

**European Geodemographics Conference
London, April 1, 2009**

Linking Spatial Data from the Web

Christian Becker, Freie Universität Berlin

Hello

■ Name

- Christian Becker

■ Job

- Partner, MES (consulting)
- PhD Student at Freie Universität Berlin

■ Semantic Web Projects

- DBpedia and DBpedia Mobile
- Marbles Browser
- BBC interlinking project
- D2RQ and D2R Server
- flickr™ wrappr

Overview

- 1. Linked Data**
- 2. DBpedia as a Geospatial Entrypoint to the Web of (Linked) Data**
- 3. Implications of Linked Data for Geodemographics**

Linked Data

We live in a world of data

■ Everything around us is online

- Things: Wikipedia, Maps, Review sites
- People: Facebook, LinkedIn, Wikipedia

■ Many things are uniquely identified:

- Every Wikipedia article
- Every street corner
- Every social network profile

Belgrave Square
From Wikipedia, the free encyclopedia

Coordinates: 51°49′N 0°15′W﻿ / ﻿51.817°N 0.250°W﻿ / 51.817; -0.250

Belgrave Square is one of the grandest 19th-century squares in London, England. Bordering Knightsbridge, it is the largest of Belgravia, and was laid out by contractor Thomas Cubitt for the Grosvenor, later the 1st Marquess of Westminster, in the 1820s. Most of the houses were occupied by 1840. The square takes its name from one of the Duke of Westminster's subsidiary titles, Viscount Belgrave. The village of **Belgravia**, **Cheshire** is two miles (3 km) from the Grosvenor family's main country seat of Eaton Hall.

The original scheme consisted of four terraces, each made up of eleven grand white stuccoed houses, apart from the south-east terrace, which has twelve; detached mansions in three of the corners; and a private central garden. The numbering is clockwise from the north: NW terrace Nos. 1 to 11; west corner mansion No. 12; SW terrace 12-23; south corner mansion No. 24; SE terrace Nos. 25-36; east corner mansion No. 37; NE

The north east side of Belgrave Square soon after construction.

Statue of Simon Bolivar

Way
11895634

	secondary road	▼
highway	secondary	✕
name	Belgrave Square	✕
oneway	yes	✕

Christian Becker

Owner at MediaEvent Services GmbH & Co. KG
Berlin Area, Germany | Computer Software

Current

- Owner at MediaEvent Services GmbH & Co. KG

Past

- Consultant XBRL at Umanis France SA
- Developer Mobile at VerdiSoft (Yahoo! Go)

