DBpedia – Extracting structured data from Wikipedia

Anja Jentzsch, Freie Universität Berlin

Köln. 24. November 2009

DBpedia

- DBpedia 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 open datasets on the Web

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

Outline

1. Extracting Structured Information from Wikipedia
2. The DBpedia Dataset
3. Use Cases
    1. Improving Wikipedia Search
    2. Data Source for other Applications and Mashups
    3. Nucleus for the Emerging Web of Data

Extracting Structured Information from Wikipedia

Title

Description

Languages

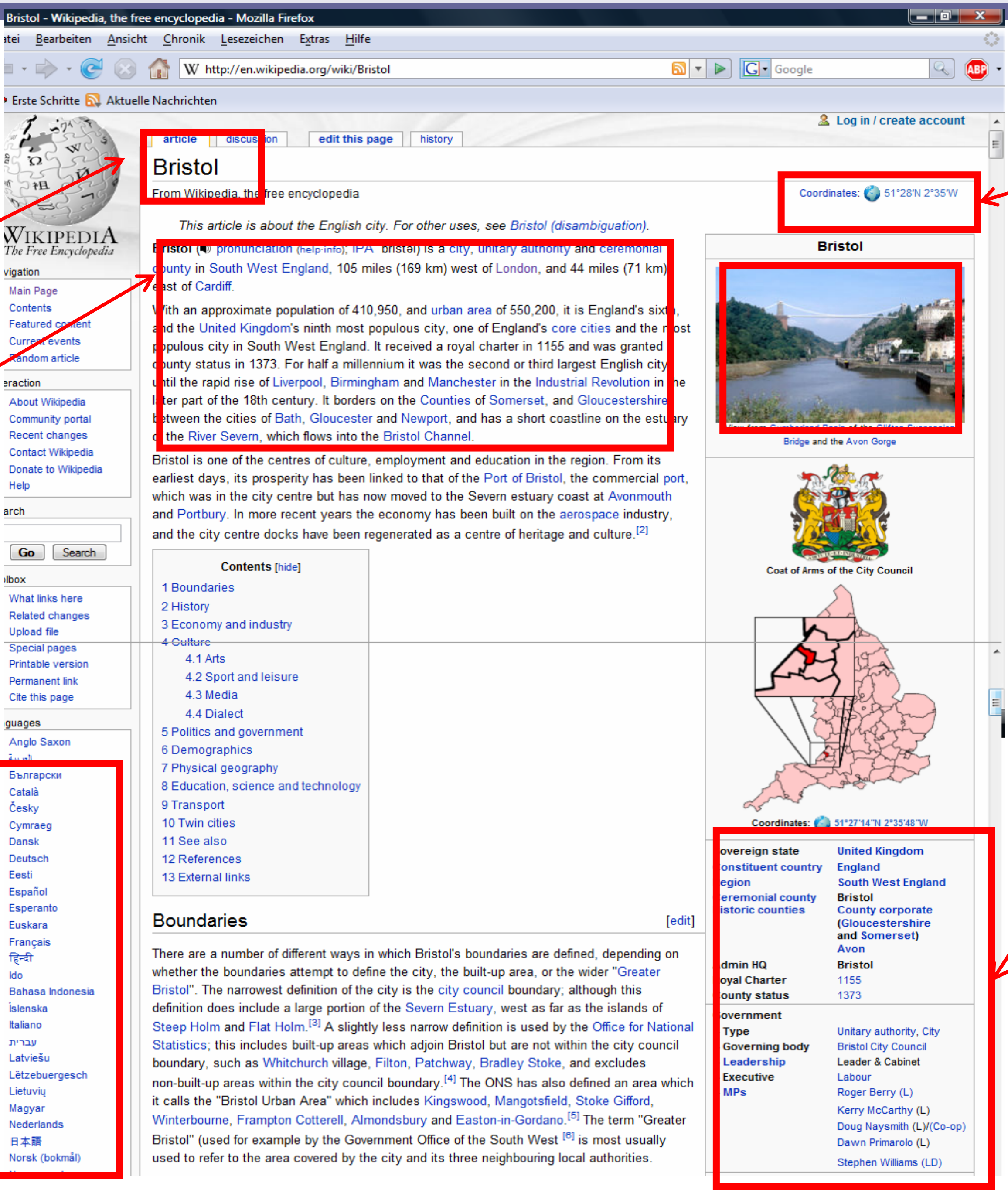
Web Links

Categorization

Domain specific Data

Images

Infoboxes



Bristol - Wikipedia, the free encyclopedia - Mozilla Firefox

Bristol

From Wikipedia, the free encyclopedia

Coordinates: 51°28′N 2°35′W

This article is about the English city. For other uses, see Bristol (disambiguation).

Bristol (🔊 pronunciation (help·info); IPA bristəl) is a city, unitary authority and ceremonial county in South West England, 105 miles (169 km) west of London, and 44 miles (71 km) east of Cardiff.

With an approximate population of 410,950, and urban area of 550,200, it is England's sixth, and the United Kingdom's ninth most populous city, one of England's core cities and the most populous city in South West England. It received a royal charter in 1155 and was granted county status in 1373. For half a millennium it was the second or third largest English city until the rapid rise of Liverpool, Birmingham and Manchester in the Industrial Revolution in the later part of the 18th century. It borders on the Counties of Somerset, and Gloucestershire between the cities of Bath, Gloucester and Newport, and has a short coastline on the estuary of the River Severn, which flows into the Bristol Channel.

Bristol is one of the centres of culture, employment and education in the region. From its earliest days, its prosperity has been linked to that of the Port of Bristol, the commercial port, which was in the city centre but has now moved to the Severn estuary coast at Avonmouth and Portbury. In more recent years the economy has been built on the aerospace industry, and the city centre docks have been regenerated as a centre of heritage and culture.[2]

Contents [hide]

Boundaries

[edit]

There are a number of different ways in which Bristol's boundaries are defined, depending on whether the boundaries attempt to define the city, the built-up area, or the wider "Greater Bristol". The narrowest definition of the city is the city council boundary; although this definition does include a large portion of the Severn Estuary, west as far as the islands of Steep Holm and Flat Holm.[3] A slightly less narrow definition is used by the Office for National Statistics; this includes built-up areas which adjoin Bristol but are not within the city council boundary, such as Whitchurch village, Filton, Patchway, Bradley Stoke, and excludes non-built-up areas within the city council boundary.[4] The ONS has also defined an area which it calls the "Bristol Urban Area" which includes Kingswood, Mangotsfield, Stoke Gifford, Winterbourne, Frampton Cotterell, Almondsbury and Easton-in-Gordano.[5] The term "Greater Bristol" (used for example by the Government Office of the South West [6] is most usually used to refer to the area covered by the city and its three neighbouring local authorities.

Bristol

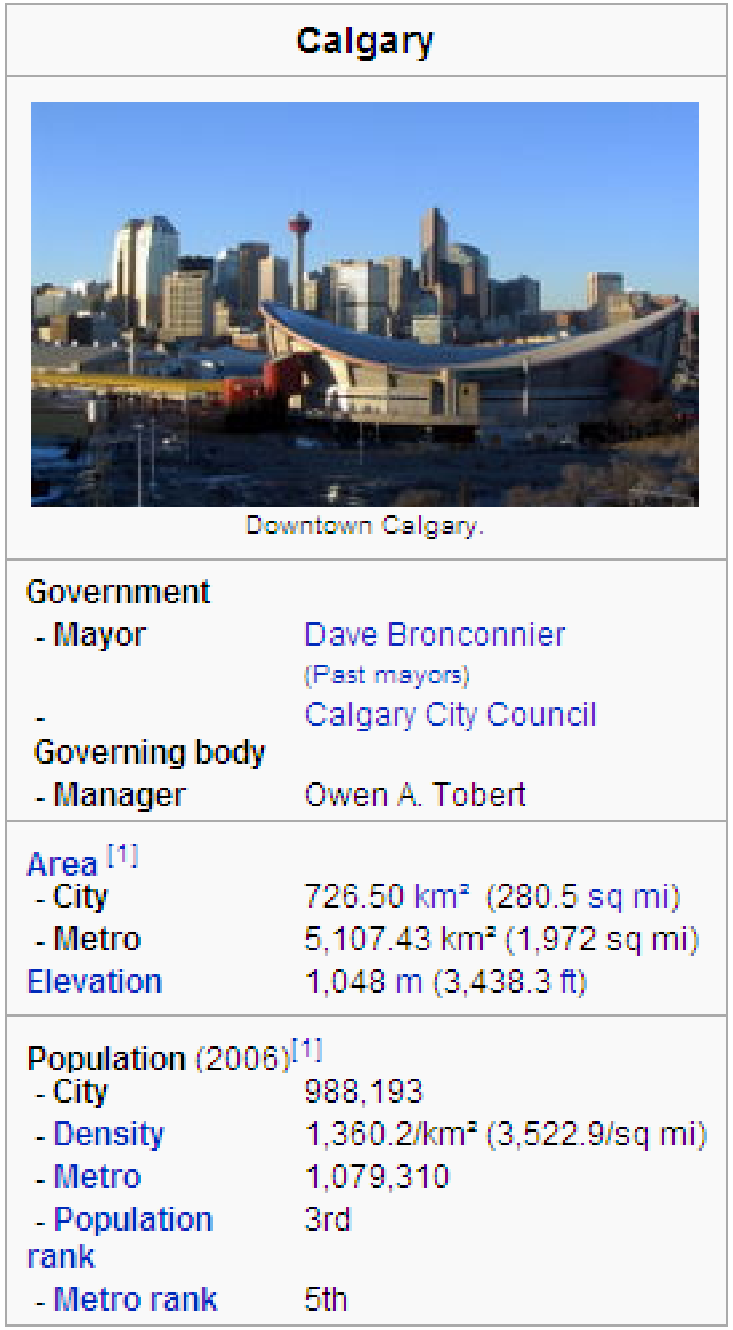
View from Cumberland Basin of the Clifton Suspension Bridge and the Avon Gorge

Coat of Arms of the City Council

Coordinates: 51°27′14″N 2°35′48″W

Sovereign state	United Kingdom
Constituent country	England
Region	South West England
Ceremonial county	Bristol
Historic counties	County corporate (Gloucestershire and Somerset) Avon
Admin HQ	Bristol
Royal Charter	1155
County status	1373
Government	
Type	Unitary authority, City
Governing body	Bristol City Council
Leadership	Leader & Cabinet
Executive	Labour
MPs	Roger Berry (L) Kerry McCarthy (L) Doug Naysmith (L)/(Co-op) Dawn Primarolo (L) Stephen Williams (LD)

Extracting Structured Information from Wikipedia



Calgary

Downtown Calgary.

