

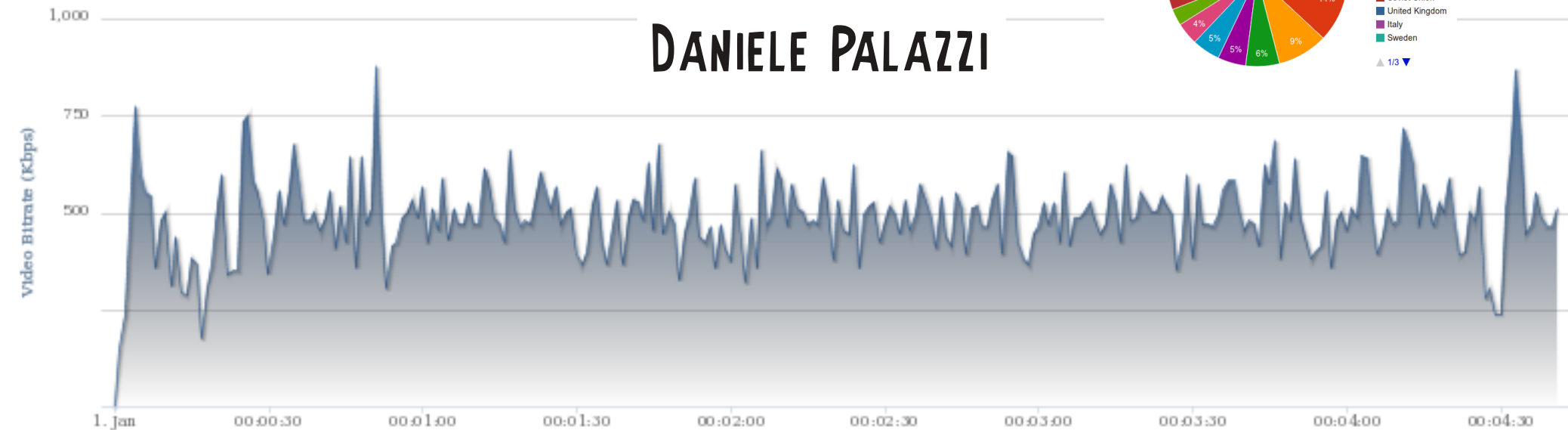
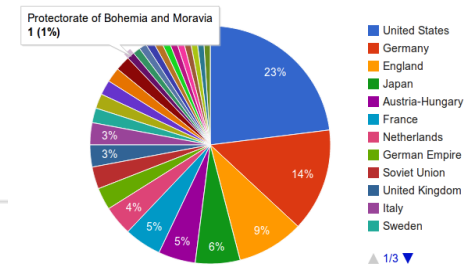
VISUALIZAÇÃO DA INFORMAÇÃO

PROFESSORA: ADRIANA VIVAQUA

PERÍODO: 2012/3

VISUALIZAÇÃO DE DADOS ESTATÍSTICOS EM LOD

ALAN TYGEL
DANIELE PALAZZI

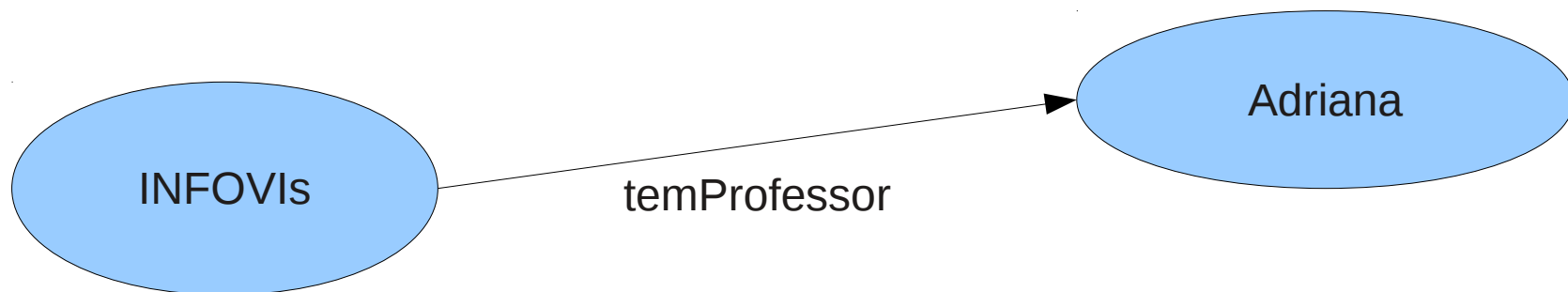


H.264

LINKED OPEN DATA - LOD

- "Linked Data é um conjunto das melhores práticas para a publicação e conexão de dados estruturados na Web, permitindo estabelecer links entre itens de diferentes fontes de dados para formar um único espaço de dados global"