Education

- Freie Universität Berlin
- Université Paris X Nanterre

Websites

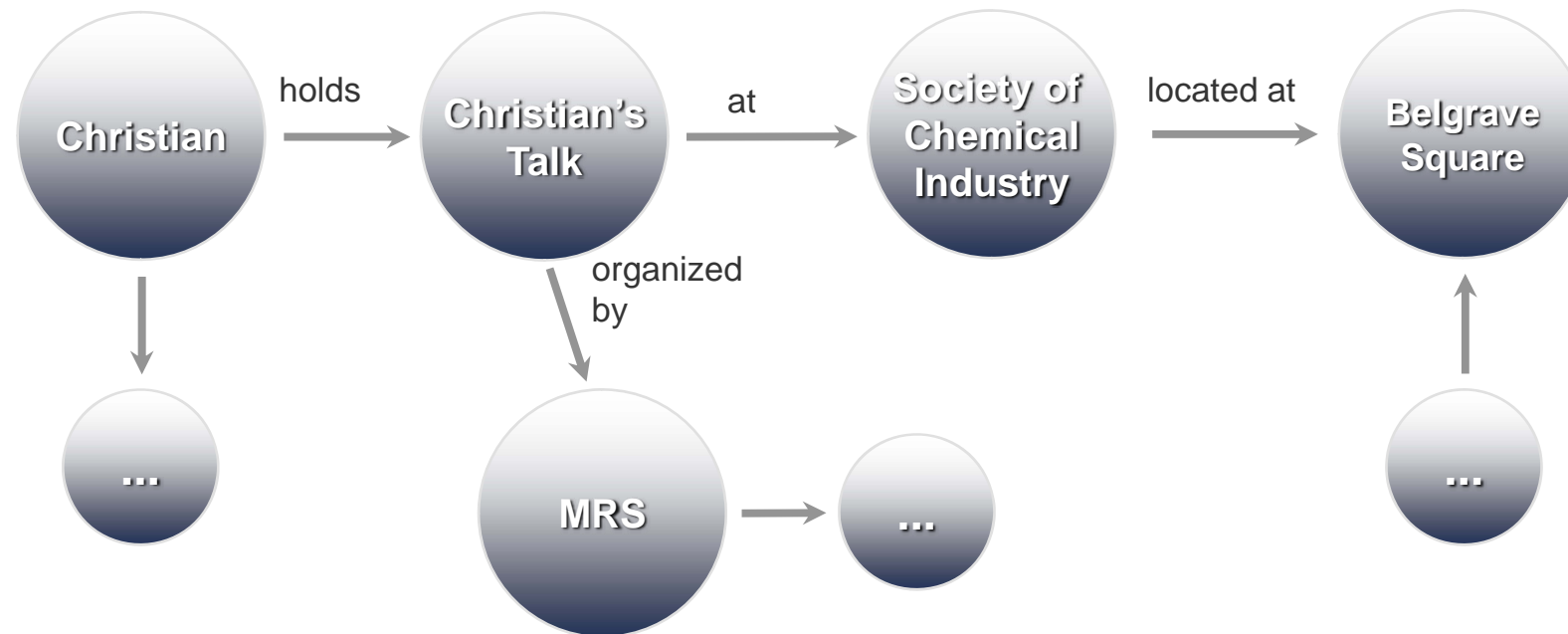
- Blog

Public Profile

<http://www.linkedin.com/in/christianhbecker>

We live in a world of data

- Now that we have identified resources, we can link them (to express information)



- *Linked Data* allows to realize this on a global scale
- Allows links to be cross-database, cross-organizational and cross-domain, much like links on websites
- A concept by Tim Berners-Lee, inventor of the Web

Linked Data as part of the Semantic Web

■ Definition

“The Semantic Web is an evolving extension of the World Wide Web *in which the semantics of information and services on the web is defined*, making it possible for the web to understand and satisfy the requests of people and machines to use the web content.” (*Wikipedia*)

■ The term “Semantic Web” encompasses general design principles and standards and has sprouted many different areas of research

■ Linked Data

- Small set of core principles
- Aims to get actual data published on the Web
- Semantic Web put to practice