Government	
- Mayor	Dave Bronconnier (Past mayors)
- Governing body	Calgary City Council
- Manager	Owen A. Tobert
Area [1]	
- City	726.50 km² (280.5 sq mi)
- Metro	5,107.43 km² (1,972 sq mi)
Elevation	1,048 m (3,438.3 ft)
Population (2006)[1]	
- City	988,193
- Density	1,360.2/km² (3,522.9/sq mi)
- Metro	1,079,310
- Population rank	3rd
- Metro rank	5th

```
http://en.wikipedia.org/wiki/Calgary
```

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

- **using a PHP extraction framework**
- **GPL license**

The DBpedia Dataset

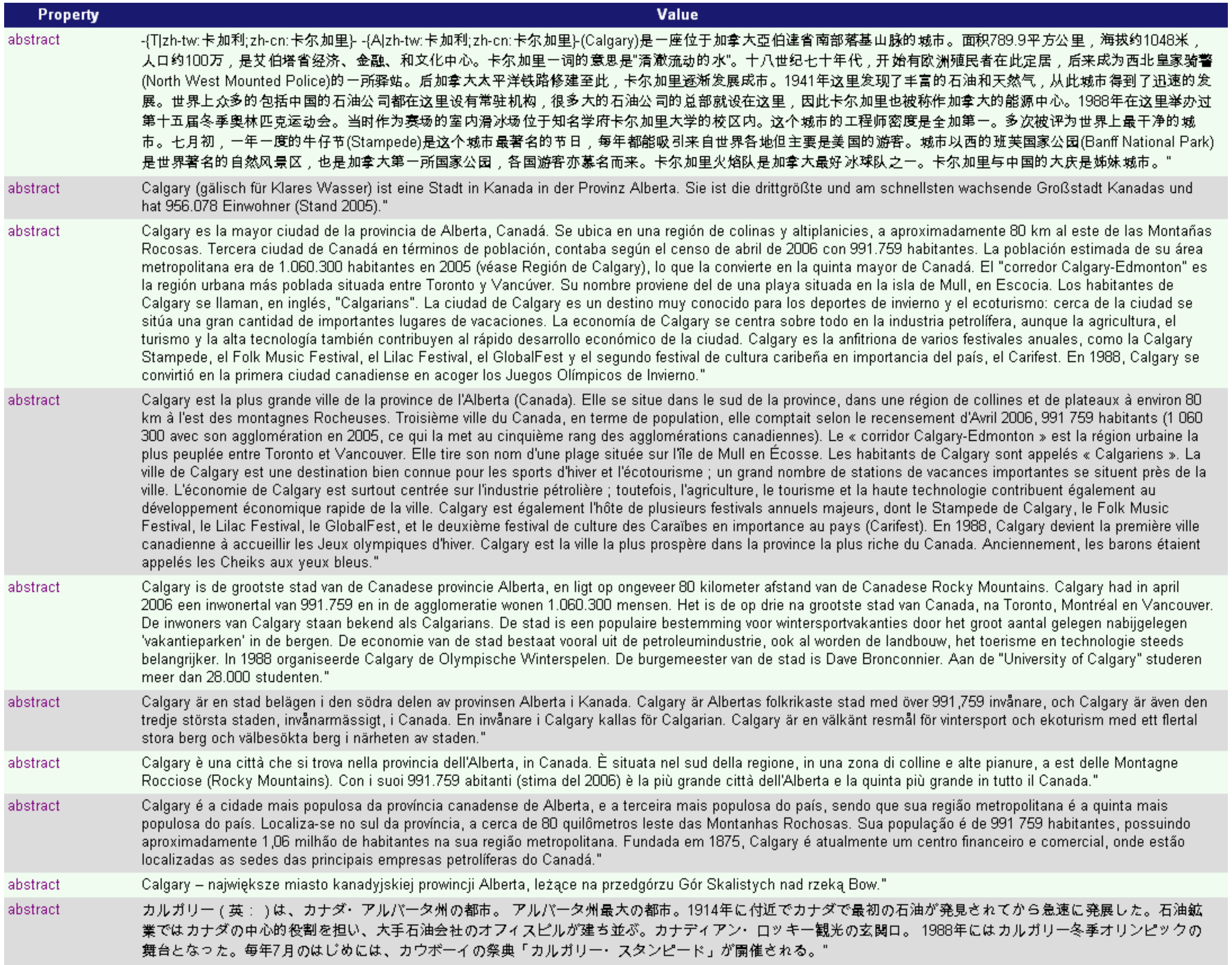
- 91 languages
- Data about 2.9 million "things"
  - 282,000 persons
  - 339,000 places
  - 119,000 organizations
  - 130,000 species
  - 88,000 music albums
  - 44,000 films
  - 19,000 books
- Altogether 479 million pieces of information (RDF triples)
  - 807,000 links to images
  - 3,840,000 links to external web pages
  - 4,878,100 data links into external RDF datasets

The DBpedia Ontology

- Hand-made
- 685 most frequently used templates ↔ 205 ontology classes
- 2800 template properties ↔ 1200 ontology properties
- Currently tracking 1,170,000 resources (articles)
- Great for cross-language fusion
  - e.g. comparing inhabitants of en:London with fr:Paris
    - en:London – settlement infobox
    - fr:Paris – Communes de France

Multi-Lingual Abstracts

- DBpedia dataset contains a short and a long abstract for each concept
- Short abstracts
  - English: 2,943,00
  - French: 500,000
  - German: 466,000
  - Polish: 382,000
  - Dutch: 367,000
  - Italian: 351,000
  - Portuguese: 332,000
  - Spanish: 308,000
  - Japanese: 255,000
  - Swedish: 198,000
  - Chinese: 143,000



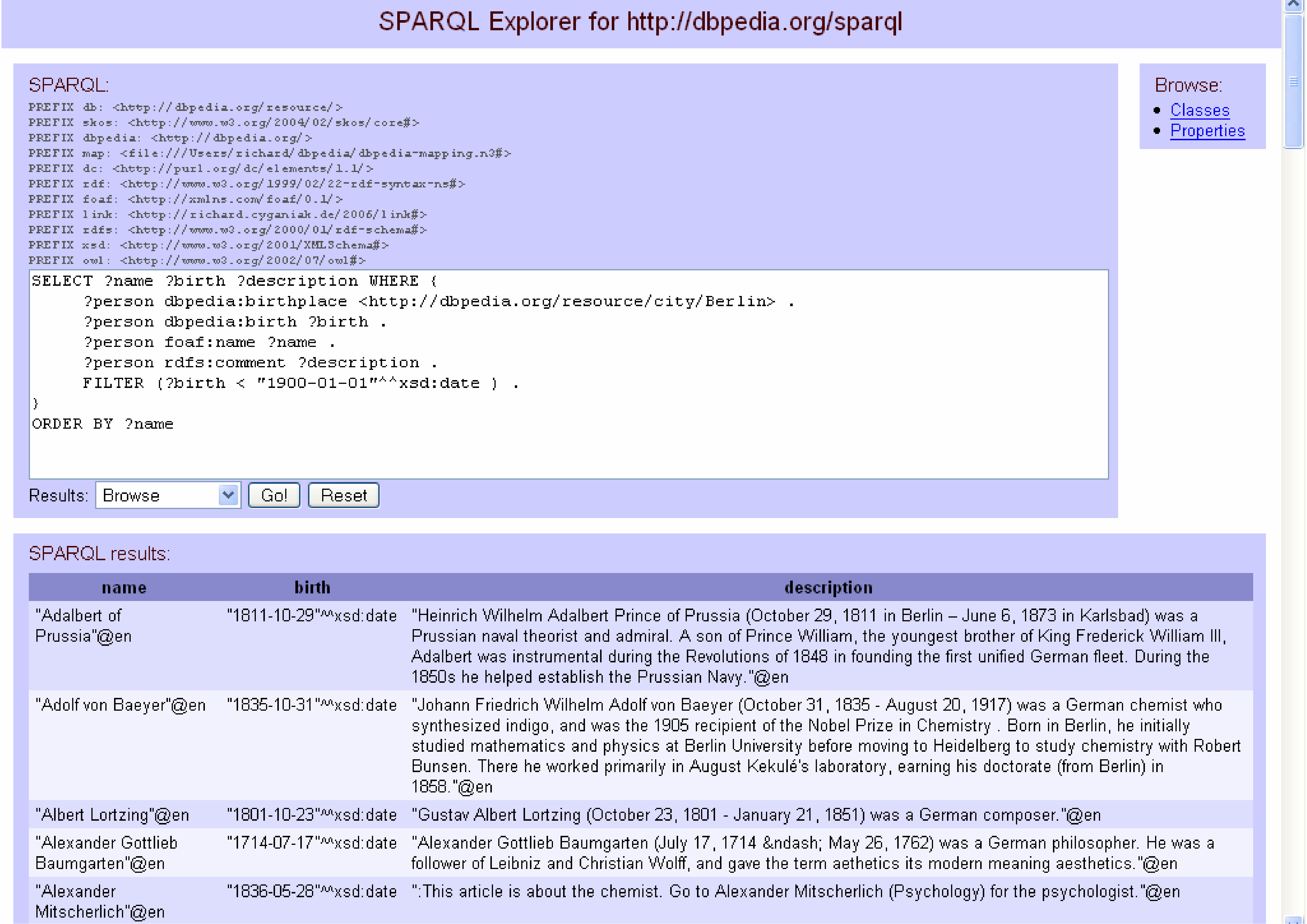
Property	Value
abstract	-{T|zh-tw:卡加利;zh-cn:卡尔加里}- -{A|zh-tw:卡加利;zh-cn:卡尔加里}-(Calgary)是一座位于加拿大亞伯達省南部落基山脉的城市。面积789.9平方公里，海拔约1048米，人口约100万，是艾伯塔省经济、金融、和文化中心。卡尔加里一词的意思是"清澈流动的水"。十八世纪七十年代，开始有欧洲殖民者在此定居，后来成为西北皇家骑警(North West Mounted Police)的一所驿站。后加拿大太平洋铁路修建至此，卡尔加里逐渐发展成市。1941年这里发现了丰富的石油和天然气，从此城市得到了迅速的发展。世界上众多的包括中国的石油公司都在这里设有常驻机构，很多大的石油公司的总部就设在这里，因此卡尔加里也被称作加拿大的能源中心。1988年在这里举办过第十五届冬季奥林匹克运动会。当时作为赛场的室内滑冰场位于知名学府卡尔加里大学的校区内。这个城市的工程师密度是全加第一。多次被评为世界上最干净的城市。七月初，一年一度的牛仔节(Stampede)是这个城市最著名的节日，每年都能吸引来自世界各地但主要是美国的游客。城市以西的班芙国家公园(Banff National Park)是世界著名的自然风景区，也是加拿大第一所国家公园，各国游客亦慕名而来。卡尔加里火焰队是加拿大最好冰球队之一。卡尔加里与中国的大庆是姊妹城市。"
abstract	Calgary (gälisch für Klares Wasser) ist eine Stadt in Kanada in der Provinz Alberta. Sie ist die drittgrößte und am schnellsten wachsende Großstadt Kanadas und hat 956.078 Einwohner (Stand 2005)."
abstract	Calgary es la mayor ciudad de la provincia de Alberta, Canadá. Se ubica en una región de colinas y altiplanicies, a aproximadamente 80 km al este de las Montañas Rocosas. Tercera ciudad de Canadá en términos de población, contaba según el censo de abril de 2006 con 991.759 habitantes. La población estimada de su área metropolitana era de 1.060.300 habitantes en 2005 (véase Región de Calgary), lo que la convierte en la quinta mayor de Canadá. El "corredor Calgary-Edmonton" es la región urbana más poblada situada entre Toronto y Vancúver. Su nombre proviene del de una playa situada 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 petrolífera, aunque la agricultura, el turismo y la alta tecnología también contribuyen al rápido desarrollo económico de la ciudad. Calgary es la anfitriona de varios festivales anuales, como la Calgary Stampede, el Folk Music Festival, el Lilac Festival, el GlobalFest y el segundo festival de cultura caribeña en importancia del país, el Carifest. En 1988, Calgary se convirtió en la primera ciudad canadiense en acoger los Juegos Olímpicos de Invierno."
abstract	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 80 km à l'est des montagnes Rocheuses. Troisième ville du Canada, en terme de population, elle comptait selon le recensement d'Avril 2006, 991 759 habitants (1 060 300 avec son agglomération en 2005, ce qui la met au cinquième rang des agglomérations canadiennes). Le « corridor 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 « Calgariens ». 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 ; toutefois, l'agriculture, le tourisme et la haute technologie contribuent 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 GlobalFest, et le deuxième festival de culture des Caraïbes en importance au pays (Carifest). En 1988, Calgary devient 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 barons étaient appelés les Cheiks aux yeux bleus."
abstract	Calgary is de grootste stad van de Canadese provincie Alberta, en ligt op ongeveer 80 kilometer afstand van de Canadese Rocky Mountains. Calgary had in april 2006 een inwonertal van 991.759 en in de agglomeratie wonen 1.060.300 mensen. Het is de op drie na grootste stad van Canada, na Toronto, Montréal en Vancouver. De inwoners van Calgary staan bekend als Calgarians. De stad is een populaire bestemming voor wintersportvakanties door het groot aantal gelegen nabijgelegen 'vakantieparken' 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 1988 organiseerde Calgary de Olympische Winterspelen. De burgemeester van de stad is Dave Bronconnier. Aan de "University of Calgary" studeren meer dan 28.000 studenten."
abstract	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, invånarmässigt, i Canada. En invånare i Calgary kallas för Calgarian. Calgary är en välkänt resmål för vintersport och ekoturism med ett flertal stora berg och välbesökta berg i närheten av staden."
abstract	Calgary è una 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 Rocciose (Rocky Mountains). Con i suoi 991.759 abitanti (stima del 2006) è la più grande città dell'Alberta e la quinta più grande in tutto il Canada."
abstract	Calgary é a cidade mais populosa da província 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 província, a cerca de 80 quilômetros leste das Montanhas Rochosas. Sua população é de 991 759 habitantes, possuindo aproximadamente 1,06 milhão de habitantes na sua região metropolitana. Fundada em 1875, Calgary é atualmente um centro financeiro e comercial, onde estão localizadas as sedes das principais empresas petrolíferas do Canadá."
abstract	Calgary – największe miasto kanadyjskiej prowincji Alberta, leżące na przedgórzu Gór Skalistych nad rzeką Bow."
abstract	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。1914年に付近でカナダで最初の石油が発見されてから急速に発展した。石油鉱業ではカナダの中心的役割を担い、大手石油会社のオフィスビルが建ち並ぶ。カナディアン・ロッキー観光の玄関口。 1988年にはカルガリー冬季オリンピックの舞台となった。毎年7月のはじめには、カウボーイの祭典「カルガリー・スタンピード」が開催される。"