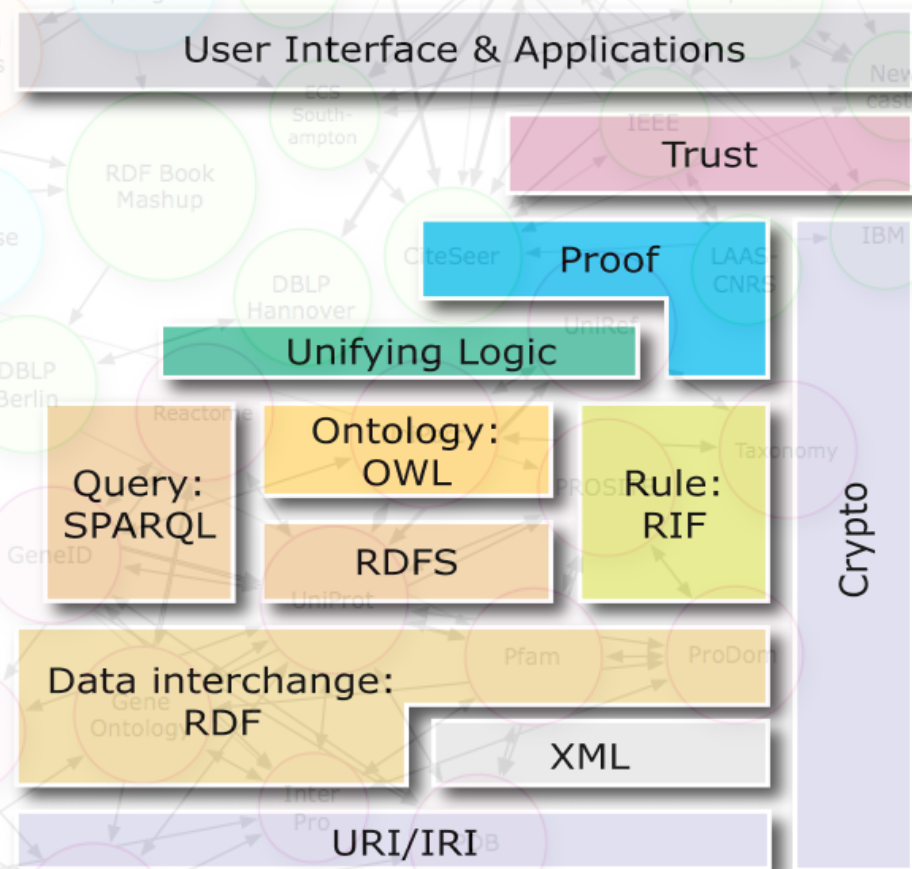
[Heath, T. and Bizer, C. 2011]



LINKED OPEN DATA - LOD

- Utilização de padrões abertos
- Dados são representados através de triplas RDF (Resource Description Framework)
- Interoperabilidade de metadados
- Abordagem mais simplista da Web Semântica

Mais e mais dados sendo gerados e potencialmente “ligáveis”



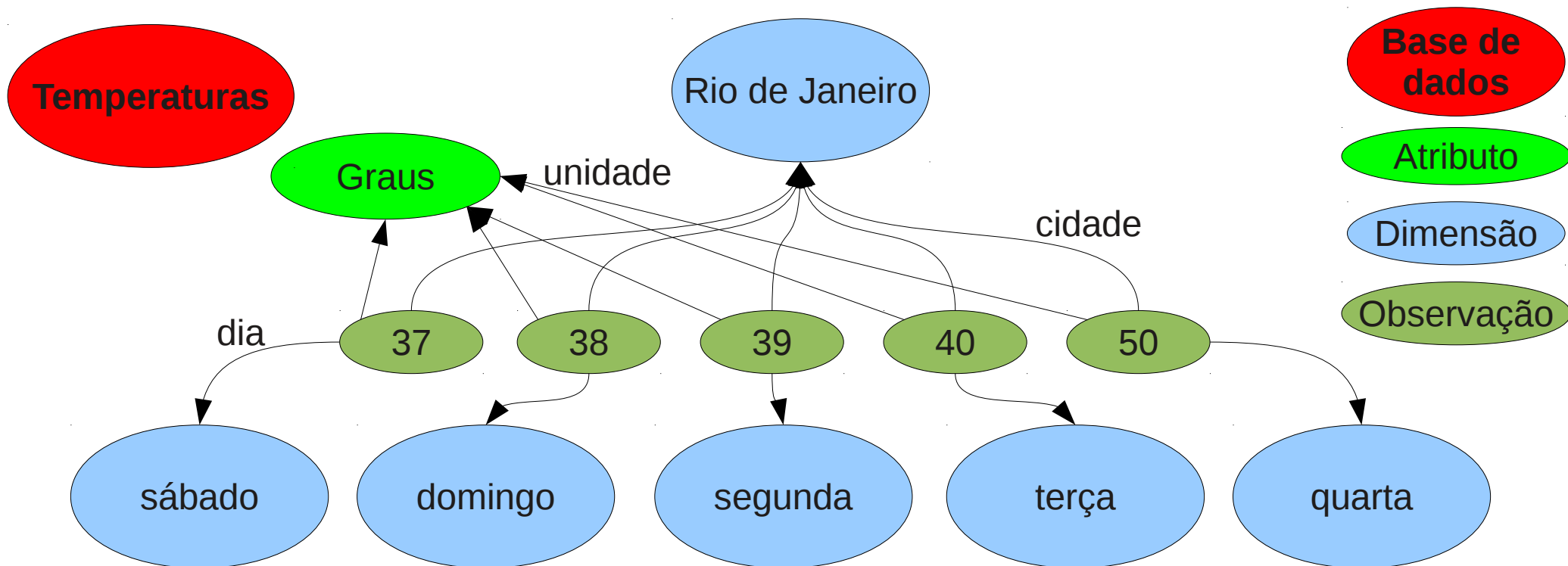
DIFICULDADES NA REPRESENTAÇÃO DE DADO ESTATÍSTICO EM LOD