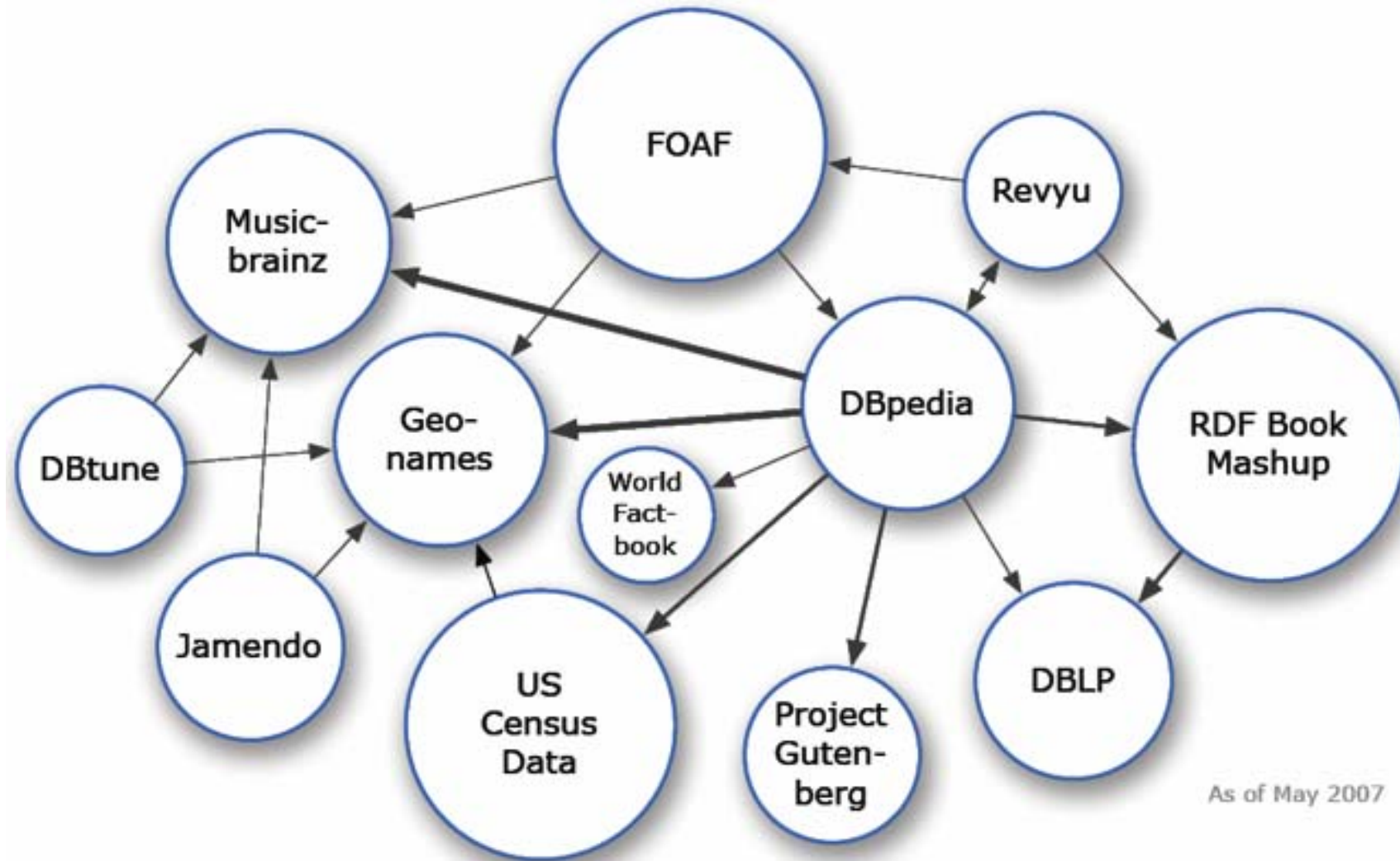
W3C Linking Open Data Project



■ 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

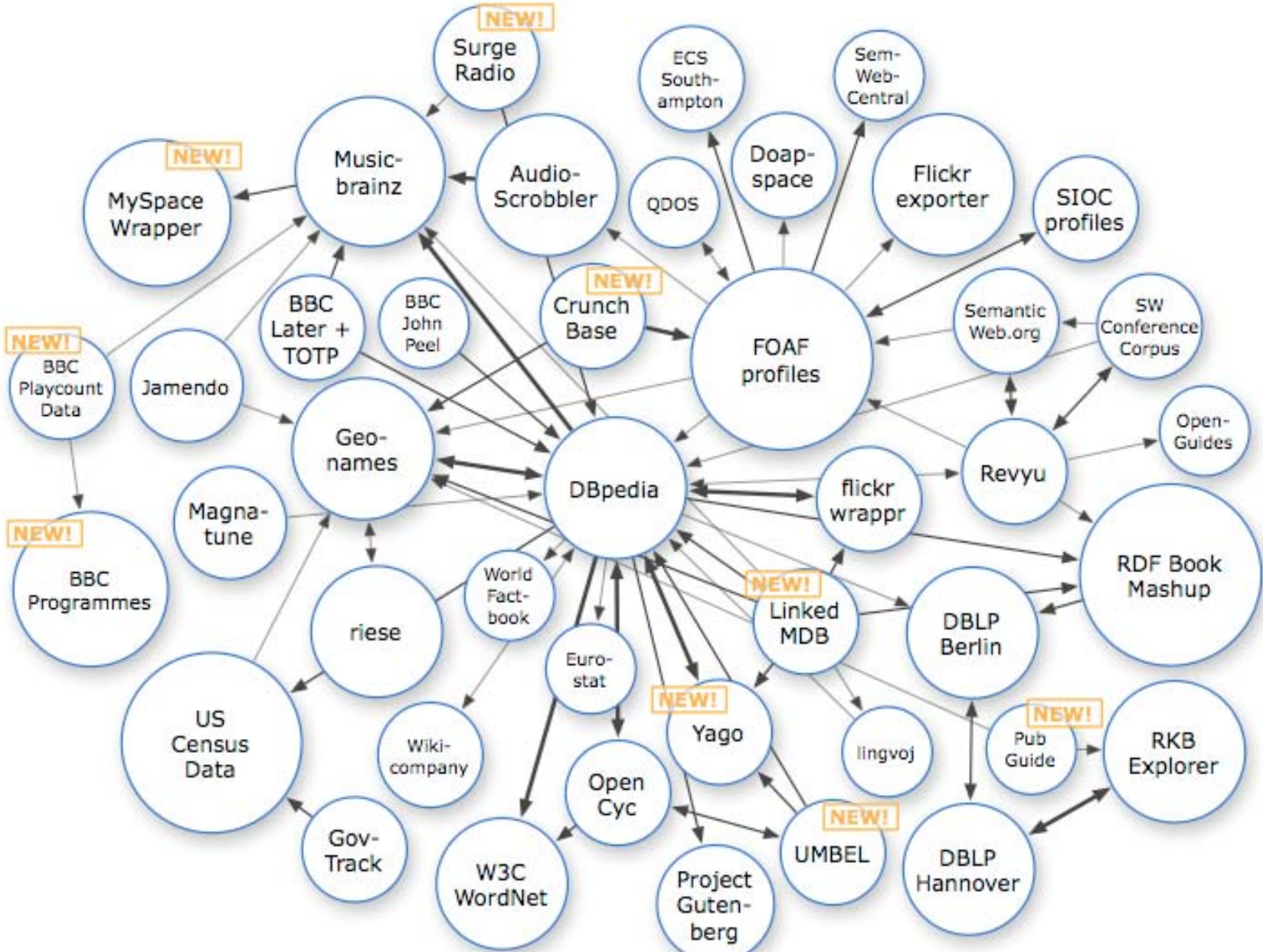


As of May 2007

500 million RDF triples

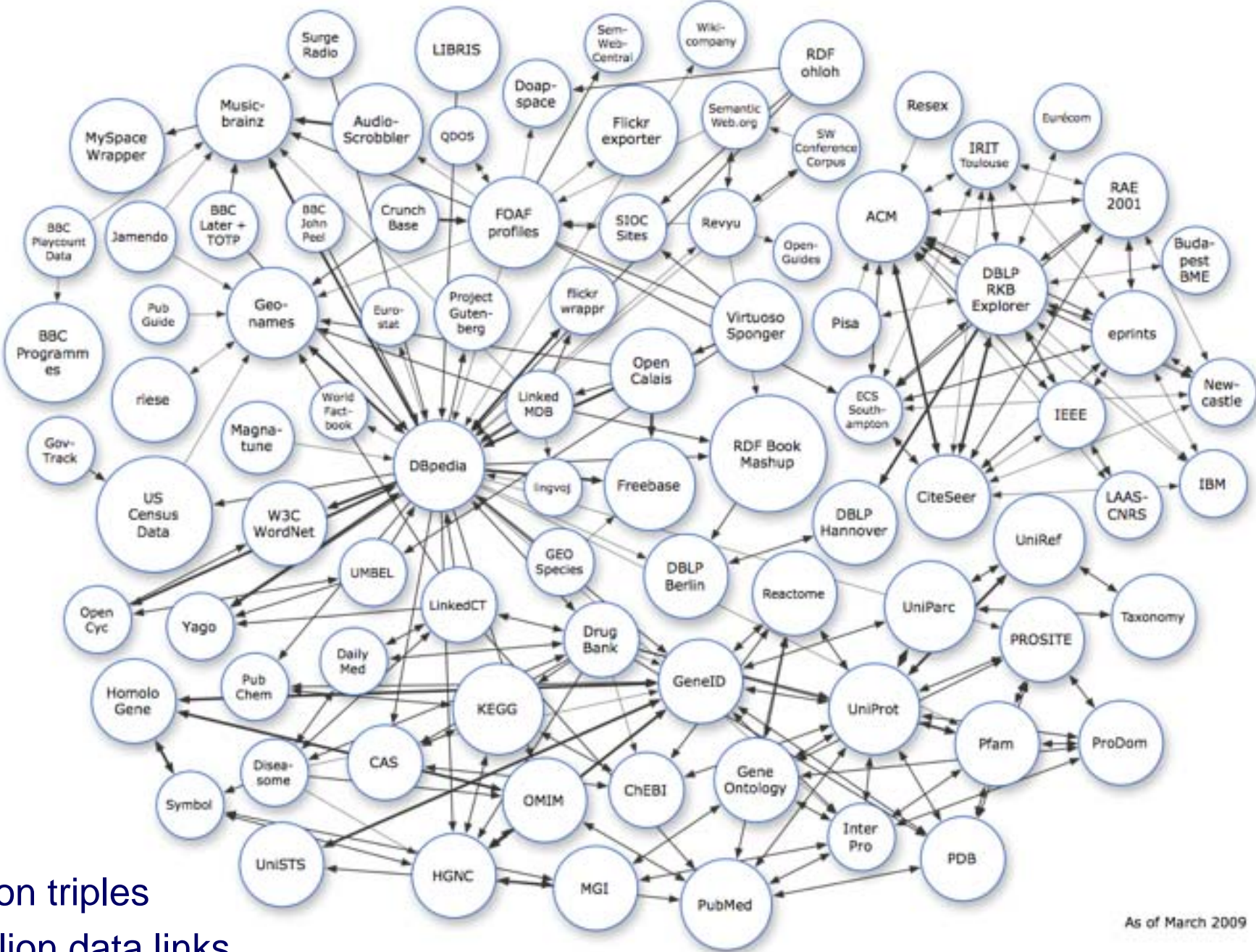
120,000 data links between data sources

LOD Datasets on the Web: September 2008



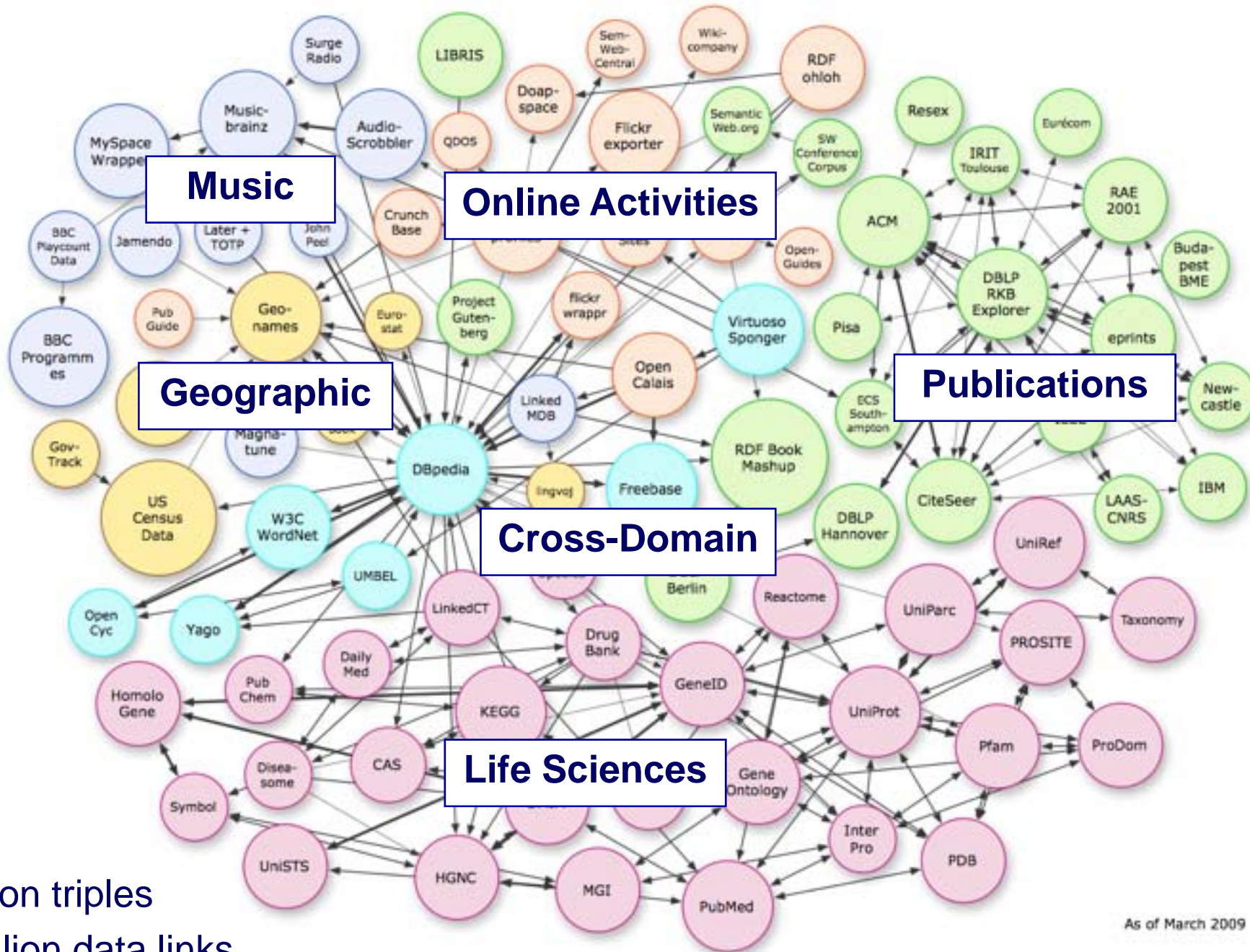
As of September 2008

LOD Datasets on the Web: March 2009



4.5 billion triples
180 million data links

LOD Datasets on the Web: March 2009



4.5 billion triples
180 million data links

DBpedia as a Geospatial Entrypoint to the Web of (Linked) Data

■ 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 open datasets on the Web

■ Contributors

- Freie Universität Berlin (Germany)
- Universität Leipzig (Germany)
- OpenLink Software (UK)



UNIVERSITÄT LEIPZIG



Extracting Data from Wikipedia

Calgary
From Wikipedia, the free encyclopedia

Coordinates: 51°02′42″N 114°03′26″W﻿ / ﻿51.044998°N 114.057220°W﻿ / 51.044998; -114.057220

This article is about the Canadian city. For other uses, see *Calgary (disambiguation)*.