DBpedia Use Cases

1. Improving Wikipedia Search
2. Data Source for other Applications and Mashups
3. Nucleus for the Emerging Web of Data

1. Improving Wikipedia Search

- DBpedia SPARQL Endpoint: http://dbpedia.org/sparql
- can answer SPARQL queries like
  - Give me all books written by authors that were born in Berlin in the 19th century?
  - All Rivers that flow into the Rhine and are longer than 50 kilometers?
  - All Actors of the American TV-series Lost that were born in England?
  - All tennis players from Moscow?
  - All soccer players with tricot number 11, playing for a club having a stadium with over 40,000 seats and is born in a country with over 10 million inhabitants?



SPARQL Explorer for http://dbpedia.org/sparql

SPARQL:

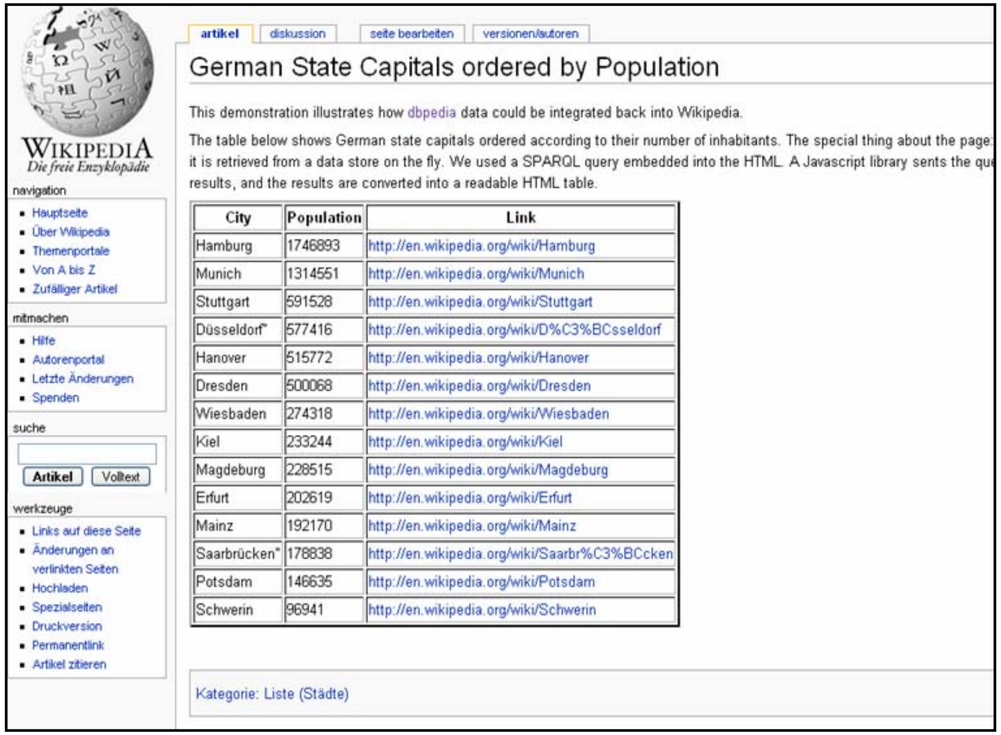
```
PREFIX db: <http://dbpedia.org/resource/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX map: <file:///Users/richard/dbpedia/dbpedia-mapping.n3#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX link: <http://richard.cyganiak.de/2006/link#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT ?name ?birth ?description WHERE {
     ?person dbpedia:birthplace <http://dbpedia.org/resource/city/Berlin> .
     ?person dbpedia:birth ?birth .
     ?person foaf:name ?name .
     ?person rdfs:comment ?description .
     FILTER (?birth < "1900-01-01"^^xsd:date ) .
}
ORDER BY ?name
```

SPARQL results:

name	birth	description
"Adalbert of Prussia"@en	"1811-10-29"^^xsd:date	"Heinrich Wilhelm Adalbert Prince of Prussia (October 29, 1811 in Berlin – June 6, 1873 in Karlsbad) was a Prussian naval theorist and admiral. A son of Prince William, the youngest brother of King Frederick William III, Adalbert was instrumental during the Revolutions of 1848 in founding the first unified German fleet. During the 1850s he helped establish the Prussian Navy."@en
"Adolf von Baeyer"@en	"1835-10-31"^^xsd:date	"Johann Friedrich Wilhelm Adolf von Baeyer (October 31, 1835 - August 20, 1917) was a German chemist who synthesized indigo, and was the 1905 recipient of the Nobel Prize in Chemistry . Born in Berlin, he initially studied mathematics and physics at Berlin University before moving to Heidelberg to study chemistry with Robert Bunsen. There he worked primarily in August Kekulé's laboratory, earning his doctorate (from Berlin) in 1858."@en
"Albert Lortzing"@en	"1801-10-23"^^xsd:date	"Gustav Albert Lortzing (October 23, 1801 - January 21, 1851) was a German composer."@en
"Alexander Gottlieb Baumgarten"@en	"1714-07-17"^^xsd:date	"Alexander Gottlieb Baumgarten (July 17, 1714 &ndash; May 26, 1762) was a German philosopher. He was a follower of Leibniz and Christian Wolff, and gave the term aethetics its modern meaning aesthetics."@en
"Alexander Mitscherlich"@en	"1836-05-28"^^xsd:date	":This article is about the chemist. Go to Alexander Mitscherlich (Psychology) for the psychologist."@en

2. Royalty-Free Data Source for other Applications

- DBpedia is published under GNU Free Documentation License
- Example use case: SPARQL generated tables within webpages



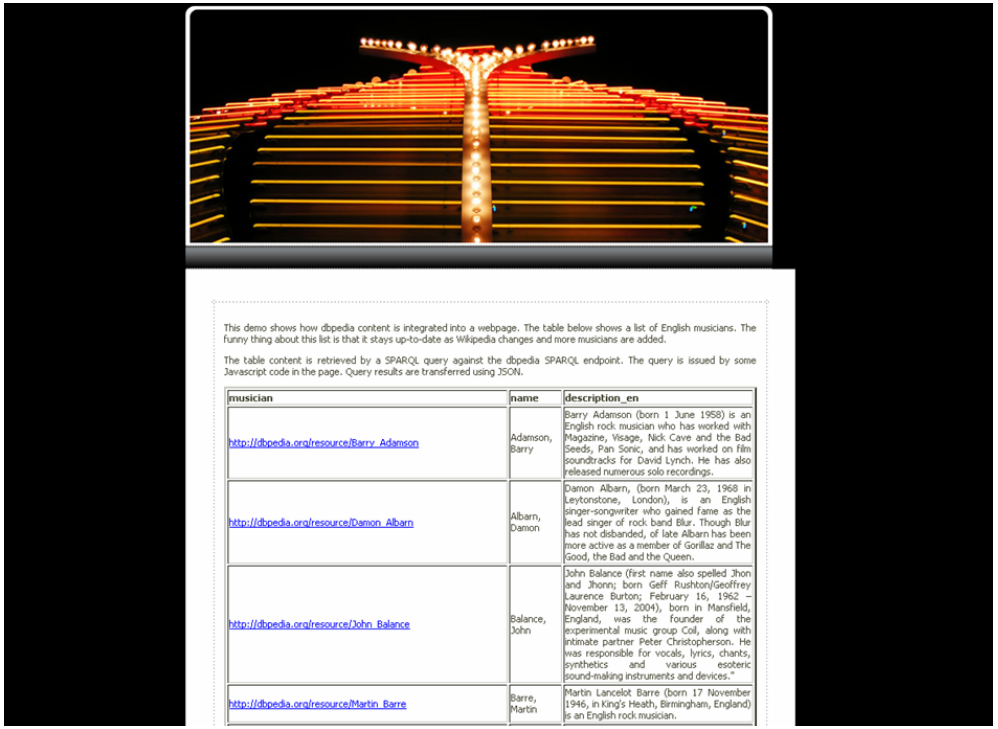
German State Capitals ordered by Population

This demonstration illustrates how dbpedia data could be integrated back into Wikipedia.

The table below shows German state capitals ordered according to their number of inhabitants. The special thing about the page: it is retrieved from a data store on the fly. We used a SPARQL query embedded into the HTML. A Javascript library sends the que results, and the results are converted into a readable HTML table.