- LOD se aplica muito bem aos dados que são ligados por essência: redes (sociais), informações sobre coisas (DBPedia)
- No entanto, para dados estatísticos a (falta de) estrutura de LOD é um problema
- Dados estatísticos são estruturados por natureza; uma observação não existe sozinha!

DADO ESTATÍSTICO

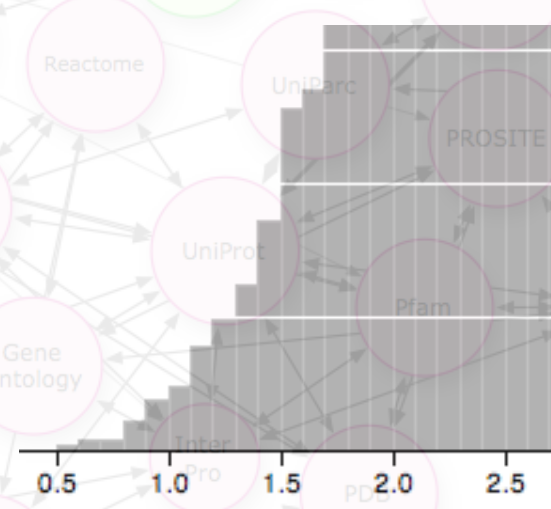
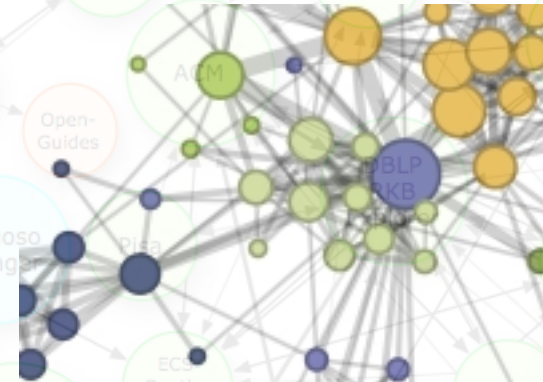
- Sequências de observações ao longo de uma ou mais dimensões (tempo, espaço, etc...)

| Dia | Temperatura | Cidade |
|---------|-------------|----------------|
| Sábado | 37 | Rio de Janeiro |
| Domingo | 38 | Rio de Janeiro |
| Segunda | 39 | Rio de Janeiro |
| Terça | 40 | Rio de Janeiro |
| Quarta | 50 | Rio de Janeiro |



TÉCNICAS DE VISUALIZAÇÃO EM LOD

- LOD
- Grafos
- Dados estatísticos
- Gráfico de barra
- Gráfico de pizza
- Histogramas



