

# The Linking Open Data W3C SWEO Community Project

Chris Bizer  
Freie Universität Berlin

## Basic Ideas

- 1. Apply the Linked Data principles to large open datasets**
  - Publish open datasets as Linked Data on the Web
  - Set RDF Links between data items from different datasets
- 2. Provide a forum for working on the tasks arising from having lots of Linked Data on the Web**
  - Browsers and other user interfaces
  - Semantic Web crawlers and search engines
  - RDF repositories and reasoning engines
  - Linked Data publishing tools
  - Record linkage algorithms
  - Data cleansing tools

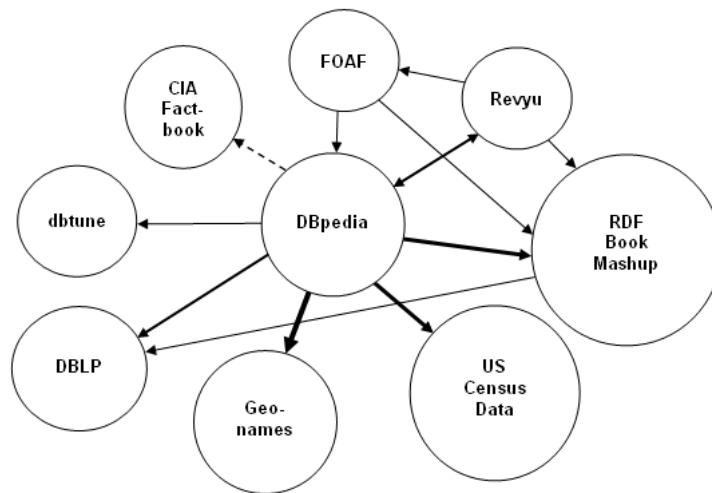
## Spirit of the Project

- Don't talk about the Semantic Web, but do it, NOW!
- Get as much Linked Data as possible online, NOW!
- Get as many people engaged as possible.
- Solve the easy problems first, do the hard ones later.

## Infrastructure

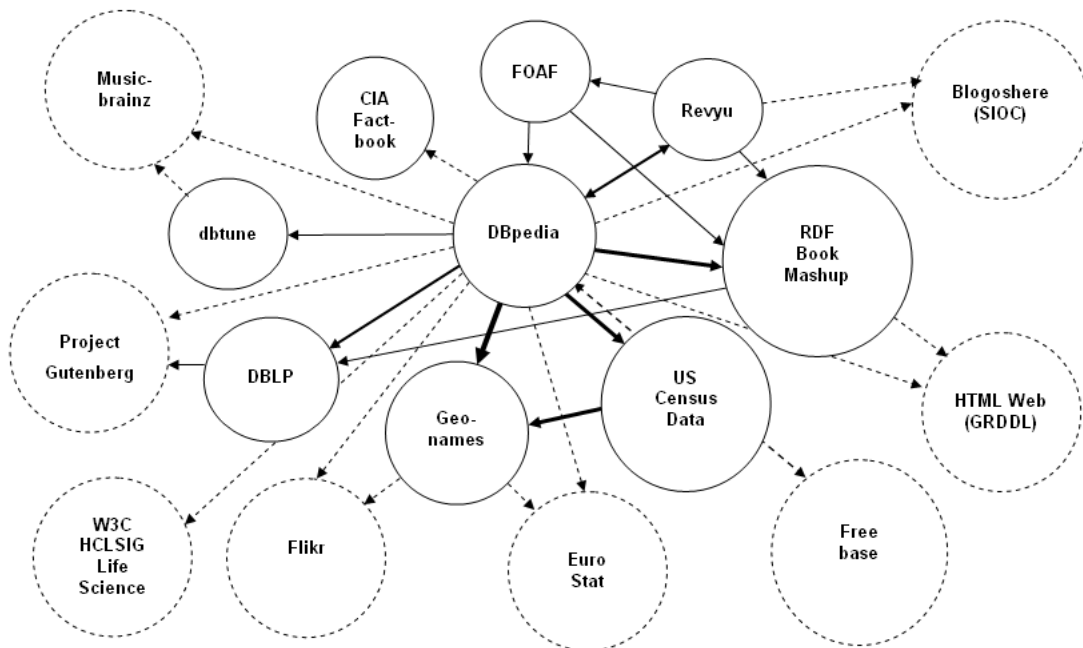
- **Wiki Page**
  - <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
- **Mailing List**
  - <http://simile.mit.edu/mailman/listinfo/linking-open-data>
- **Participating in the project**
  - Put your name on the Wiki page
  - Subscribe to the mailing list
  - Do something useful