City	Population	Link
Hamburg	1746893	http://en.wikipedia.org/wiki/Hamburg
Munich	1314551	http://en.wikipedia.org/wiki/Munich
Stuttgart	591528	http://en.wikipedia.org/wiki/Stuttgart
Düsseldorf	577416	http://en.wikipedia.org/wiki/D%C3%BCsseldorf
Hanover	515772	http://en.wikipedia.org/wiki/Hanover
Dresden	500068	http://en.wikipedia.org/wiki/Dresden
Wiesbaden	274318	http://en.wikipedia.org/wiki/Wiesbaden
Kiel	233244	http://en.wikipedia.org/wiki/Kiel
Magdeburg	228515	http://en.wikipedia.org/wiki/Magdeburg
Erfurt	202619	http://en.wikipedia.org/wiki/Erfurt
Mainz	192170	http://en.wikipedia.org/wiki/Mainz
Saarbrücken	178838	http://en.wikipedia.org/wiki/Saarbr%C3%BCcken
Potsdam	146635	http://en.wikipedia.org/wiki/Potsdam
Schwerin	96941	http://en.wikipedia.org/wiki/Schwerin

Kategorie: Liste (Städte)



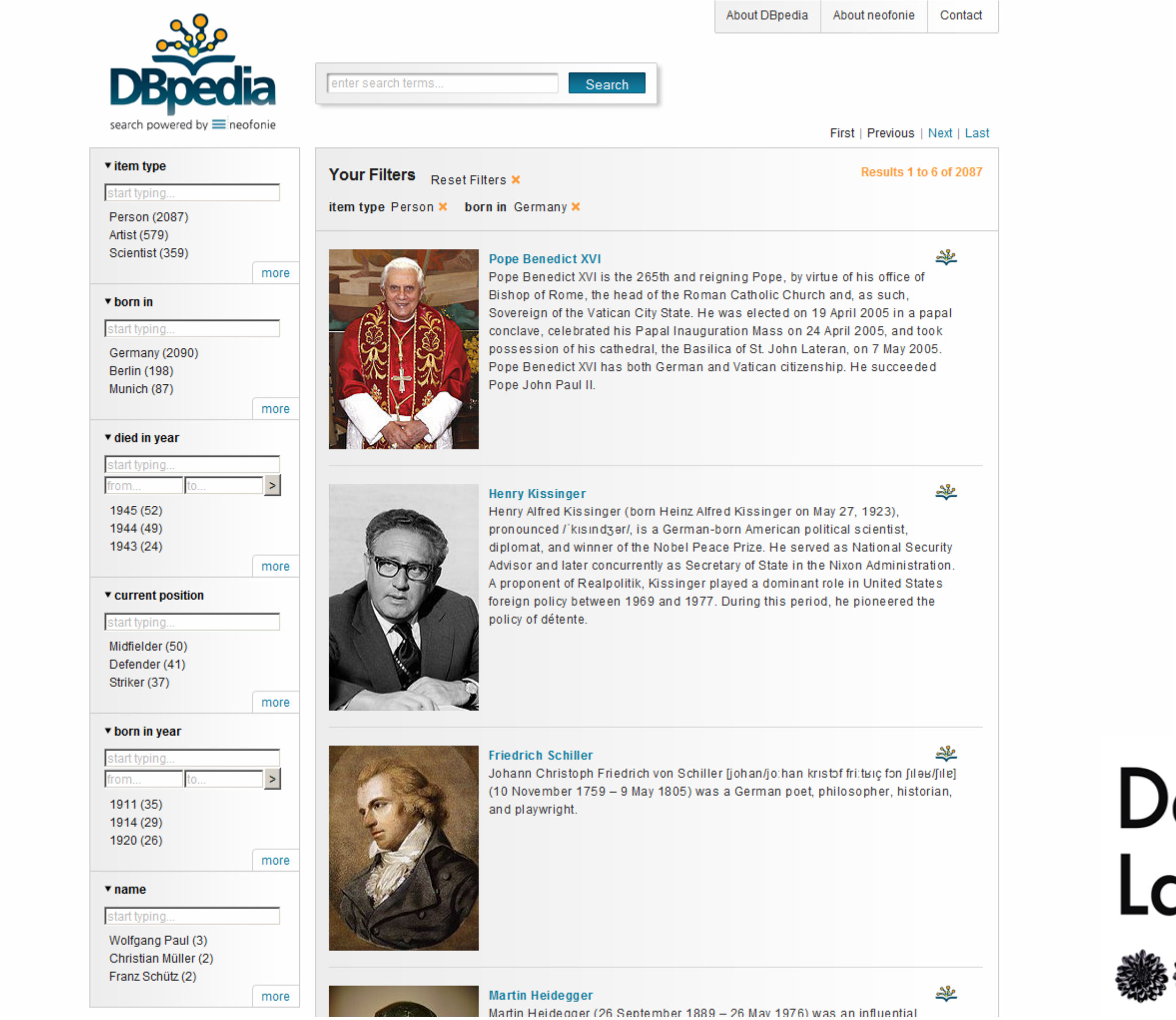
This demo shows how dbpedia content is integrated into a webpage. The table below shows a list of English musicians. The funny thing about this list is that it stays up-to-date as Wikipedia changes and more musicians are added.

The table content is retrieved by a SPARQL query against the dbpedia SPARQL endpoint. The query is issued by some Javascript code in the page. Query results are transferred using JSON.

musician	name	description_en
http://dbpedia.org/resource/Barry_Adamson	Adamson, Barry	Barry Adamson (born 1 June 1958) is an English rock musician who has worked with Magazine, Visage, Nick Cave and the Bad Seeds, Pan Sonic, and has worked on film soundtracks for David Lynch. He has also released numerous solo recordings.
http://dbpedia.org/resource/Damon_Albarn	Albarn, Damon	Damon Albarn, (born March 23, 1968 in Leytonstone, London), is an English singer-songwriter who gained fame as the lead singer of rock band Blur. Though Blur has not disbanded, of late Albarn has been more active as a member of Gorillaz and The Good, the Bad and the Queen.
http://dbpedia.org/resource/John_Balance	Balance, John	John Balance (first name also spelled Jhon and Jhonn; born Geff Rushton/Geoffrey Laurence Burton; February 16, 1962 – November 13, 2004), born in Mansfield, England, was the founder of the experimental music group Coil, along with intimate partner Peter Christopherson. He was responsible for vocals, lyrics, chants, synthetics and various esoteric sound-making instruments and devices."
http://dbpedia.org/resource/Martin_Barre	Barre, Martin	Martin Lancelot Barre (born 17 November 1946, in King's Heath, Birmingham, England) is an English rock musician.

Faceted Browsing Interface

- Cooperation with Neofonie (Berlin search engine company)
- Faceted Browsing and Free-Text Search



DBpedia
search powered by neofonie

About DBpedia | About neofonie | Contact

enter search terms... Search

First | Previous | Next | Last

▾ item type
start typing...
Person (2087)
Artist (579)
Scientist (359)
more

▾ born in
start typing...
Germany (2090)
Berlin (198)
Munich (87)
more

▾ died in year
start typing...
from... to... >
1945 (52)
1944 (49)
1943 (24)
more

▾ current position
start typing...
Midfielder (50)
Defender (41)
Striker (37)
more

▾ born in year
start typing...
from... to... >
1911 (35)
1914 (29)
1920 (26)
more

▾ name
start typing...
Wolfgang Paul (3)
Christian Müller (2)
Franz Schütz (2)
more

Your Filters Reset Filters ×

Results 1 to 6 of 2087

item type Person × born in Germany ×

Pope Benedict XVI
Pope Benedict XVI is the 265th and reigning Pope, by virtue of his office of Bishop of Rome, the head of the Roman Catholic Church and, as such, Sovereign of the Vatican City State. He was elected on 19 April 2005 in a papal conclave, celebrated his Papal Inauguration Mass on 24 April 2005, and took possession of his cathedral, the Basilica of St. John Lateran, on 7 May 2005. Pope Benedict XVI has both German and Vatican citizenship. He succeeded Pope John Paul II.

Henry Kissinger
Henry Alfred Kissinger (born Heinz Alfred Kissinger on May 27, 1923), pronounced /ˈkɪsɪndʒər/, is a German-born American political scientist, diplomat, and winner of the Nobel Peace Prize. He served as National Security Advisor and later concurrently as Secretary of State in the Nixon Administration. A proponent of Realpolitik, Kissinger played a dominant role in United States foreign policy between 1969 and 1977. During this period, he pioneered the policy of détente.

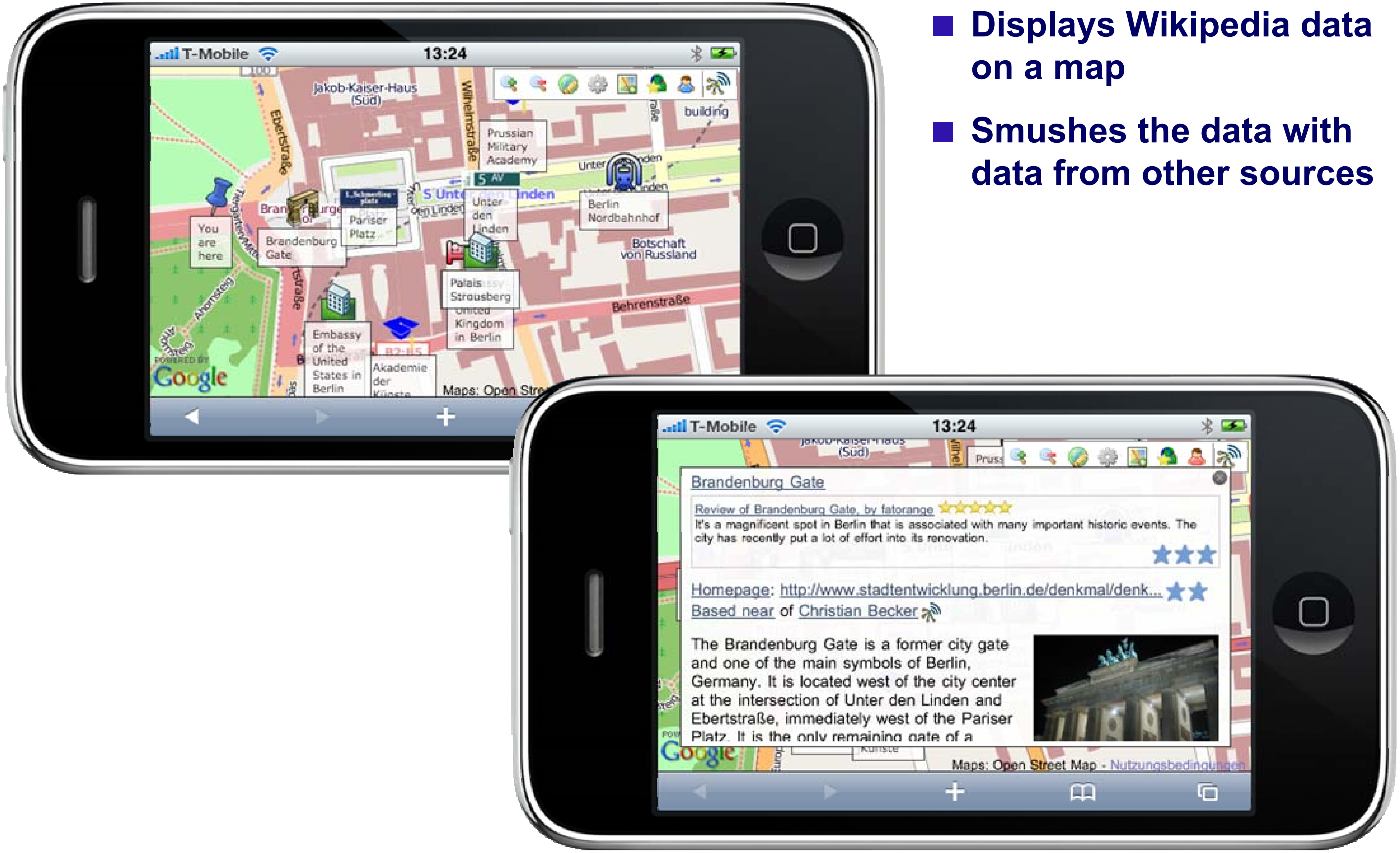
Friedrich Schiller
Johann Christoph Friedrich von Schiller [johan/joːhan krɪstɔf friːtʁɪç fɔn ʃɪləʁ/ʃɪlɐ] (10 November 1759 – 9 May 1805) was a German poet, philosopher, historian, and playwright.