Calgary (pronounced /kælɡrɪ/) is the largest city in the province of Alberta, Canada. It is located in the south of the province, in an area of foothills and high plains, approximately 80 km (50 mi) east of the front ranges of the Canadian Rockies. The city is located in the Parkland region of Alberta. Calgary is the third largest civic municipality, by population, in Canada. As of the 2006 civic census, Calgary's population was 1,042,892.^[1] The metropolitan population (CMA) was 1,162,310 in 2006,^[2] making Greater Calgary the fifth largest census metropolitan area in the country after Toronto, Montreal, Vancouver, and Ottawa. Because it is located 298 km (185 mi) due south of Edmonton, statisticians define the narrow populated area between these cities as the "Calgary-Edmonton Corridor."^[3] Calgary is the largest Canadian metropolitan area between Toronto and Vancouver.

Calgary is well-known as a destination for winter sports and recreation, with a number of major mountain resorts near the city and metropolitan area. Economic activity in Calgary is mostly centred on the petroleum industry; however, agriculture, tourism, and high-tech industries contribute to the city's fast economic growth. Calgary holds many annual festivals which include the Calgary Stampede, the Folk Festival, the Lilac Festival, One Yellow Rabbit High Performance Arts Festival, Calgary's International Festival of the Arts, Wordfest: Banff-Calgary International Writers Festival, Calgary International Spoken Word Festival, One World Festival (GlobalFest), and the fourth largest Caribbean festival in the country (Carifest). In 1988, Calgary became the first Canadian city to host the Olympic Winter Games, and one of the fastest long track speed skating ice rinks in the world was built at the University of Calgary to accommodate these Games.

Calgary was ranked the world's cleanest city by Mercer Quality of Living in a survey published in 2007 by *Forbes* magazine.^[4]

Contents [hide]

- History
 - 1.1 First settlement
 - 1.2 The oil boom
 - 1.3 Recent history
- Geography
 - 2.1 Calgary's neighbourhoods
 - 2.2 Climate
- Fauna and flora
- Culture

