

Svetovni splet podatkov

Iztok Savnik

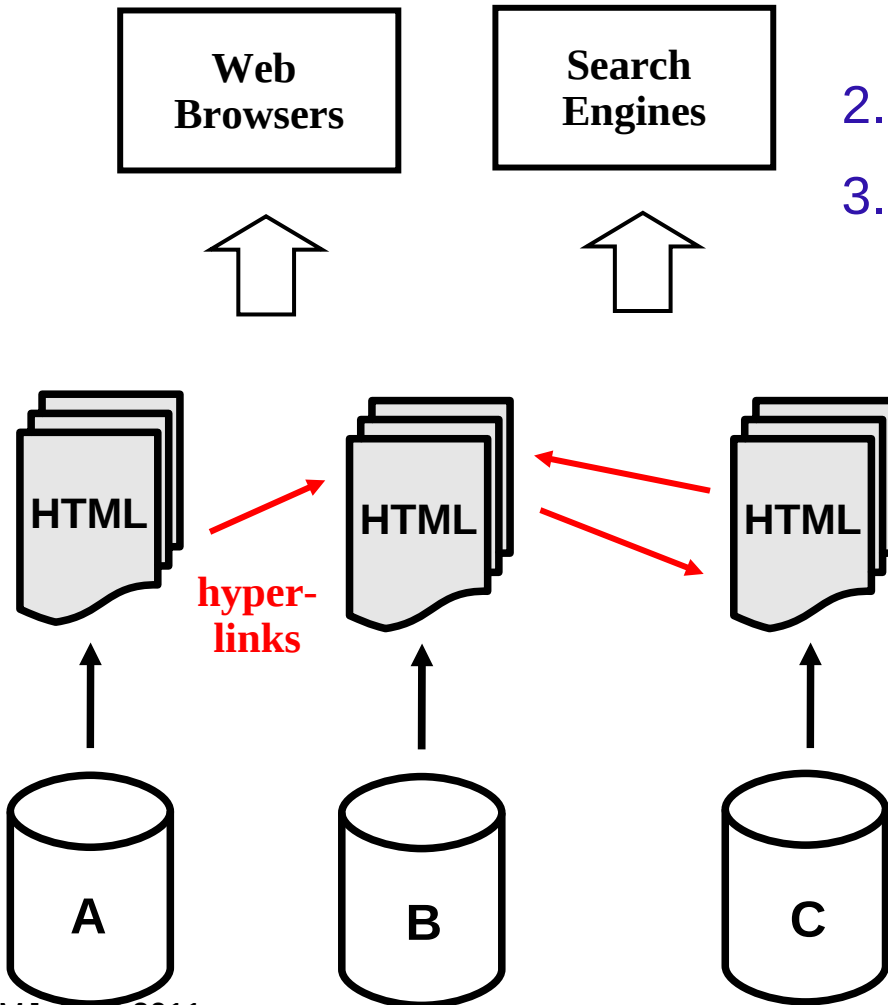
Linked data

1. Od spleta dokumentov do spleta podatkov
 - Spletni API-ji in Linked Data
2. Linked Data implementacija = RDF+OWL na spletu
 - Kateri podatki obstajajo?
3. Aplikacije
 - Kaj se dogaja s podatki?
4. Naslednji koraki
 - Kaj manjka?

Klasični splet

En sam globalen inform. prostor

1. URL za:
 - Globalni unikatni IDs
 - Poizvedovalni mehanizmi
2. HTML kot skupna oblika vsebine
3. Hyper-povezave



Problem in rešitev

Problem

Ker je vsebina spleta zelo šibko strukturirana, aplikacije težko implementirajo pametne operacije.

Rešitev

Povečaj strukturo vsebine spleta.