Martin Heidegger
Martin Heidegger (26 September 1889 – 26 May 1976) was an influential

Deutschland
Land der Ideen

DBpedia Mobile

- Displays Wikipedia data on a map
- Smushes the data with data from other sources



Brandenburg Gate

Review of Brandenburg Gate, by fatorange

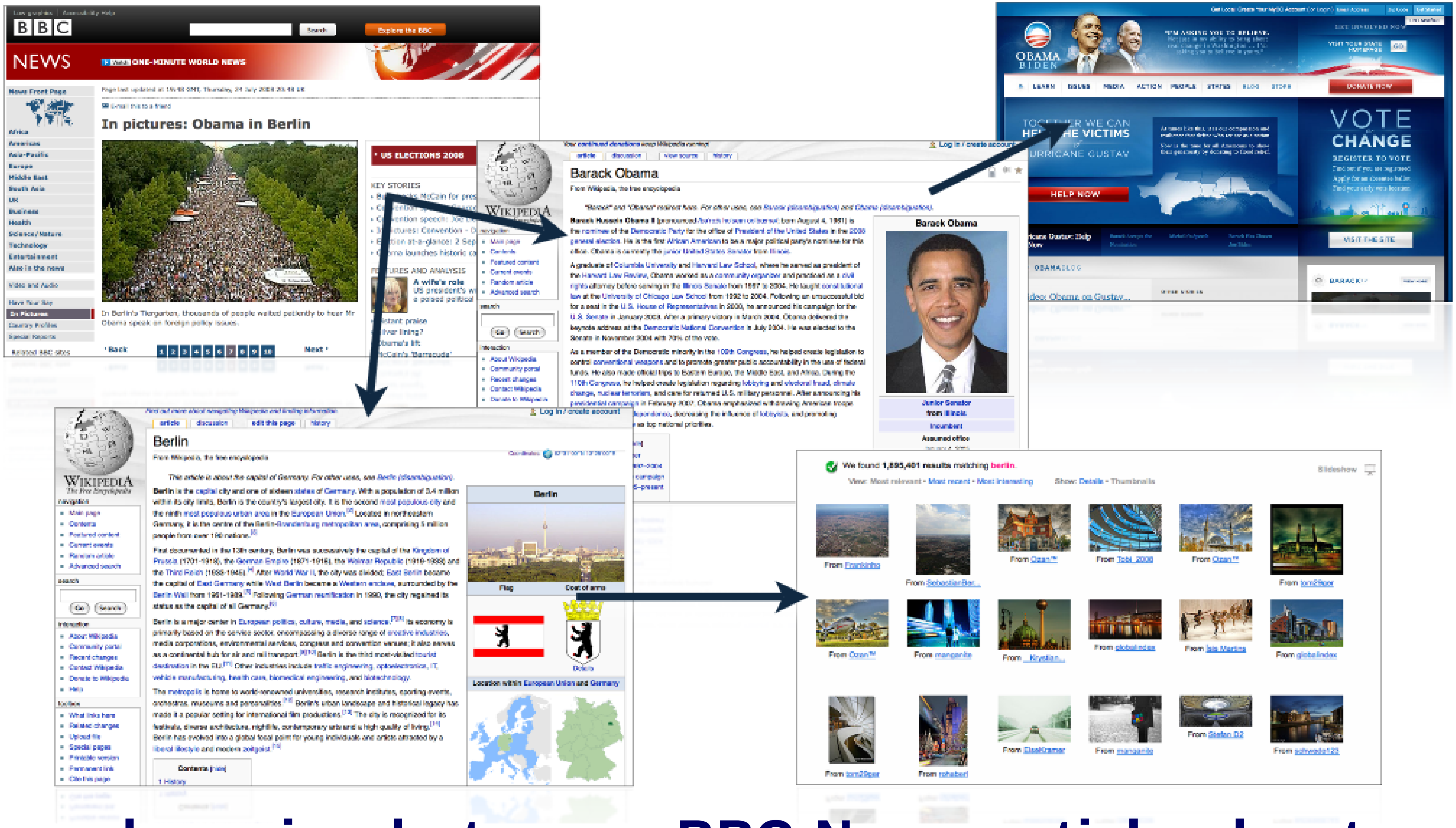
It's a magnificent spot in Berlin that is associated with many important historic events. The city has recently put a lot of effort into its renovation.

Homepage: http://www.stadtentwicklung.berlin.de/denkmal/denk...

Based near of Christian Becker

The Brandenburg Gate is a former city gate and one of the main symbols of Berlin, Germany. It is located west of the city center at the intersection of Unter den Linden and Ebertstraße, immediately west of the Pariser Platz. It is the only remaining gate of a

Example: Integrating Wikipedia content into BBC News

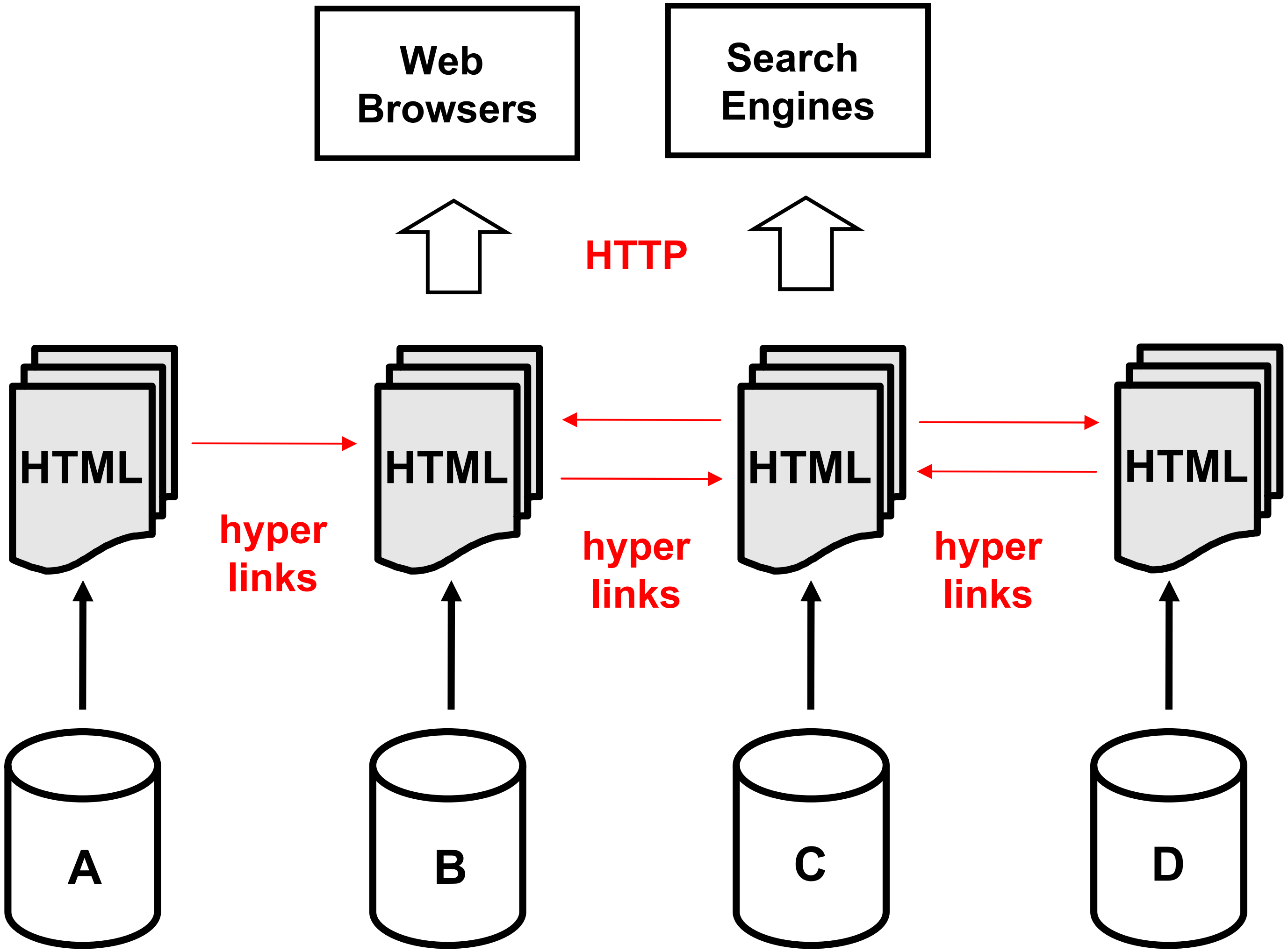

BBC NEWS
In pictures: Obama in Berlin
US ELECTIONS 2008
Barack Obama
From Wikipedia, the free encyclopedia
Berlin
From Wikipedia, the free encyclopedia
OBAMA BIDEN
VOTE CHANGE
HELP NOW
We found 1,895,401 results matching berlin.

Based on simple tags, a BBC News article about Barack Obama in Berlin could automatically import Berlin photos and Weblinks for Barack Obama

3. Nucleus for the Emerging Web of Data

The Web of Documents

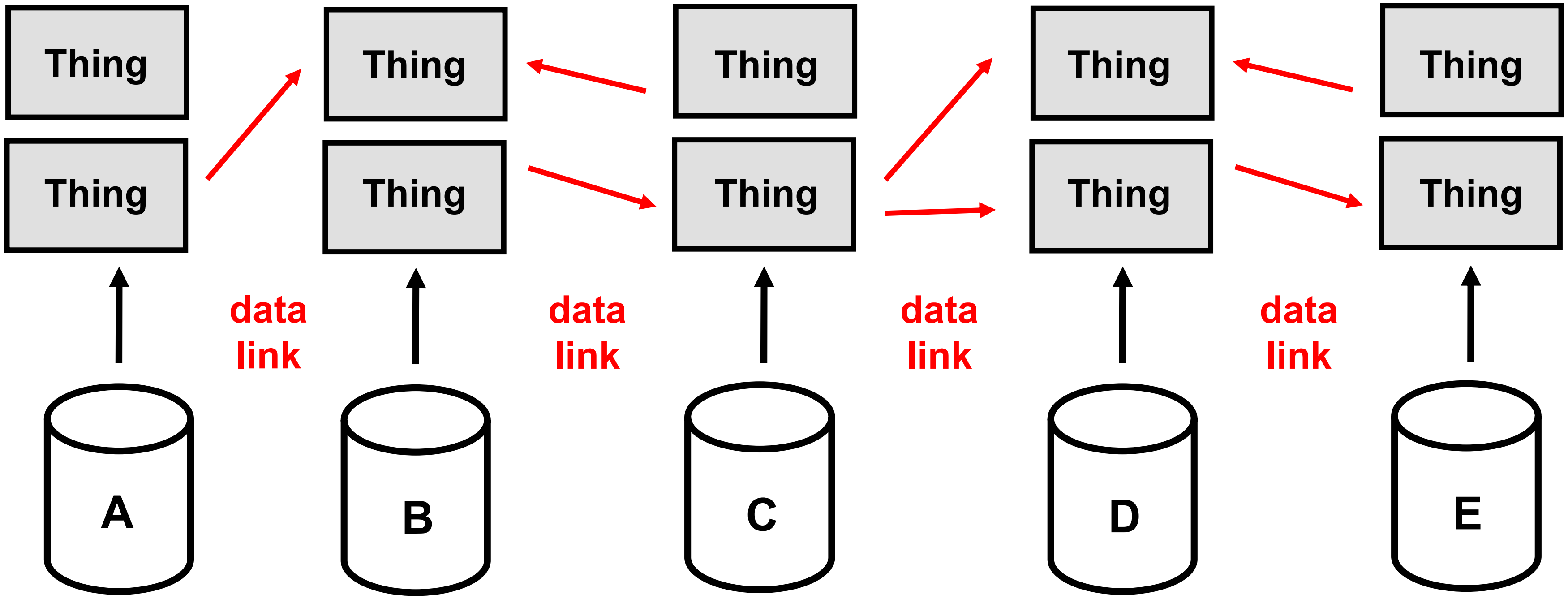
The Web is a single information space build on open standards and hyperlinks.