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

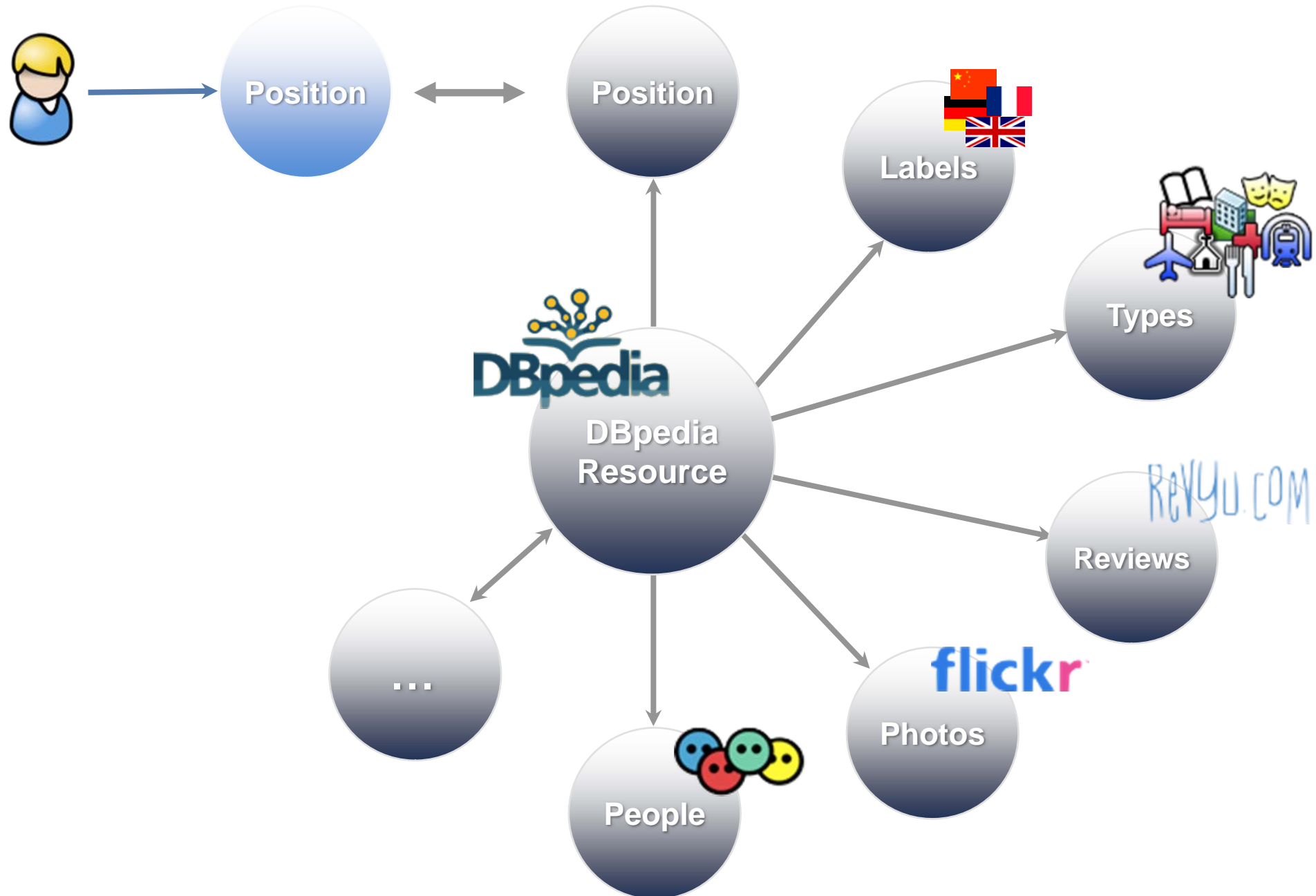
```
<http://dbpedia.org/resource/Calgary>  
dbpedia:native_name "Calgary" ;  
dbpedia:altitude "1048" ;  
geo:lat "51.044998" ;  
geo:long "-114.057220" ;  
dbpedia:population_city "988193" ;  
dbpedia:population_metro "1079310" ;  
mayor_name  
    dbpedia:Dave_Bronconnier ;  
governing_body  
    dbpedia:Calgary_City_Council ;  
...
```



The DBpedia Dataset

- **Web-scale identifiers for 2.6 million “things”**
 - including at least 213,000 persons
 - 328,000 places
- **Altogether 274 million pieces of information (RDF triples)**
 - 29 million triples extracted from infoboxes
 - 609,000 links to pictures
 - 3,150,000 links to relevant external web pages
 - 4,900,000 links to other LOD datasets
- **Categorizations**
 - DBpedia Ontology with 170 classes and 940 properties, based on the most frequently used classes
 - 415,000 Wikipedia categories
 - 75,000 YAGO categories