Spletni API-ji in prepletene storitve



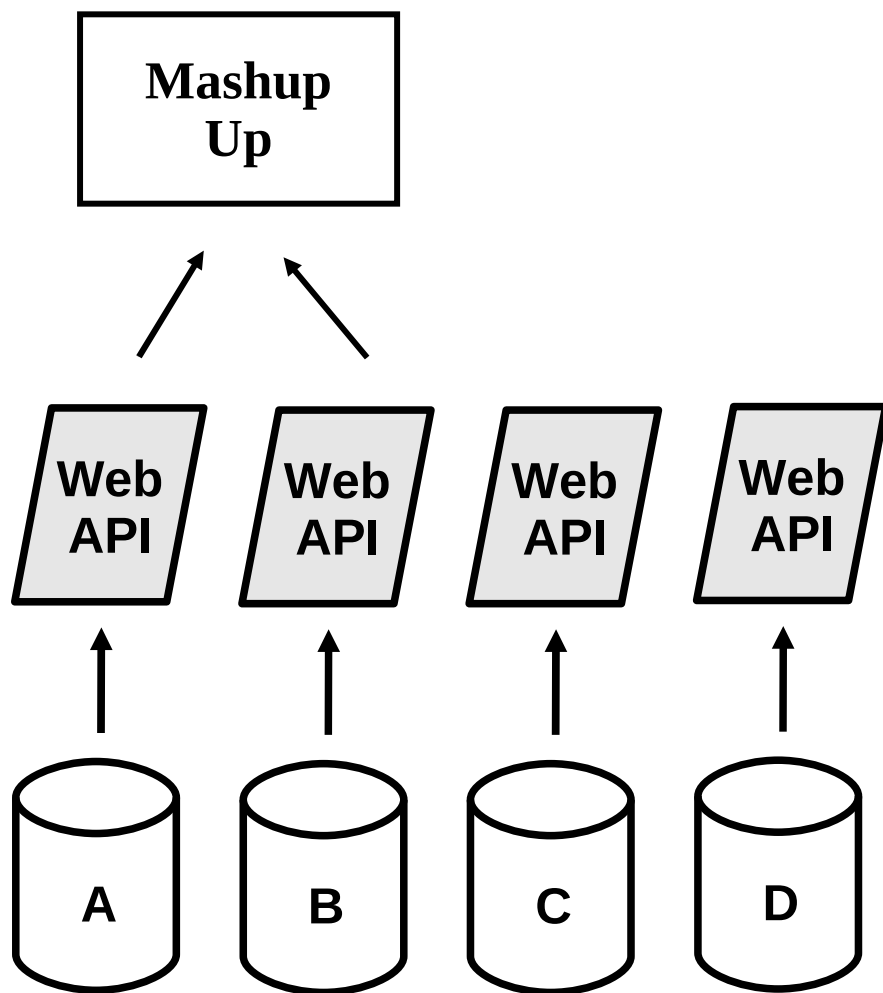
Weather.com interface showing a map of the United States with weather icons and a sidebar with weather details.

SecretPrices.com interface showing a search bar, navigation menu, and a list of latest secret product deals.

Wiiseeker interface showing a map of Lancaster, PA, with a search for 'circuit city' and a list of nearby electronics stores.

Flickr Sudoku puzzle game interface, showing a 9x9 grid with numbers and puzzle pieces.