LOD E DADOS ESTATÍSTICOS - REPRESENTAÇÃO

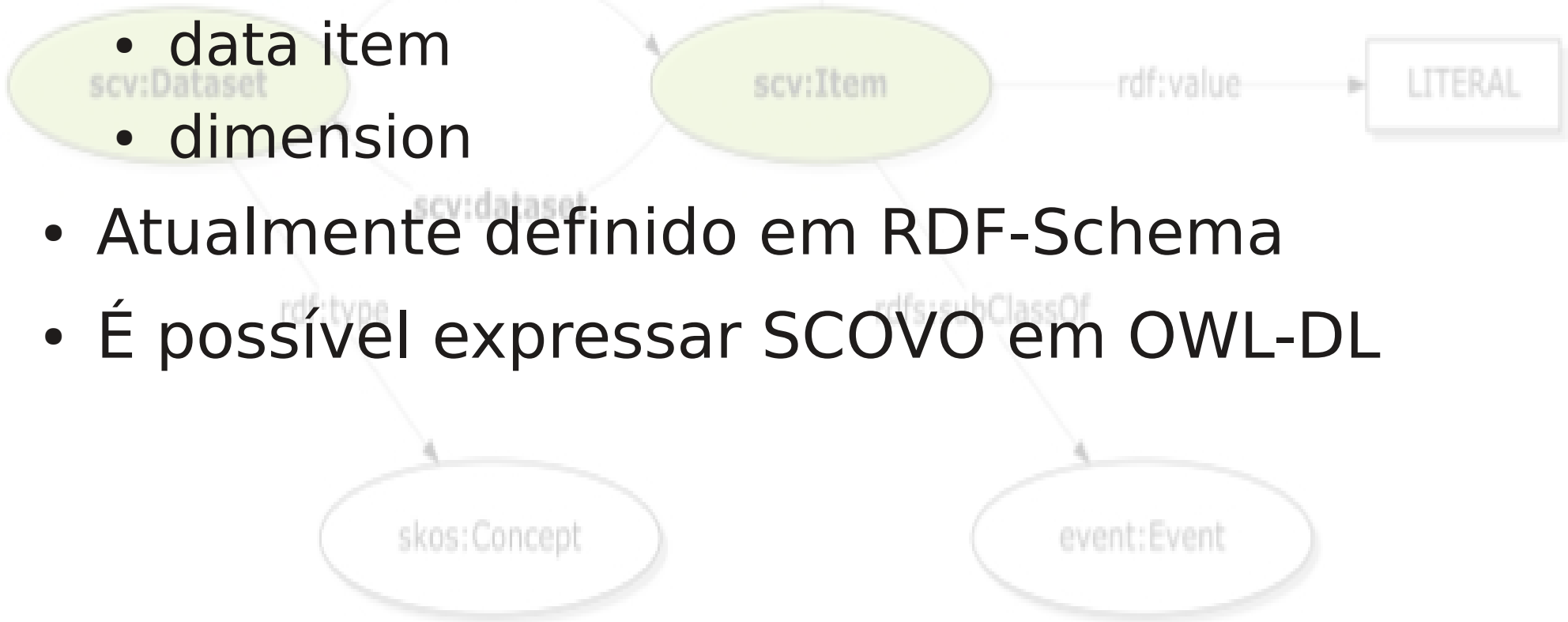
- SCOVO
- SMDX
- Data Cube
- RDF



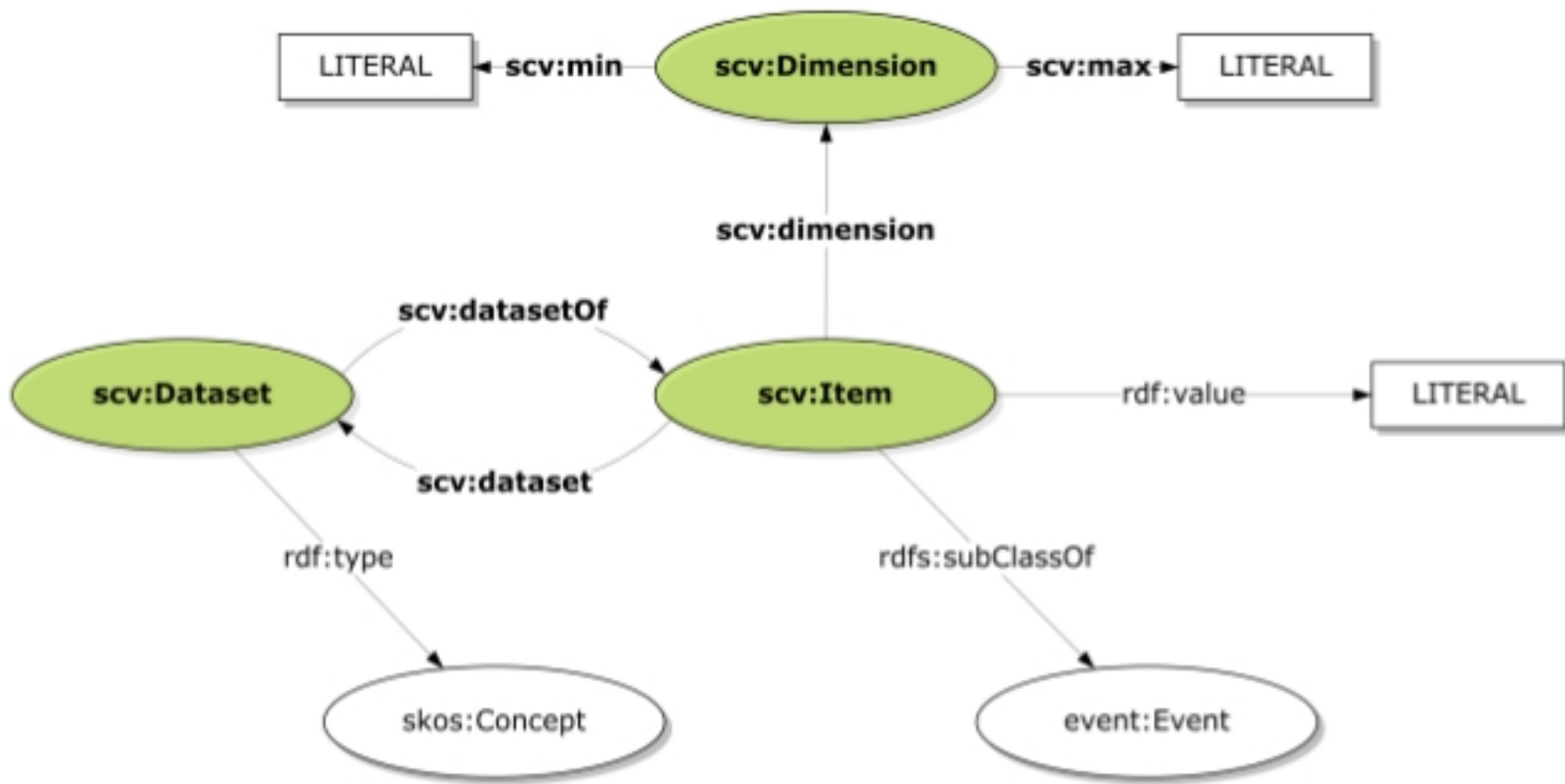
SCOVO



- SCOVO - Statistical Core Vocabulary
- Conceitos básicos: `scv:dimension`
 - dataset
 - data item
 - dimension
- Atualmente definido em RDF-Schema
- É possível expressar SCOVO em OWL-DL



SCOVO



DATA CUBE

- Permite que a informação seja representada usando RDF e publicada seguindo os princípios de dados ligados
- Baseado em SDMX
- Vocabulário focado exclusivamente na publicação de dados multidimensionais
- Um cubo é organizado de acordo com um conjunto de dimensões, atributos e medidas

DATA CUBE

- Dimensões, atributos e medidas
- Slices

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

RDF - RESOURCE DESCRIPTION FRAMEWORK

- Linguagem para representar informação sobre recursos na Web
- Padrão W3C