DBpedia as a geospatial entrypoint to the web of data



Potential uses for DBpedia in a mobile context

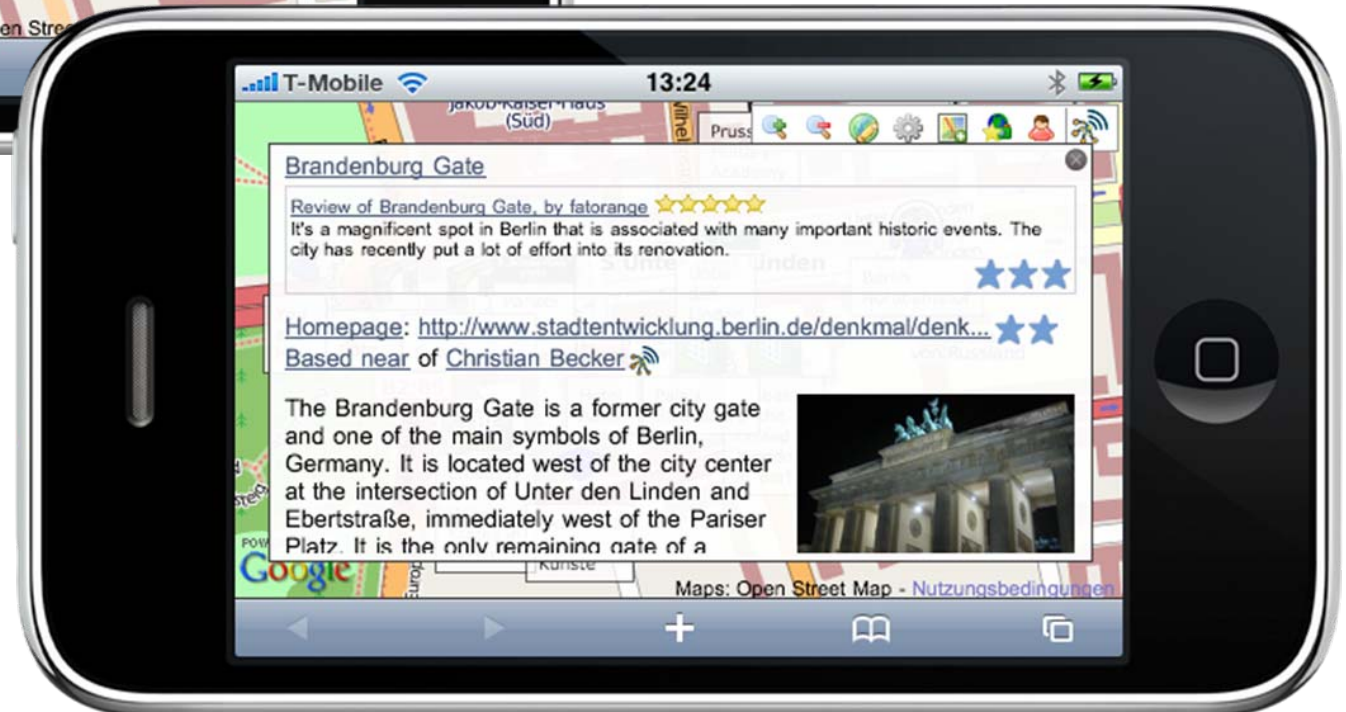
■ Active

- Explore the area (DBpedia Mobile)
- Local search
- Complex queries
 - e.g. show me the stops of a specific train line

■ Passive

- Detect the user's location - not simply as coordinates, but as a data resource that can be analyzed and act accordingly
 - Museum: Silence the phone
 - Airport: Show current flights
 - Pub, concert: Show nearby friends

DBpedia Mobile



Use DBpedia locations as starting points ...



Choose between 14 languages ...



Get more details ...