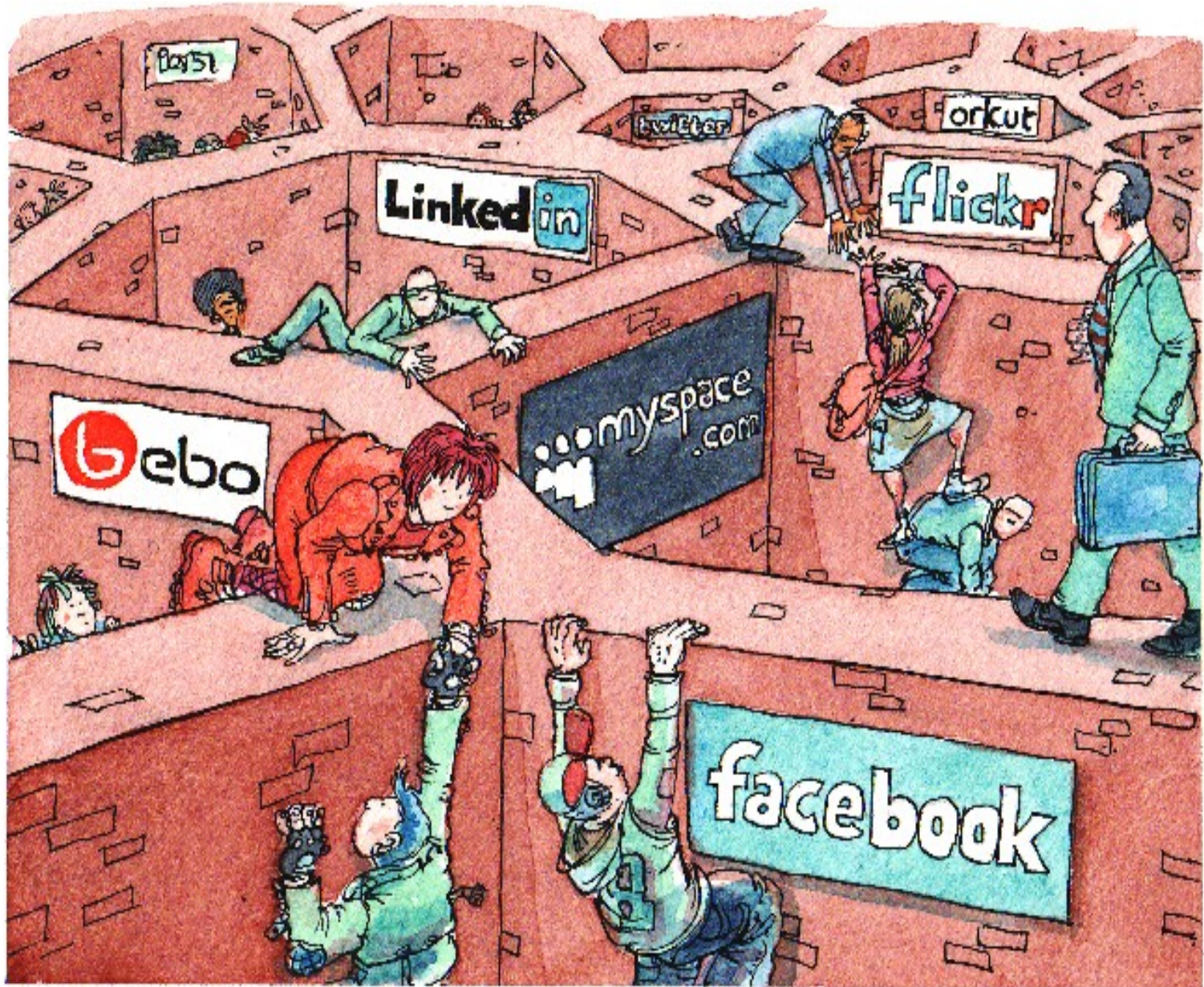
Spletni API-ji in mashup-i



Slabe lastnosti:

- 0) API-ji nudijo privatne vmesnike.
- 1) Mashup-i temeljijo na fiksni množici podatkovnih virov.
- 2) Ne moremo definirati povezav med podatkovnimi objekti.

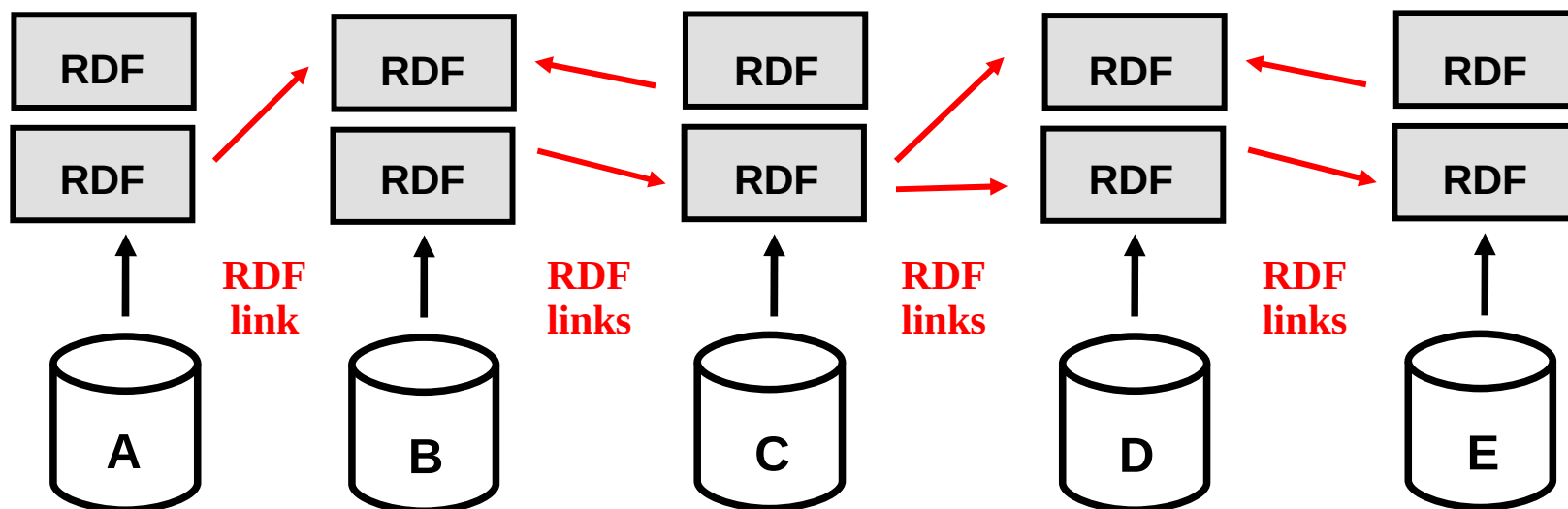
Spletni API-ji razdelijo splet na vrtove z ograjami