- As informações são representadas por **declarações** (statements) contendo **sujeito, predicado e objeto**.



LOD e DADOS ESTATÍSTICOS - VISUALIZAÇÃO

- Foram encontradas duas ferramentas razoavelmente maduras que oferecem opções de visualização de dados armazenados em RDF
- Ambas em estágio inicial de desenvolvimento, com pouca documentação e poucos usuários
- Ontowiki → Cubeviz
- Visualbox

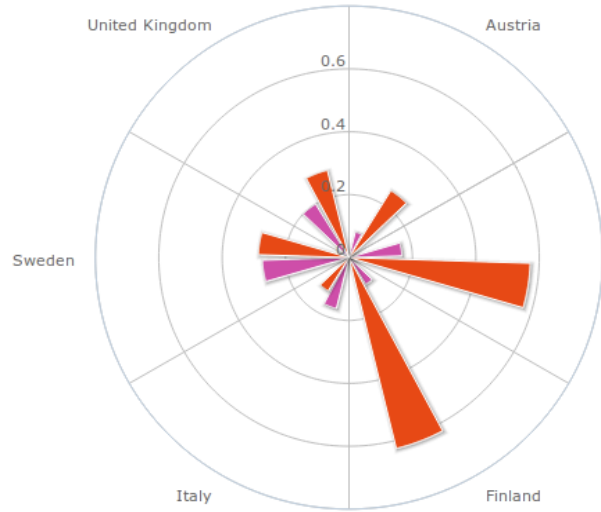
CUBEVIZ

- Desenvolvido pelo Agile Knowledge Engineering and Semantic Web (AKSW), da Alemanha → LOD2
- É um plugin do Ontowiki
- Funciona em conjunto com o Stats2RDF, que converte arquivos CSV para Datacube (semiautomático)
- Permite visualização facetada em gráficos e tabelas descritos em Datacube
- Não traz nada além do que uma ferramenta de visualização de dados estruturados traria
- Desenvolvimento: <http://aksw.org/Projects/CubeViz.html>
- Exemplos: <http://cubeviz.aksw.org/>