Get all details ... and navigate along data links



Implications of Linked Data for Geodemographics

Implications of Linked Data

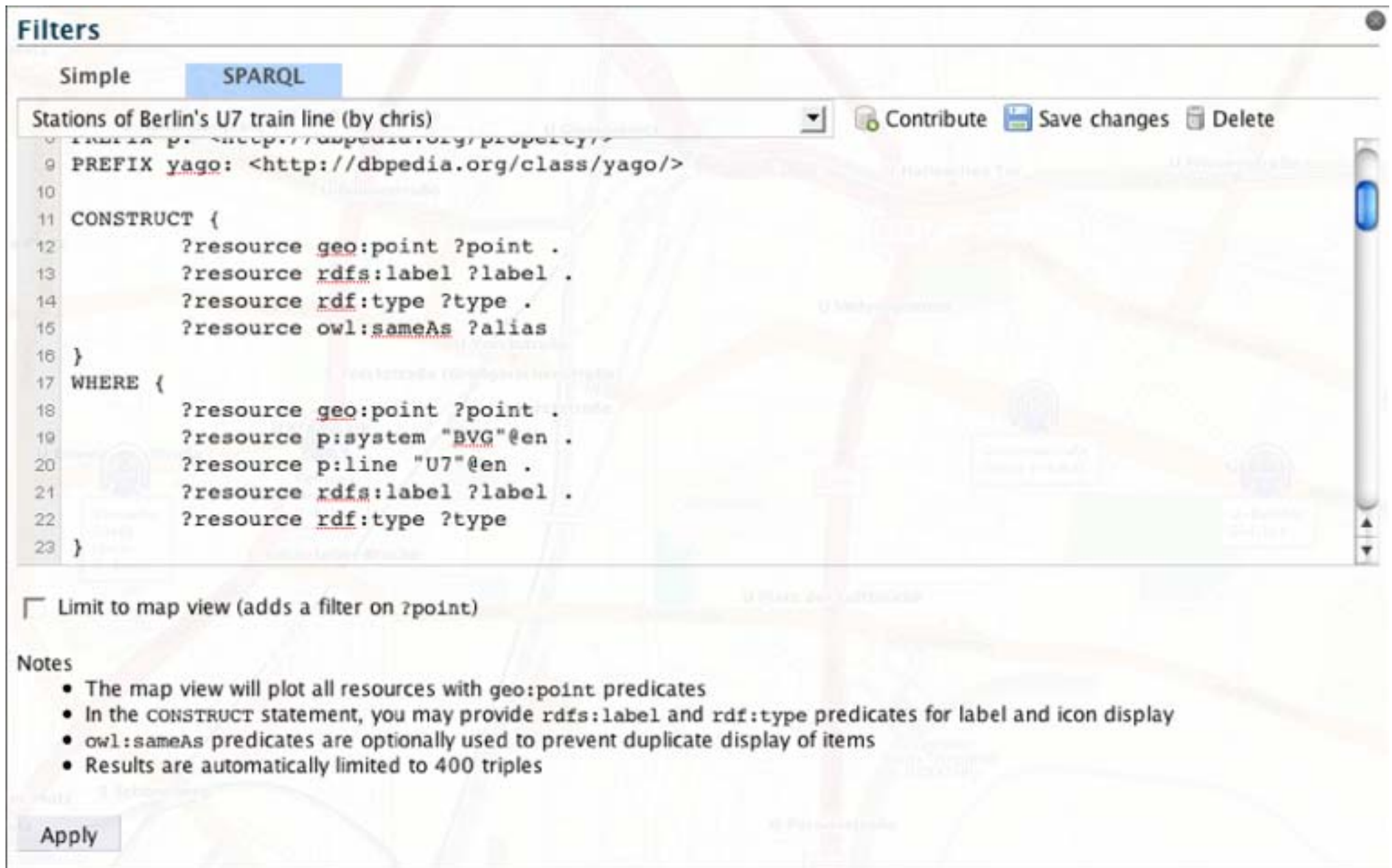
■ For professional data consumers

- Reduce data integration costs
- Reduce licensing costs
- Combine with organization-internal data
- Cross-domain use
 - “Find parking spots near UK airports in cities with more than 1 million inhabitants”
 - “Find hotels in regions with high disposable income”
 - “Find competitor locations and reviews”

■ For end users

- Richer, more up to date navigation
- Better detection of context
- Better context-related offerings

Complex queries: Find stops of a specific train line



The screenshot shows a web interface for running SPARQL queries. The title is "Filters" and the query is titled "Stations of Berlin's U7 train line (by chris)". The query is written in SPARQL and is used to find the stops of the U7 train line in Berlin. The query includes a PREFIX for 'yago' and a CONSTRUCT statement to generate labels and types for the results. The WHERE clause filters for resources that are geo:point, belong to the BVG system, are on the U7 line, and have an rdfs:label. The interface also includes a checkbox for "Limit to map view (adds a filter on ?point)", a "Notes" section with four bullet points, and an "Apply" button.

```
9 PREFIX yago: <http://dbpedia.org/class/yago/>
11 CONSTRUCT {
12     ?resource geo:point ?point .
13     ?resource rdfs:label ?label .
14     ?resource rdf:type ?type .
15     ?resource owl:sameAs ?alias
16 }
17 WHERE {
18     ?resource geo:point ?point .
19     ?resource p:system "BVG"@en .
20     ?resource p:line "U7"@en .
21     ?resource rdfs:label ?label .
22     ?resource rdf:type ?type
23 }
```

Limit to map view (adds a filter on ?point)

Notes

- The map view will plot all resources with geo:point predicates
- In the CONSTRUCT statement, you may provide rdfs:label and rdf:type predicates for label and icon display
- owl:sameAs predicates are optionally used to prevent duplicate display of items
- Results are automatically limited to 400 triples

Apply

Complex queries: Find stops of a specific train line



OpenStreetMap: England 2006-2008



Credits: GeoFabrik

Christian Becker: Linking Spatial Data from the Web (London, 04/01/2009)

Open Data is growing!

- **Vast, steadily growing amounts of publicly accessible data**
- **Rich user-generated geo data for Europe due to its vivid editor community, notably in the UK and Germany**
- **Currently available geographic datasets as part of Linking Open Data:**
 - **Wikipedia (DBpedia, rich data about 328,000 places)**
 - **OpenStreetMap (1.2 million points of interest)**
 - post boxes
 - traffic lights
 - ...
 - **GeoNames (administrative hierarchies)**
 - **US Census**
 - **EuroStat (General and Economics datasets)**
 - **World Factbook**

Current Status of Linked Data

- **W3C-approved standards for cross-database, cross-organizational, cross-domain interoperability**
- **Datasets and links are not perfect, but graduating beyond a research stage**
- **Adoption by data providers and support by major players**
 - **BBC**
 - **Thomson Reuters**
 - **Cyc Foundation**
 - **Life Sciences community**
 - **W3C**
- **Large amounts of data that are yet untapped**
 - **Open government data**
 - **User-generated data such as reviews and locations**

Thanks

■ Questions?

■ References

- Tim Berners-Lee's TED Talk

http://www.ted.com/index.php/talks/tim_berniers_lee_on_the_next_web.html

- LinkedData.org

- OpenStreetMap.org

Linked Data version: LinkedGeoData.org

- DBpedia.org

- DBpedia Mobile:

<http://beckr.org/DBpediaMobile>