Linked Data

Use RDF and HTTP to

1. publish structured data on the Web,
2. set data links between data from one data source to data within other data sources.

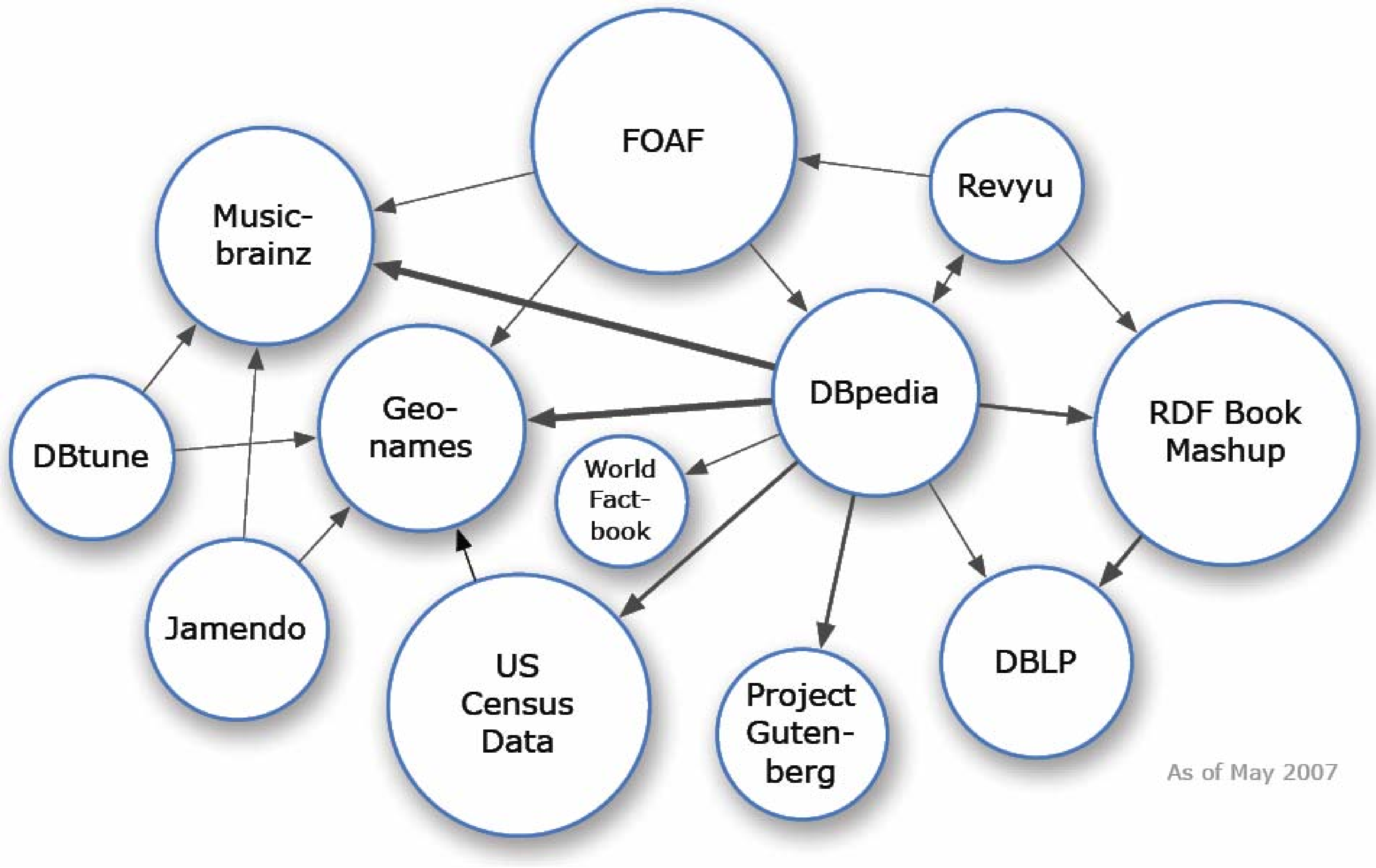


W3C Linking Open Data Project

LINKINGOPENDATA
W3C SWEO Community Project

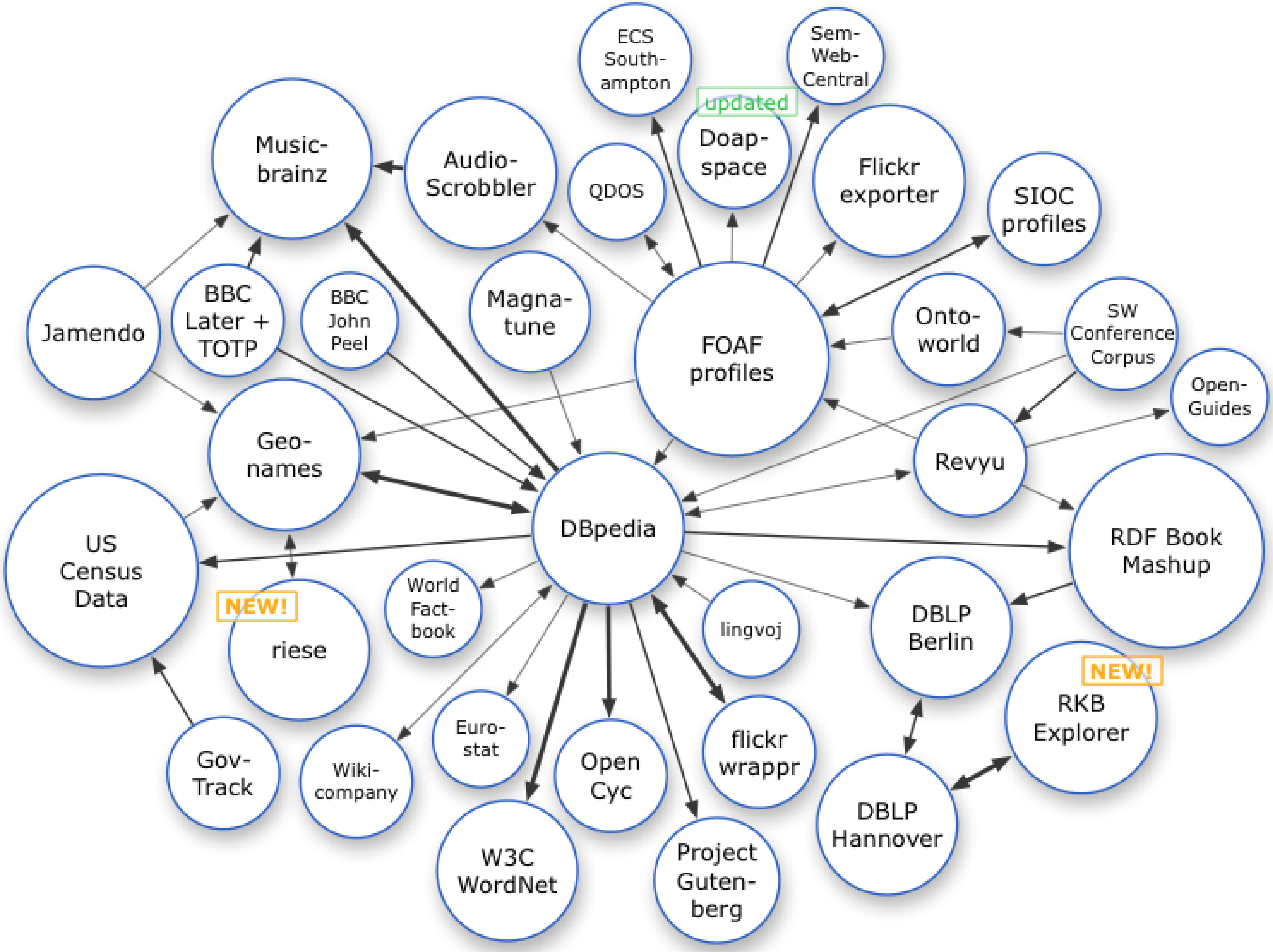
- Community effort to
  - publish existing open license datasets as Linked Data on the Web
  - interlink things between different data sources

LOD Datasets on the Web: May 2007



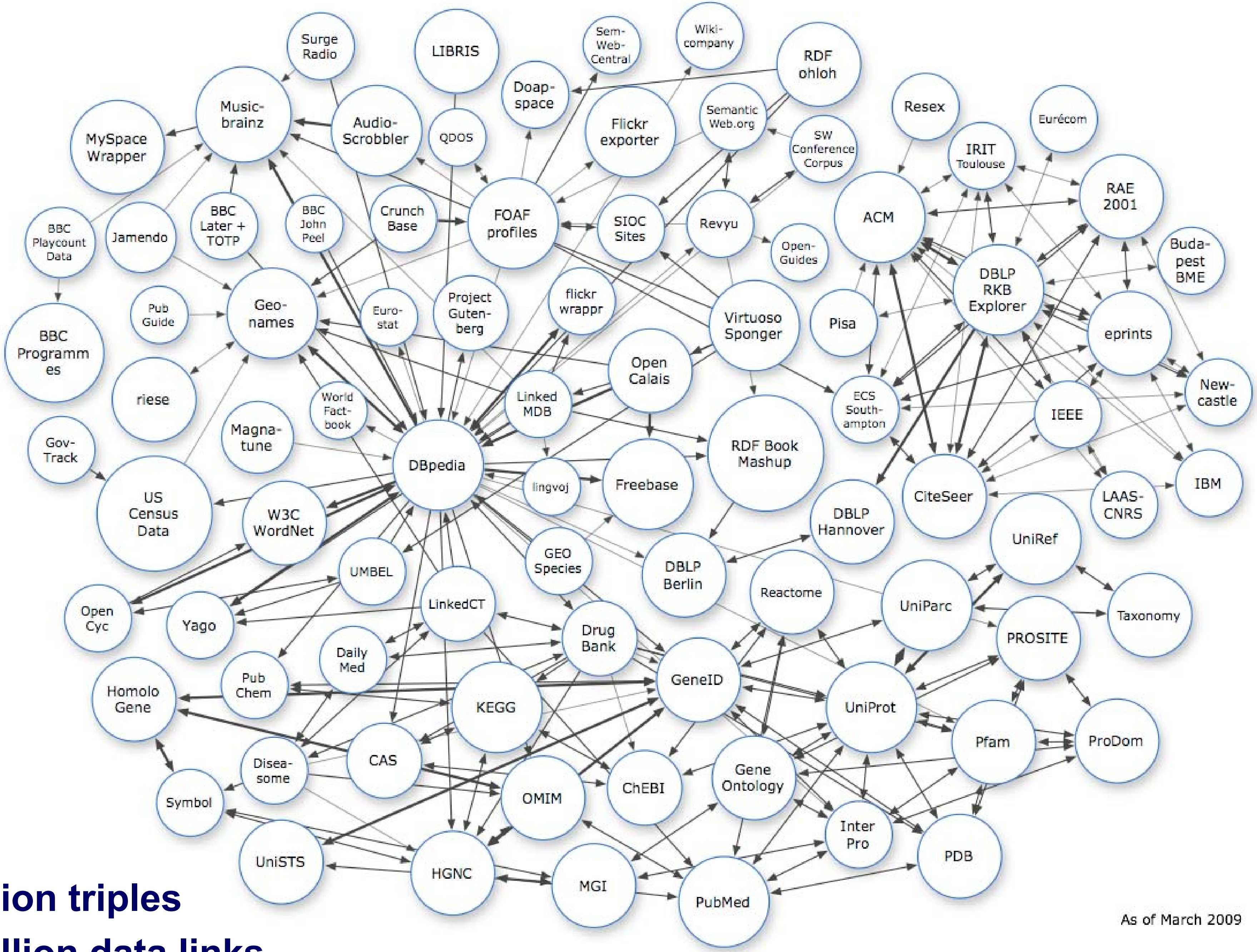
- Over 500 million RDF triples.