CUBEVIZ

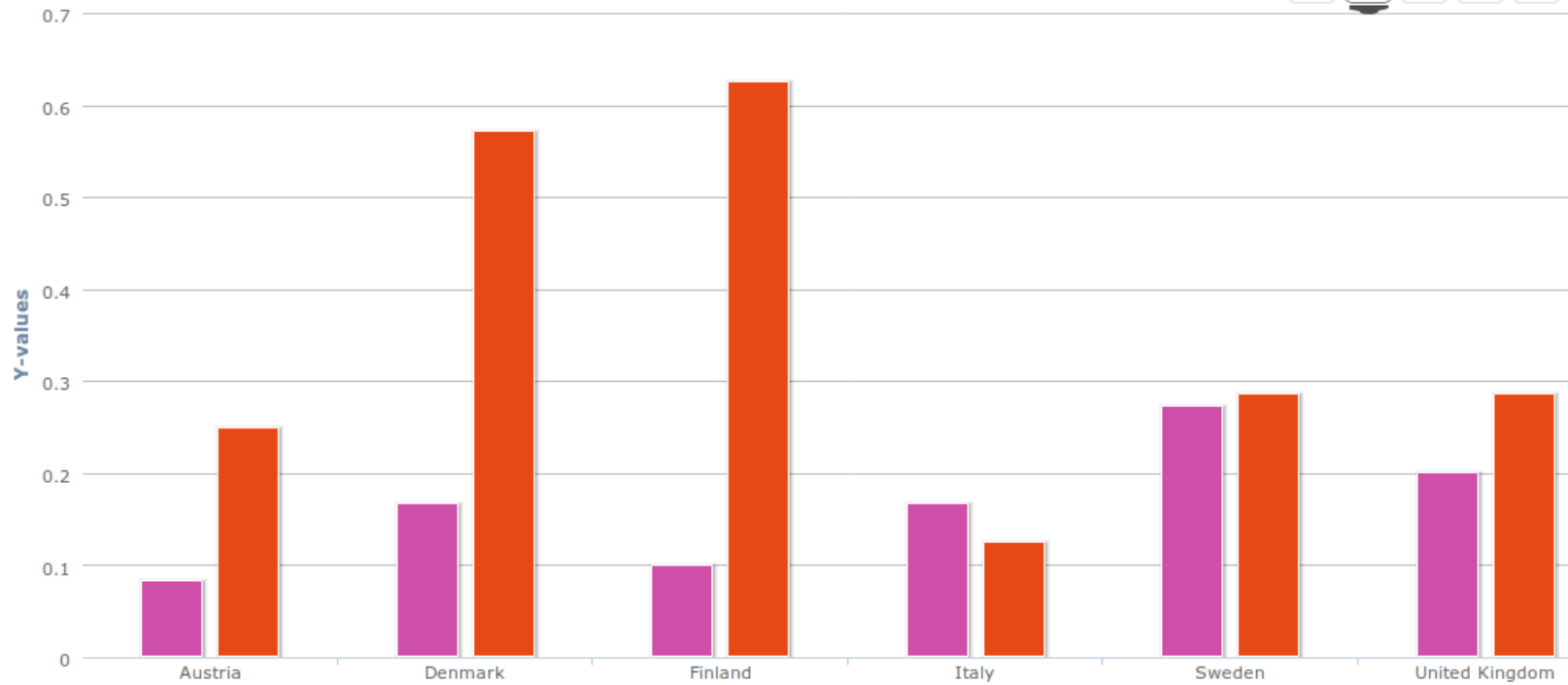
Observations

| Label | Country | Indicator | Value |
|---|-------------------------|---|--------------|
| Availability of eGovernment services - citizens for Austria in 2001 | Austria | % of basic public services for citizens, which are fully available online | 0.0833333333 |
| Availability of eGovernment services - citizens for Denmark in 2001 | Denmark | % of basic public services for citizens, which are fully available online | 0.1666666667 |
| Availability of eGovernment services - citizens for Finland in 2001 | Finland | % of basic public services for citizens, which are fully available online | 0.1 |
| Availability of eGovernment services - citizens for Italy in 2001 | Italy | % of basic public services for citizens, which are fully available online | 0.1666666667 |
| Availability of eGovernment services - citizens for Sweden in 2001 | Sweden | % of basic public services for citizens, which are fully available online | 0.2727272727 |



Visualization for <http://data.lod2.eu/scoreboard/>

per Year - 2001



■ % of basic public services for citizens, which are fully available online
 ■ % of basic public services for enterprises, which are fully available online

Main navigation

- Getting started
- HowTo
- Examples
- FAQ
- CubeViz Home
- Project Homepage

Knowledge Bases

View

<http://data.lod2.eu/scoreboard/>

Data Selection

Data Structures

per Year

Data Sets

2001

Dimensions

Country (6 / 17)

Indicator (2 / 2)

Permalink Update visualization

VISUALBOX

- Versão simplificada do LODSPeaKr, feita por um desenvolvedor chileno (Alvaro Graves)
- Ferramenta genérica de visualização de dados em LOD a partir de um endpoint qualquer
- Workflow simples permite utilizar uma base qualquer e gerar a visualização desejada
- Configuração:
 - Endpoint – base de dados em RDF acessível
 - Query – Query SPARQL para pegar os dados que deseja visualizar
 - Template – A partir de opções de visualização dos dados (grafo, gráficos, tabelas) cria-se um template HTML para ser inserido em um sistema
- Desenvolvimento: <http://alangrafu.github.com/visualbox/>
- Demo: <http://visualbox.org/demo/index.html>

VISUALBOX

Pré-definidas

Filtros de Visualização

COMPONENTS new

- cities_by_area
- heinlein_and_bradbury
- instances
- kubrick_filmography**
- lamborghini_cars
- namedGraphs
- nobel_laureates_by_country
- people_influenced_by_neruda
- places_in_rensselear_county_by_area
- rush
- schoolmap
- search
- triples_in_a_table

Views

- html**

Models new

- main**

View component

Embed component

- Visualization Filters
- D3CirclePacking
 - D3Dendrogram
 - D3ForceGraph
 - D3ParallelCoordinates
 - D3WordCloud
 - GoogleMaps
 - GoogleVizBarChart
 - GoogleVizColumnChart
 - GoogleVizLineChart
 - GoogleVizPieChart
 - GoogleVizScatterChart
 - GoogleVizTable
 - Timeknots

Template

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="utf-8">
<title>My visualization</title>
</head>
<body>
<!--
The width=1000, color=steelblue and dateFormat='%Y' are optional
-->
{{models.main|Timeknot:"releaseDate,label,color=steelblue,width=1000,dateFormat='%Y'"}} </body>
</html>
```

Save

Template HTML

Query

```
#Get all the movies directed by Stanley Kubrick that have a release date
#The MAX function is a tricky way of obtaining just one value in case there are multiple dates associated
PREFIX dbp: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?movie ?label MAX(?r) AS ?releaseDate WHERE{
?movie dbo:director dbp:Stanley_Kubrick ;
rdfs:label ?label;
dbo:releaseDate ?r .

