://www.w3.org/2014/data-shapes/

ShEx Community portal http://shex.io

List of issues:

https://github.com/shexSpec/shex/issues