LOD Datasets on the Web: April 2008



- Over 2 billion RDF triples.

LOD Datasets on the Web: March 2009

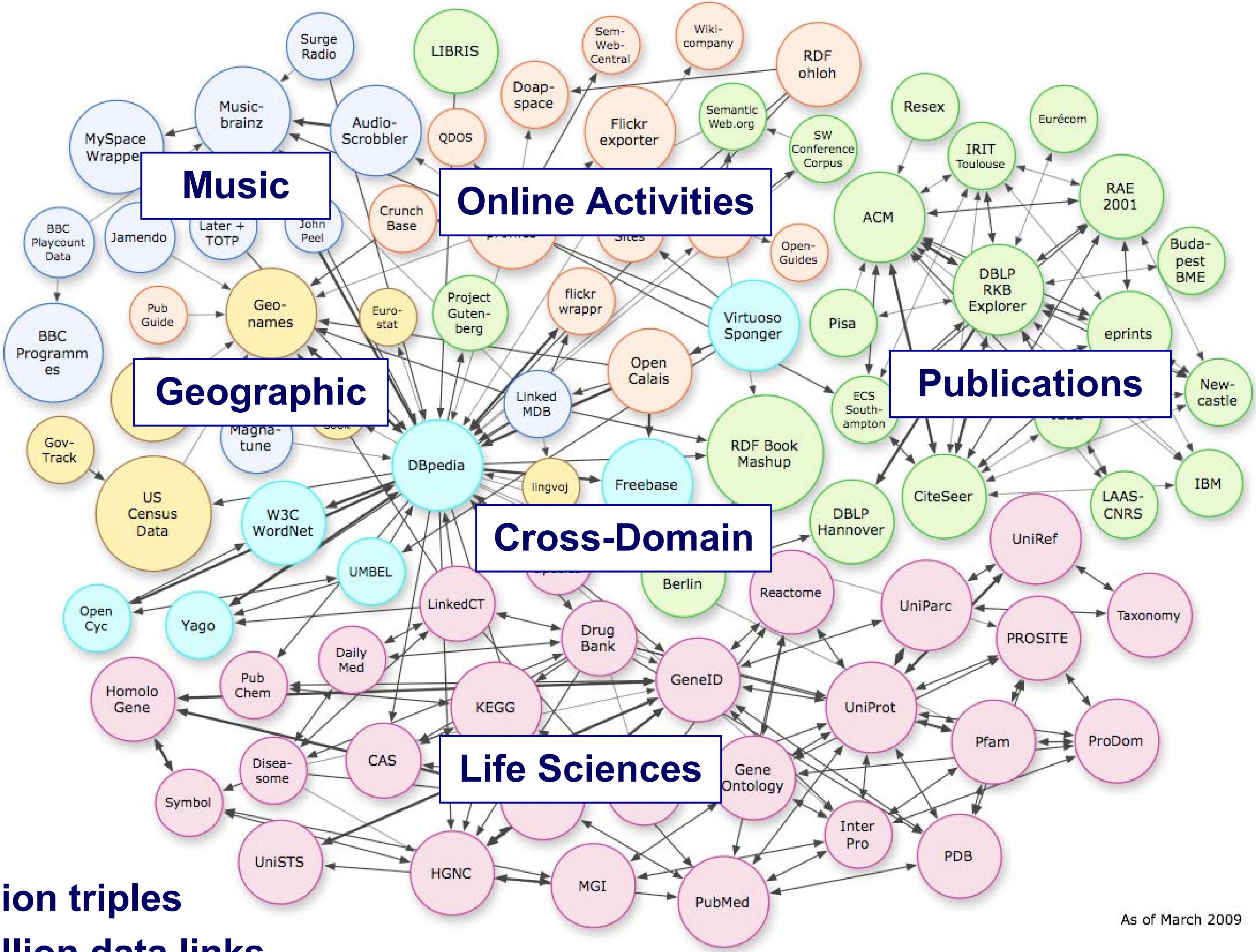


4.5 billion triples

180 million data links

As of March 2009

LOD Datasets on the Web: March 2009



As of March 2009

4.5 billion triples

180 million data links

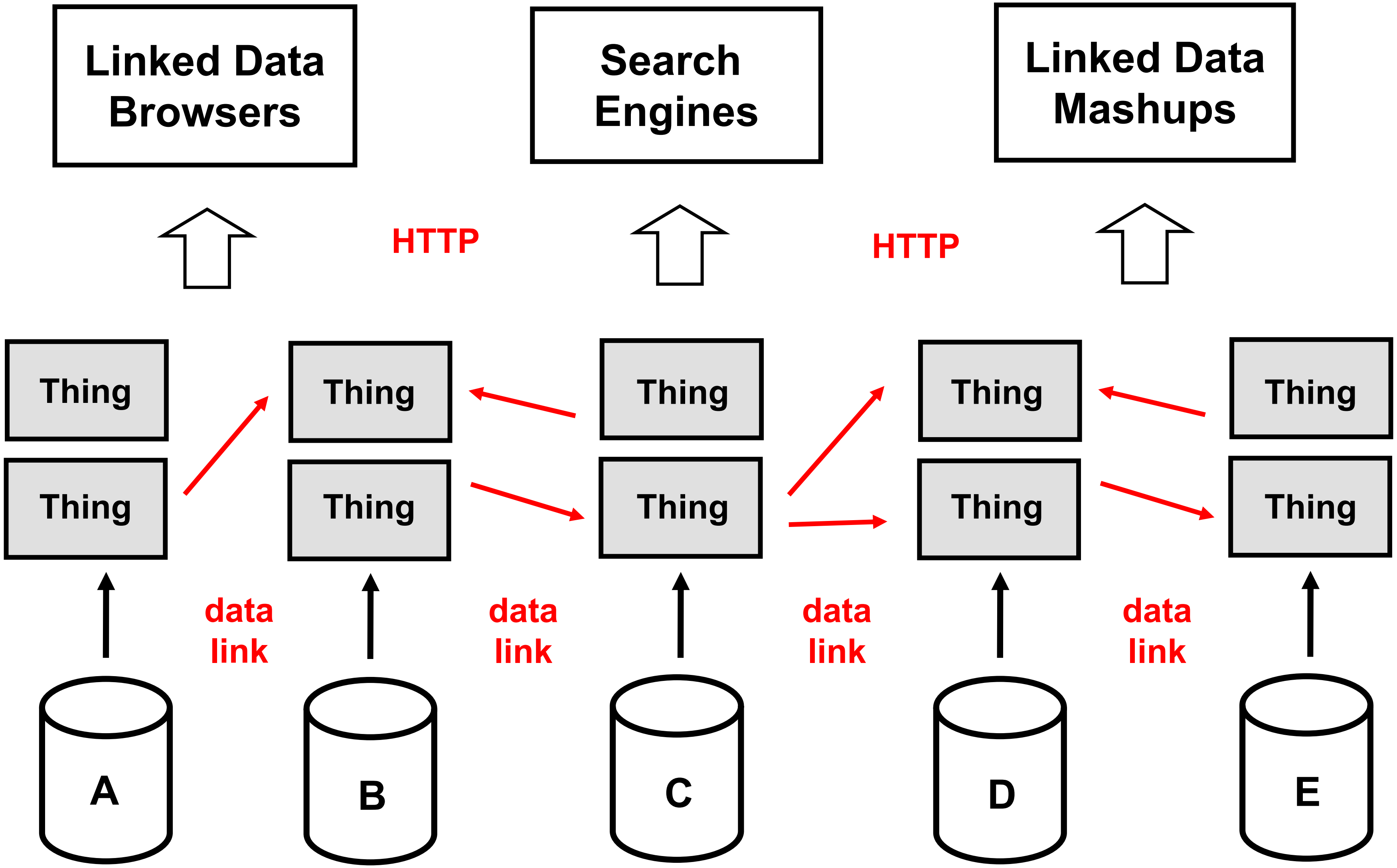
Linked Data Adoption

- **Who's publishing / supporting Linked Data? (excerpt)**
  - BBC
  - Cyc Foundation
  - Freebase
  - MusicBrainz
  - Thomson Reuters
  - UK Ordnance Survey
  - Zemanta
  - Life Sciences Community
  - New York Times
- **Yahoo! and Google have started to crawl Linked Data in its RDFa serialization as well as Microformats.**

DBpedia interlinks (exemplary)

- MusicBrainz
- OpenStreetMap (via LinkedGeoData.org)
- BBC Programmes, Music
- Semantic CrunchBase
- Geonames
- WordNet
- World Factbook
- EuroStat
- Flickr (via flickrwrappr)
- FreeBase
- OpenCyc
- US Census
- Amazon, Google Base (via Book Mashup)
- Dailymed, Diseasome, Drugbank, Sider

What can I do with this?