#Filter forces to get only the label in english
FILTER( LANG(?label) = "en")
}GROUP BY ?movie ?label
```

Save

Test this query against

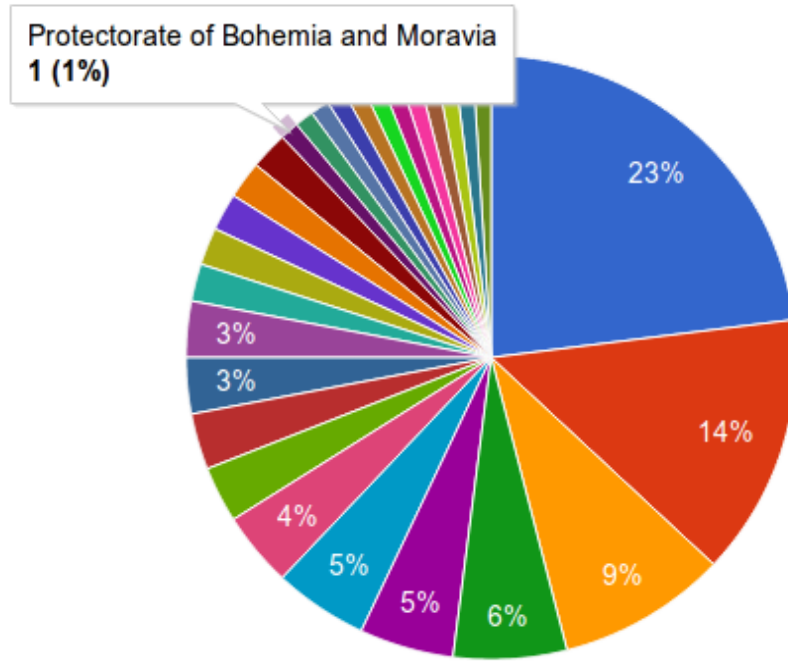
local (http://localhost:8080/openi

Consulta SPARQL

SPARQL Endpoint

VISUALBOX

Protectorate of Bohemia and Moravia
1 (1%)

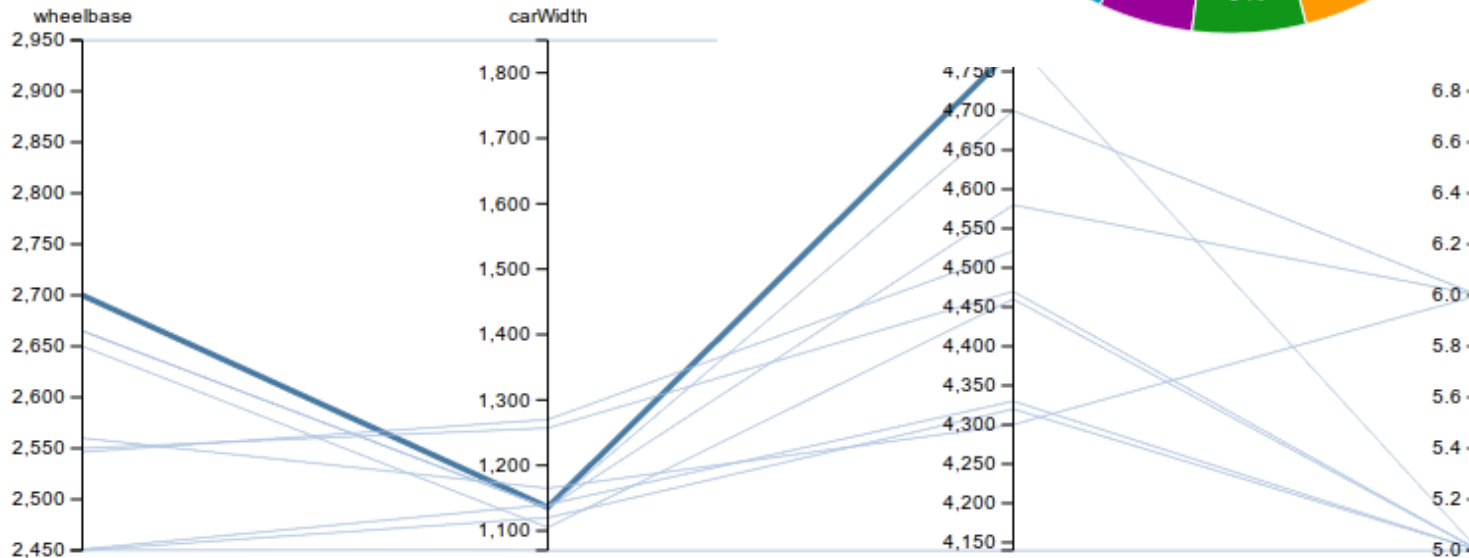


- United States
- Germany
- England
- Japan
- Austria-Hungary
- France
- Netherlands
- German Empire
- Soviet Union
- United Kingdom
- Italy
- Sweden

▲ 1/3 ▼

Measures of different models of Lamborghini cars

Lamborghini Aventador



NAVEGADORES LINKED DATA

- Possibilita navegação entre fontes de dados expressas em triplas RDF

- Disco
- Marbles
- Tabulator
- IsaViz
- RDF Gravity
- ...

DISCO

Christian Bizer | Disco - Hyperdata Browser - Windows Internet Explorer

URI: <http://www4.wiwiss.fu-berlin.de/dblp/resource/person/315759> Go!

| Property | Value | Sources |
|-----------------|---|---------|
| more data | List of all instances. People ↗ | G2 |
| type | http://xmlns.com/foaf/0.1/Person ↗ | G2 |
| label | Christian Bizer ↗ | G2 |
| sourceURL | Christian Bizer ↗ | G1 |
| name | Christian Bizer ↗ | G2 |
| is Creator of | Business to Consumer Markets on the Semantic Web. ↗ | G2 G7 |
| is Creator of | Semantic Web Publishing using Named Graphs. ↗ | G2 G7 |
| is Creator of | The Impact of Semantic Web Technologies on Job Recruitment Processes. ↗ | G2 G4 |
| is Creator of | D2R MAP - A Database to RDF Mapping Language. ↗ | G2 G7 |
| is Creator of | Using context- and content-based trust policies on the semantic web. ↗ | G2 G6 |
| is Creator of | Named graphs, provenance and trust. ↗ | G2 G7 |
| is Creator of | A Web Service Market Model based on Dependencies. ↗ | G2 G7 |
| is Creator of | Named graphs. ↗ | G2 G5 |
| is sameAs of | Chris Bizer ↗ | G3 |
| is sourceURL of | Christian Bizer ↗ | G1 |

Sources
Displayed information originates from the following RDF graphs:

- G1. <http://localhost/provenanceInformation> [↗](#)
- G2. <http://www4.wiwiss.fu-berlin.de/dblp/resource/person/315759> [↗](#)
- G3. <http://sites.wiwiss.fu-berlin.de/sub/bizer/foaf.rdf> [↗](#)
- G4. <http://www4.wiwiss.fu-berlin.de/dblp/resource/record/conf/wirtschaftsinformatik/BizerHMOTE05> [↗](#)