Povezani podatki

Uporabi tehnologije semantičnega spleta za:

- Publiciranje strukturiranih podatkov na splet,
- Definicijo povezav med podatki iz enega spletnega mesta s podatki v drugih spletnih mestih.



Principi povezanih podatkov

1. Uporabi URI-je kot imena za objekte
2. Uporabi HTTP URI, da lahko ljudje dostopajo do teh imen
3. Ko nekdo dostopa do URI zagotovi uporabne RDF podatke
4. Vključi RDF stavke drugih URI, da lahko ljudje odkrijejo sorodne objekte.

Tim Berners-Lee 2007

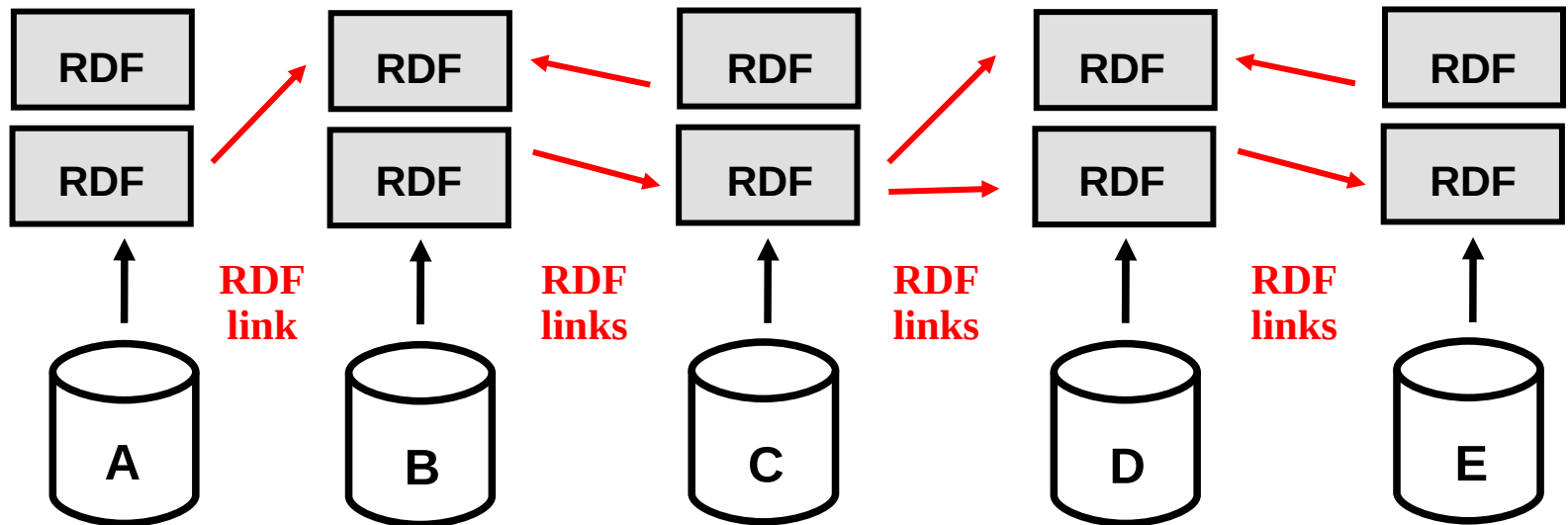
<http://www.w3.org/DesignIssues/LinkedData.html>

Lastnosti spleta povezanih podatkov

- Kdorkoli lahko publicira podatke na spletu povezanih podatkov
- Entitete so povezane s povezavami
 - Kreacija globalnega podatkovnega grafa, ki povezuje podatkovne vire in omogoča odkrivanje novih virov.
- Podatki so samo-opisni
 - Če aplikacija dobi podatke, ki so predstavljeni z nepoznanim besednjakom, mora aplikacija identificirati URI-je, ki identificirajo slovarje z definicijami RDFS in OWL izrazov.
- Splet podatkov je odprt
 - To pomeni, da lahko aplikacije odkrivajo nove podatkovne vire v času izvajanja.