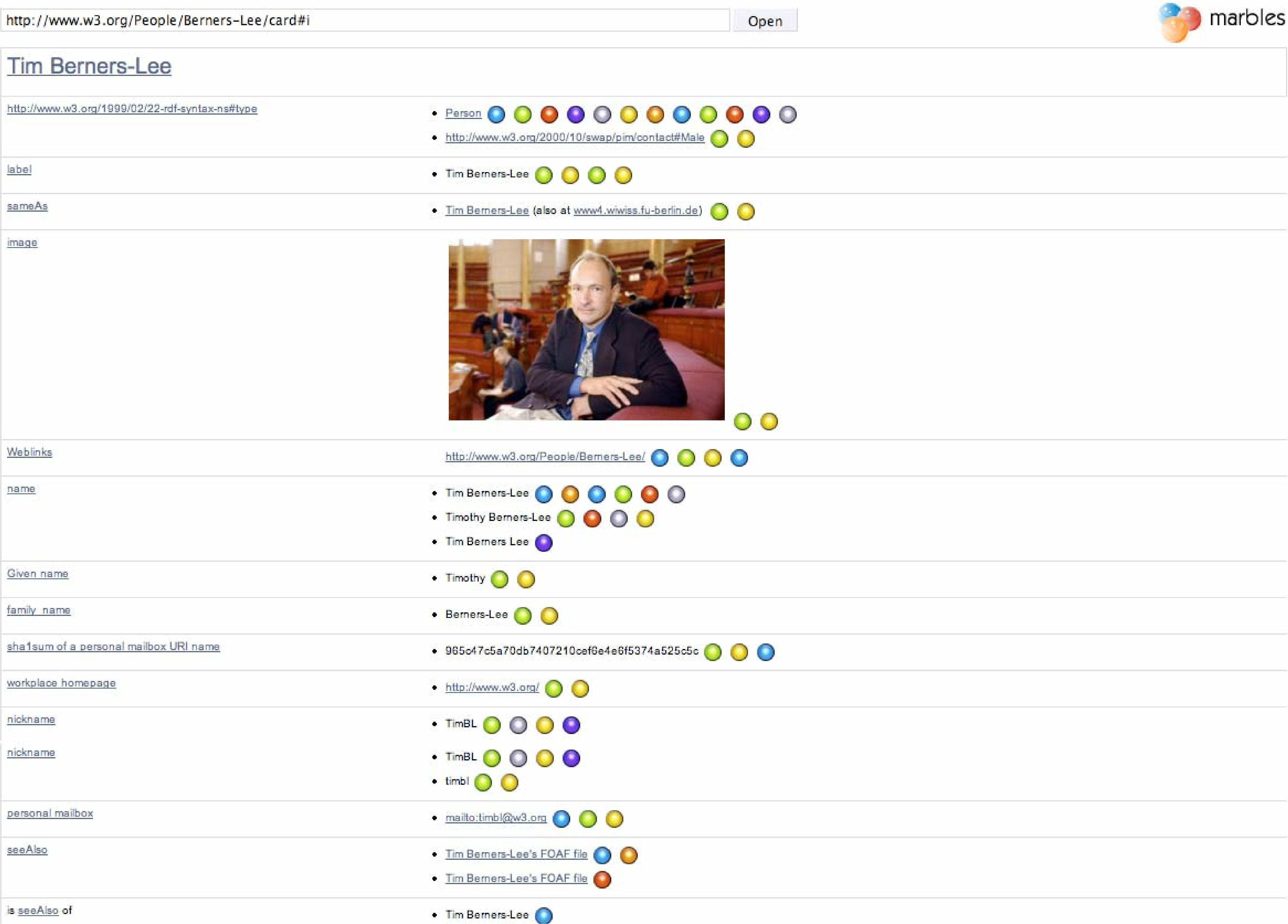
Accessing the DBpedia dataset

- Linked Data
  - e.g. http://dbpedia.org/resource/London
- SPARQL Interface
  - http://dbpedia.org/sparql
- Download RDF Dumps
  - http://wiki.dbpedia.org/Downloads
- Amazon Public Datasets

Linked Data Interface

- The project follows the Linked Data principles
  - All concepts are identified using URI references
  - All URIs are dereferencable over the Web into a small RDF snippet

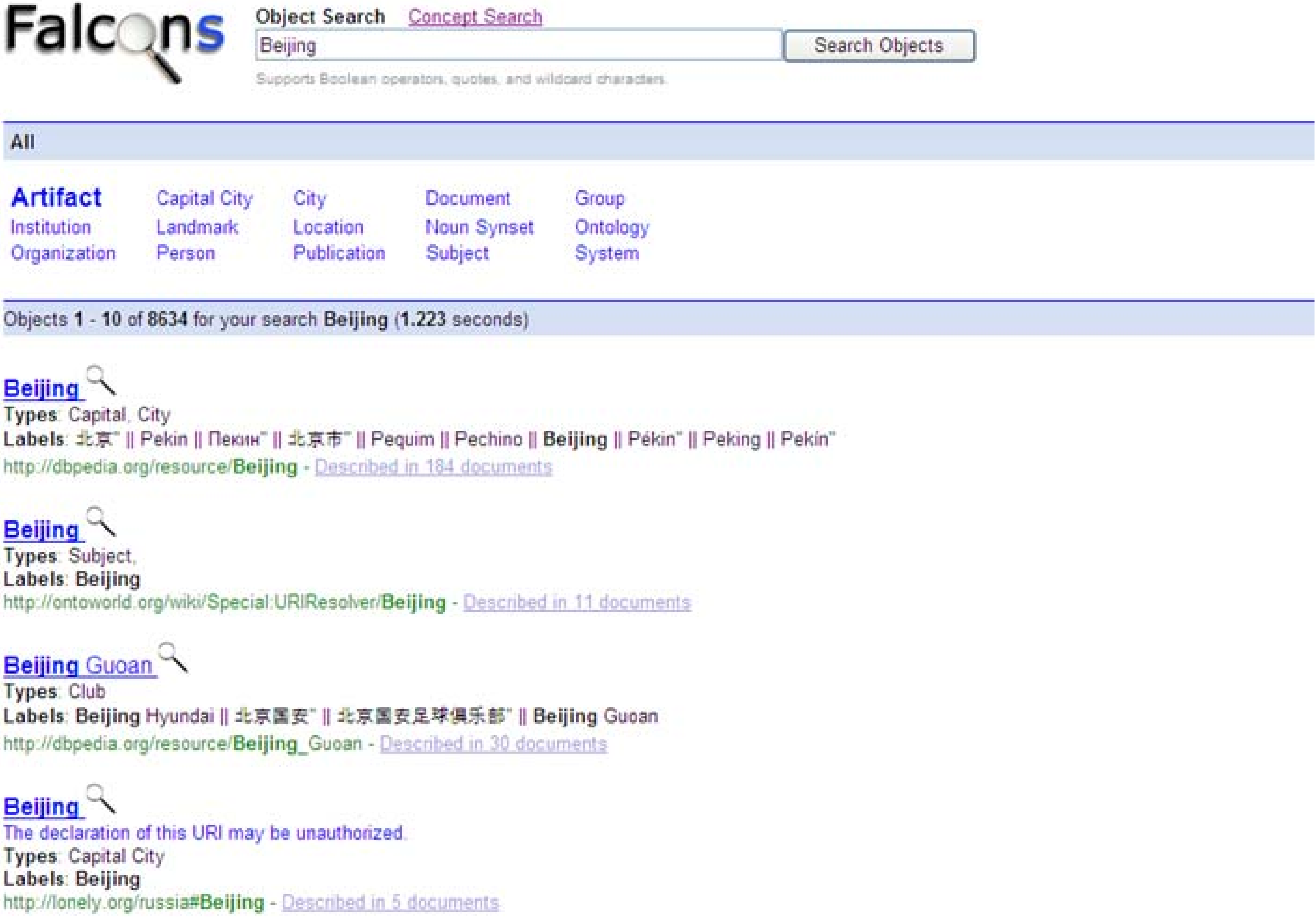
- The Linked Data interface can be used by
  - Semantic Web Browsers, like
    - Tabulator Browser
    - Marbles
    - OpenLink RDF Browser
  - Semantic Web Crawlers, like
    - Zitgist (Zitgist LLC, USA)
    - SWSE (DERI, Ireland)
    - Swoogle (UMBC, USA)



http://www.w3.org/People/Berners-Lee/card#i

Tim Berners-Lee

http://www.w3.org/1999/02/22-rdf-syntax-ns#type	Person 🔵 🟢 🔴 🟣 ⚪ 🟡 🟠 🔵 🟢 🔴 🟣 ⚪ http://www.w3.org/2000/10/swap/pim/contact#Male 🟢 🟡
label	Tim Berners-Lee 🟢 🟡 🟢 🟡
sameAs	Tim Berners-Lee (also at www4.wiwiss.fu-berlin.de) 🟢 🟡
image	🟢 🟡
Weblinks	http://www.w3.org/People/Berners-Lee/ 🔵 🟢 🟡 🔵
name	Tim Berners-Lee 🔵 🟠 🔵 🟢 🔴 ⚪ Timothy Berners-Lee 🟢 🔴 ⚪ 🟡 Tim Berners Lee 🟣
Given name	Timothy 🟢 🟡
family name	Berners-Lee 🟢 🟡
sha1sum of a personal mailbox URI name	965c47c5a70db7407210cef6e4e6f5374a525c5c 🟢 🟡 🔵
workplace homepage	http://www.w3.org/ 🟢 🟡
nickname	TimBL 🟢 ⚪ 🟡 🟣
nickname	TimBL 🟢 ⚪ 🟡 🟣 timbl 🟢 🟡
personal mailbox	mailto:timbl@w3.org 🔵 🟢 🟡
seeAlso	Tim Berners-Lee's FOAF file 🔵 🟠 Tim Berners-Lee's FOAF file 🔴
is seeAlso of	Tim Berners-Lee 🔵



Objects **1** - **10** of **8634** for your search **Beijing** (**1.223** seconds)

Beijing
Types: Capital, City
Labels: 北京" || Pekin || Пекин" || 北京市" || Pequim || Pechino || **Beijing** || Pékin" || Peking || Pekín"
http://dbpedia.org/resource/**Beijing** - Described in 184 documents

Beijing
Types: Subject,
Labels: **Beijing**
http://ontoworld.org/wiki/Special:URIResolver/**Beijing** - Described in 11 documents

Beijing Guoan
Types: Club
Labels: **Beijing** Hyundai || 北京国安" || 北京国安足球俱乐部" || **Beijing** Guoan
http://dbpedia.org/resource/**Beijing**_Guoan - Described in 30 documents

Beijing
The declaration of this URI may be unauthorized.
Types: Capital City
Labels: **Beijing**
http://lonely.org/russia#**Beijing** - Described in 5 documents

3. What is next for DBpedia?

Improve the Quality of Extracted Data

- Problem
  - chaotic usage of infoboxes within Wikipedia
- Solution
  - smarter version of the infobox extractor
  - smushes multiple properties with the same meaning
  - smushes different infoboxes for the same class
  - uses knowledge about property ranges
  - generates a cleaner class hierarchy
- Status
  - First release of the DBpedia “Ontology” in November 2008
  - Still improve the mappings and extraction code

Cross-Language Data Fusion

- Opportunity
  - there are 264 Wikipedia Editions in different languages
  - there are cross-language links
  - the Italian Wikipedia knows more about Italian villages then the English one
  - the German Wikipedia contains more person infoboxes than the English one
- Idea
  - Augment the infobox dataset with facts from other Wikipedia editions
- Result
  - A much richer DBpedia dataset

Augment DBpedia with Data from External Sources

Opportunity

- the Linking Open Data cloud provides lots of useful data which is not contained in Wikipedia yet
- For instance:
  - EuroStat provides additional statistical information about countries
  - Musicbrainz contains additional information about other bands
  - Geonames provides additional information about locations

Idea

- Augment DBpedia with additional data from external sources

Result

- A much richer DBpedia dataset

Live Update

- **Current Situation**
  - DBpedia update cycle: 2-3 months
  - Wikipedia provides us with access to the live update stream
- **Opportunity**
  - Increase the frequency of the DBpedia dataset using this update stream
  - Move ontology schema definition and mappings to Wikipedia
- **Result**
  - DBpedia in synchronization with Wikipedia
  - DBpedia as a semantic mirror of Wikipedia
  - Wikipedia community can edit the ontology and mappings

Contribute back to the Wikipedia Community

Opportunity

- augmentation with data from the LOD cloud makes the DBpedia dataset richer than Wikipedia itself
- infobox data is extracted from Wikipedia editions in various languages

Idea

- Extend the Wikipedia authoring environment with
  - Suggestions for infobox values
  - Cross-language consistency checking for infoboxes

Initialize Wikipedia Clean-Up Cycles

- Data-driven search interfaces expose the weaknesses of Wikipedia template system
- Preferred items not showing up in end-user interfaces may motivate Wikipedia editors to use templates more stringently

Thanks!

References

- DBpedia
  http://dbpedia.org/About
- DBpedia Faceted Browser
  http://dbpedia.neofonie.de
- W3C Linking Open Data Project
  http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
- Tutorial: How to Publish Linked Data on the Web
  http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/