## What did we achieve so far?



- About 600 million triples served as Linked Data
- About 100 000 RDF links between different data sets

## Where will we be in 2 month?



## Example Data Source: DBpedia

### ■ DBpedia.org is a community effort to

- extract structured information from Wikipedia
- make this information available on the Web under an open license
- interlink the DBpedia dataset with other datasets on the Web

### ■ Contributors

- Freie Universität Berlin (Germany)
- Universität Leipzig (Germany)
- OpenLink Software (UK)
- Linking Open Data Community (W3C SWEO)



## Extracting Structured Data from Wikipedia

<http://en.wikipedia.org/wiki/Calgary>

```
<http://dbpedia.org/resource/Calgary>
  dbpedia:native_name "Calgary" ;
  dbpedia:altitude "1048" ;
  dbpedia:population_city "988193" ;
  dbpedia:population_metro "1079310" ;
  mayor_name
    dbpedia:Dave_Bronconnier ;
  governing_body
    dbpedia:Calgary_City_Council ;
  ...
```

- Altogether 9,100,000 RDF triples extracted from 754,000 infoboxes

| Calgary  |   |
|--|---|
|  |   |
| Downtown Calgary.  |   |
| <b>Government</b>  |   |
| - Mayor  | Dave Bronconnier<br>(Past mayors)       |
| - Governing body   | Calgary City Council                    |
| - Manager  | Owen A. Tobert                          |
| <b>Area</b> <sup>[1]</sup>   |   |
| - City   | 726.50 km <sup>2</sup> (280.5 sq mi)    |
| - Metro  | 5,107.43 km <sup>2</sup> (1,972 sq mi)  |
| <b>Elevation</b>   | 1,048 m (3,438.3 ft)                    |
| <b>Population</b> (2006) <sup>[1]</sup>  |   |
| - City   | 988,193                                 |
| - Density  | 1,360.2/km <sup>2</sup> (3,522.9/sq mi) |
| - Metro  | 1,079,310                               |
| - Population rank  | 3rd                                     |
| - Metro rank   | 5th                                     |

# The DBpedia Dataset

- 1,600,000 concepts
- including
  - 58,000 persons
  - 70,000 places
  - 35,000 music albums
  - 12,000 films
- described by 91 million triples
- using 8,141 different properties.
  
- abstracts in 10 different languages
- 557,000 links to pictures
- 1,300,000 links to relevant external web pages
- 207,000 Wikipedia categories
- 75,000 YAGO categories

# Multi-Lingual Abstracts

- The dataset contains a short and a long abstract for each concept.

## ■ Short abstracts

- English: 1,637,622
- German: 246,791
- French: 206,085
- Dutch: 133,746
- Polish: 118,874
- Italian: 113,950
- Spanish: 112,417
- Japanese: 106,610
- Portuguese: 104,842
- Swedish: 100,267
- Chinese: 54,991