- G5. <http://www4.wiwiss.fu-berlin.de/dblp/resource/record/journals/ws/CarrollBHS05> [↗](#)
- G6. <http://www4.wiwiss.fu-berlin.de/dblp/resource/record/conf/www/BizerO04> [↗](#)
- G7. <http://www4.wiwiss.fu-berlin.de/dblp/resource/record/conf/otm/TalksdorBEH03> [↗](#)

Session Cache
[Display all RDF graphs that are currently in your session cache.](#)

Label of the displayed resource

Navigation box

Resource description

List of all source graphs

Link for displaying the session cache

Go to URI button

Sources of each piece of information

CONCLUSÃO

- O que gostaríamos de ter:
 - Formas de representação de dados estatísticos que pudessem aliar a simplicidade do modelo relacional à flexibilidade dos dados ligados
 - Formas de visualização que fossem além do que já existe para análise gráfica de dados estatísticos, explorando a semântica, as ligações, e a flexibilidade dos dados abertos
- O que temos:
 - Datacube representa um grande avanço em termos de representação, pois adapta o estado da arte em metadados estatísticos (SDMX) ao mundo LOD
 - Entretanto, a representação dos dados em LOD representa custos de armazenamento e processamento muito altos em relação aos BDs relacionais, que têm mais de 30 anos de otimização
 - As ferramentas de visualização ainda não exploram o potencial semântico e de ligação dos dados em LOD.
 - Exemplos de melhorias:
 - _ Visualização com dados de várias bases
 - _ Navegação a partir da visualização, e visualização a partir da navegação

REFERÊNCIAS

• Visualização de LOD

- Dadzi e Rowe, Approaches to Visualising Linked Data: A Survey
- Stuhr et al., LODWheel - JavaScript-based Visualization of RDF Data

• LOD Estatístico

- Zapilko e Mathiak, Defining and Executing Assessment Tests on Linked Data for Statistical Analysis
- Zapilko et al., Enriching and Analysing Statistics with Linked Open Data

• Visualização da Dados estatísticos em LOD

- Bruneti et al., The Linked Data Visualization Model
- Salas et al., Publishing Statistical Data on the Web

• Ferramentas

- Ontowiki
- SDMX
- CubeViz / Datacube
- SCOVO

• Outros protótipos de visualização de dados estatístico em LOD:

- <http://semwiq.faw.uni-linz.ac.at/node/9>
- <http://lodvisualization.appspot.com/>