| Property | Value   |
|----------|---|
| abstract | <p>([zh-hw:卡加利], [en:Calgary]) «[zh-hw:卡加利] [en:Calgary] 係一座位于加拿大艾伯塔省南部的城市，面积709平方公里，海拔1045米，人口约100万。是艾伯塔省经济、金融、政治中心。卡加利是「四大城市」中人口第二多的城市。十九世纪七十年代，开凿到太平洋的加拿大铁路，成为城市发展的关键因素 (North West Mounted Police) 的一部份。在加拿大太平洋铁路修建前，卡加利是连接艾伯塔省、1941年这里发现了丰富的石油天然气，从此城市开始「石油的繁荣」。世界上大部分的中国石油公司都在这里设有分支机构。很多大的石油公司总部就设在这里。因此卡加利是加拿大石油业的中心，1909年在这里举办了世界石油工业大会，这个城市也因此成为石油工业的总部。这个城市的工作岗位完全依赖于石油工业。卡加利是加拿大最大的城市。七月四日，一年一度的(Stampede)是这个城市最盛大的节日，每年都有超过百万的游客来这里参加这个节日。城市以西的皇家国家公园(Royal National Park)是世界著名的自然风景区。在加拿大第一国家公园，自然历史博物馆。卡加利是加拿大最安全的大城市之一。卡加利是中国的十大最佳旅游目的地。」</p>   |
| abstract | <p>Calgary (deutscher Name: Wassen) ist eine Stadt in Kanada in der Provinz Alberta. Sie ist die drittgrößte und am schnellsten wachsende Großstadt Kanadas und hat 952.079 Einwohner (Stand 2005).</p>   |
| abstract | <p>Calgary est la mayor ciudad de la provincia de Alberta, Canadá. Se sitúa en una región de colinas y altiplanicios, a aproximadamente 100 km al este de las Montañas Rocosas. Tercera ciudad de Canadá en términos de población, contaba según el censo de abril de 2005 con 991.759 habitantes. La población estimada de su área metropolitana era de 1.000.300 habitantes en 2005 (nueva Región de Calgary), lo que la convierte en la quinta mayor de Canadá. El "condado Calgary-Edmonton" es la región urbana más poblada ubicada entre Toronto y Vancouver. Su nombre proviene del de una playa ubicada en la isla de Mull, en Escocia. Los habitantes de Calgary se llaman, en inglés, "Calgarians". La ciudad de Calgary es un destino muy conocido para los deportes de invierno y el ecoturismo; cerca de la ciudad se sitúa una gran cantidad de importantes lugares de vacaciones. La economía de Calgary se centra sobre todo en la industria petrolera, aunque la agricultura, el turismo y la alta tecnología también contribuyen al tejido económico de la ciudad. Calgary es la sede de varios festivales anuales, como la Calgary Stampede, el Folk Music Festival, el Lilac Festival, el Obfest y el segundo festival de cultura caribeña en importancia del país, el Carifest. En 1980, Calgary se convirtió en la primera ciudad canadiense en acoger los Juegos Olímpicos de Invierno.</p>  |
| abstract | <p>Calgary est la plus grande ville de la province de l'Alberta (Canada). Elle se situe dans le sud de la province, dans une région de collines et de plateaux à environ 100 km à l'est des montagnes Rocheuses. Troisième ville du Canada, en terme de population, elle comptait selon le recensement d'Avril 2005, 991 759 habitants (1 000 300 avec son agglomération en 2005, ce qui la met au cinquième rang des agglomérations canadiennes). La « comté Calgary-Edmonton » est la région urbaine la plus peuplée entre Toronto et Vancouver. Elle tire son nom d'une plage située sur l'île de Mull en Écosse. Les habitants de Calgary sont appelés « Calgarians ». La ville de Calgary est une destination bien connue pour les sports d'hiver et l'écotourisme : un grand nombre de stations de vacances importantes se situent près de la ville. L'économie de Calgary est surtout centrée sur l'industrie pétrolière, l'industrie, l'agriculture, le tourisme et la haute technologie contribuant également au développement économique rapide de la ville. Calgary est également l'hôte de plusieurs festivals annuels majeurs, dont le Stampede de Calgary, le Folk Music Festival, le Lilac Festival, le Obfest et le deuxième festival de culture des Caraïbes en importance au pays (Carifest). En 1980, Calgary devint la première ville canadienne à accueillir les Jeux olympiques d'hiver. Calgary est la ville la plus prospère dans la province la plus riche du Canada. Anciennement, les bœufs étaient appelés les Cheeks aux yeux bleus.</p> |
| abstract | <p>Calgary is de grootste stad van de Canadese provincie Alberta, en ligt op ongeveer 100 kilometer afstand van de Canadese Rocky Mountains. Calgary had in april 2005 een momenteel van 991 759 en in de agglomeratie waren 1 000 300 mensen. Het is de op drie na grootste stad van Canada, na Toronto, Montréal en Vancouver. De mensen van Calgary staan bekend als Calgarians. De stad is een populair bestemming voor wintersportvakanties door het grote aantal skiënde nabijgelegen 'backcountry' in de bergen. De economie van de stad bestaat vooral uit de petroleumindustrie, ook al worden de landbouw, het toerisme en technologie steeds belangrijker. In 1980 organiseerde Calgary de Olympische Winterspelen. De belangrijkste van de stad is Daw Bioscience. Aan de 'University of Calgary' studeren meer dan 20.000 studenten.</p>   |
| abstract | <p>Calgary är en stad belägen i den södra delen av provinsen Alberta i Kanada. Calgary är Albertas folkrikaste stad med över 991 759 invånare, och Calgary är även den tredje största staden i hela Kanada. Calgary är belägen i Calgary kullar i Kanada. Calgary är en viktig region för Calgary. Calgary är en viktig region för Calgary med ett stort stors berg och väskviktiga berg i närheten av staden.</p>  |
| abstract | <p>Calgary é uma città che si trova nella provincia dell'Alberta, in Canada. È situata nel sud della regione, in una zona di colline e alte pianure, a est delle Montagne Roccose (Rocky Mountains). Con i suoi 991 759 abitanti (stima del 2005) è la più grande città dell'Alberta e la quarta più grande in tutto il Canada.</p>   |
| abstract | <p>Calgary é a cidade mais populosa da provincia canadense de Alberta, e a terceira mais populosa do país, sendo que sua região metropolitana é a quinta mais populosa do país. Localiza-se no sul da provincia, a cerca de 100 quilômetros leste das Montanhas Rochosas. Sua população é de 991 759 habitantes, provendo aproximadamente 1,00 milhão de habitantes na sua região metropolitana. Fundada em 1875, Calgary é atualmente um centro financeiro e comercial, sendo ainda localizada às redes das principais empresas petrolíferas de Canadá.</p>  |
| abstract | <p>Calgary – největší město kanadského provincie Alberta, ležící na předpolí Otů Skalnatých nad rovkou Dne.</p>   |
| abstract | <p>カルガリー (英) は、カナダ・アルバータ州の都市。アルバータ州最大の都市。1944年に建設された石油の産地として急速に発展した。石油産業でカナダの中核的役割を担い、大衆運輸会社のフィーストが拠点を置き、カナダ・アン・ロイヤル・バンクの支店が、1980年にはカルガリー・ホーン・オブ・カナダの拠点となった。毎年7月のはじめには、カワプーイの祭典「カルガリー・スタンピング」が開催される。</p>  |

## Example RDF Links

### ■ Out-Bound RDF Link

```
<http://dbpedia.org/resource/Berlin> owl:sameAs  
<http://sws.geonames.org/2950159> .
```

```
<http://dbpedia.org/resource/Tim_Berners-Lee> owl:sameAs  
<http://www4.wiwiw.fu-berlin.de/dblp/resource/person/100007> .
```

### ■ In-Bound RDF Links

```
<http://www.w3.org/People/Berners-Lee/card#i> owl:sameAs  
<http://dbpedia.org/resource/Tim_Berners-Lee> .
```

```
<http://richard.cyganiak.de/foaf.rdf#cygri> foaf:topic_interest  
<http://dbpedia.org/resource/Semantic_Web> .
```

## Browsing DBpedia together with Linked Data

The screenshot shows the DBpedia search results for "Tim Berners-Lee". At the top, there is a search bar with "Search DBpedia.org" and "Explore Linked Data". Below the search bar, the profile for Tim Berners-Lee is displayed. It includes a small photo of him, a list of properties (like Name, Birth, Description), a "Description" section with a paragraph of text, and a "Reference" section with a list of links. On the right side, there are two sections: "Linked Datasets" with a checkbox for "DBLP" and "Creator" with a list of links. Two callout boxes point to these sections: "Linked Datasets" points to the "DBLP" checkbox, and "Linked Data from DBLP" points to the "Creator" section.

Linked Datasets

Linked Data from DBLP

# Browsing In and Out of DBpedia with DISCO

The screenshot shows the DISCO Hyperdata Browser interface in a Windows Internet Explorer window. The browser title is "Christian Bizer | Disco - Hyperdata Browser". The address bar shows the URL: [http://localhost:20203-of\\_browser/Dbrowse\\_url=http%3A%2F%2Fwww4.wiwiw4.fu-berlin.de/dblp/resource/person/315759](http://localhost:20203-of_browser/Dbrowse_url=http%3A%2F%2Fwww4.wiwiw4.fu-berlin.de/dblp/resource/person/315759). The main content area displays the resource "Christian Bizer" with a table of properties and values, a list of sources, and a session cache section.

**Label of the displayed resource**: Christian Bizer

**Navigation box**: The URL input field and the "Go!" button.

**Resource description**: The table of properties and values.

**List of all source graphs**: The "Sources" section.

**Link for displaying the session cache**: The "Session Cache" section.

**Go to URI button**: The "Go!" button.

**Sources of each piece of information**: The "Sources" column in the table.

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

**Sources**  
 Displayed information originates from the following RDF graphs:  
 G1 <http://localhost/provenanceinformation#>  
 G2 <http://www4.wiwiw4.fu-berlin.de/dblp/resource/person/315759#>  
 G3 <http://sites.wiwiw4.fu-berlin.de/suhl/bizerfoaf.rdf#>  
 G4 <http://www4.wiwiw4.fu-berlin.de/dblp/resource/record/conf/wirtschaftsinformatik/BizerHMOT05#>  
 G5 <http://www4.wiwiw4.fu-berlin.de/dblp/resource/record/journals/wa/CarrotBHS05#>  
 G6 <http://www4.wiwiw4.fu-berlin.de/dblp/resource/record/conf/wiwi/Bizer04#>  
 G7 <http://www4.wiwiw4.fu-berlin.de/dblp/resource/record/conf/otm/Talk-ado#BEH03#>

**Session Cache**  
 Display all RDF graphs that are currently in your session cache.