Implementacija povezanih podatkov na spletu

- Je to realno?



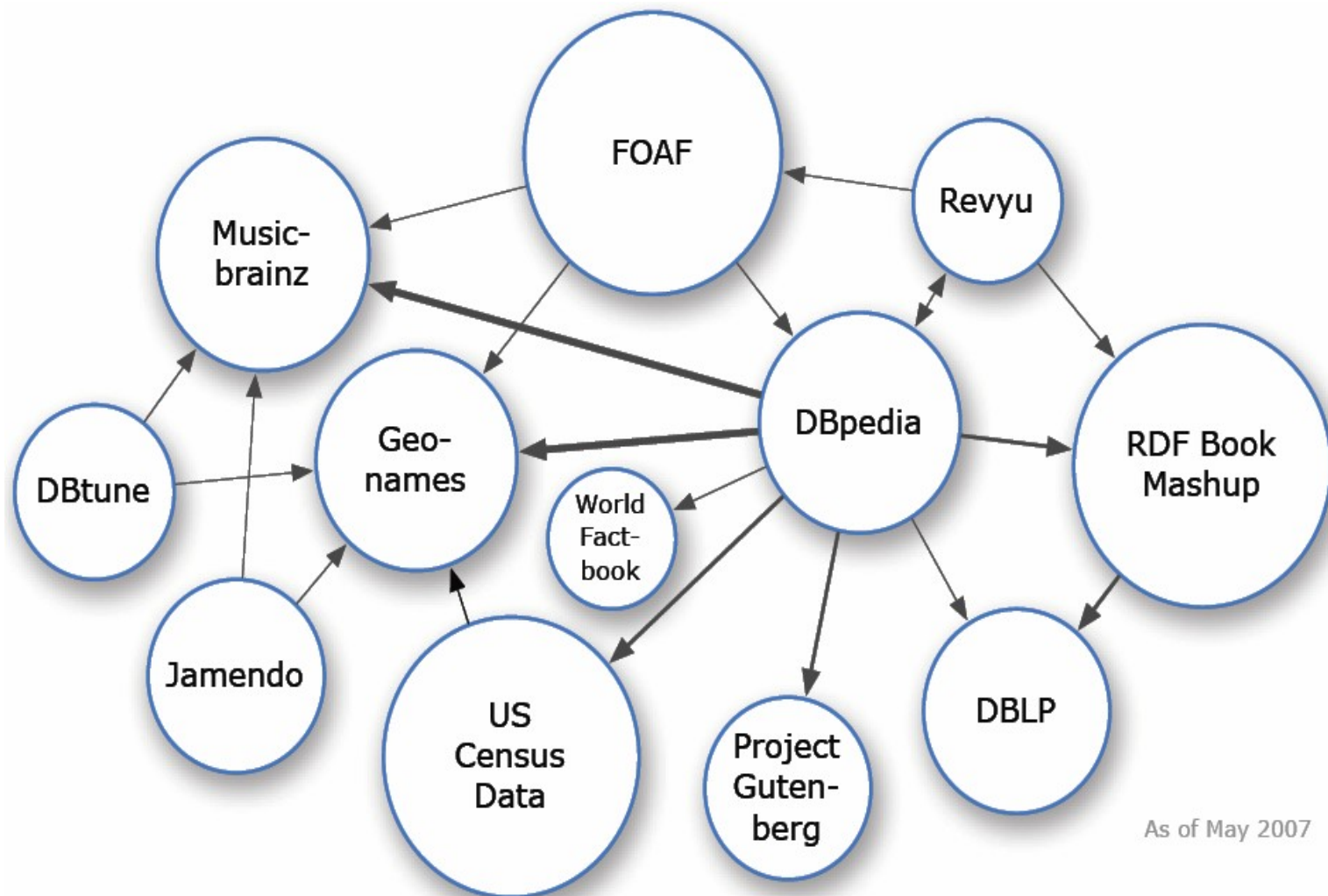
W3C Linking Open Data



■ Skupnost:

- Publiciranje obstoječih odprtih podatkovnih zbirk na spletu.
- Povezovanje stvari med različnimi podatkovnimi viri

LOD zbirke na spletu: Maj 2007



■ Nad 500 milijonov RDF trojic

■ Okoli 120,000 RDF povezav med pod.viri

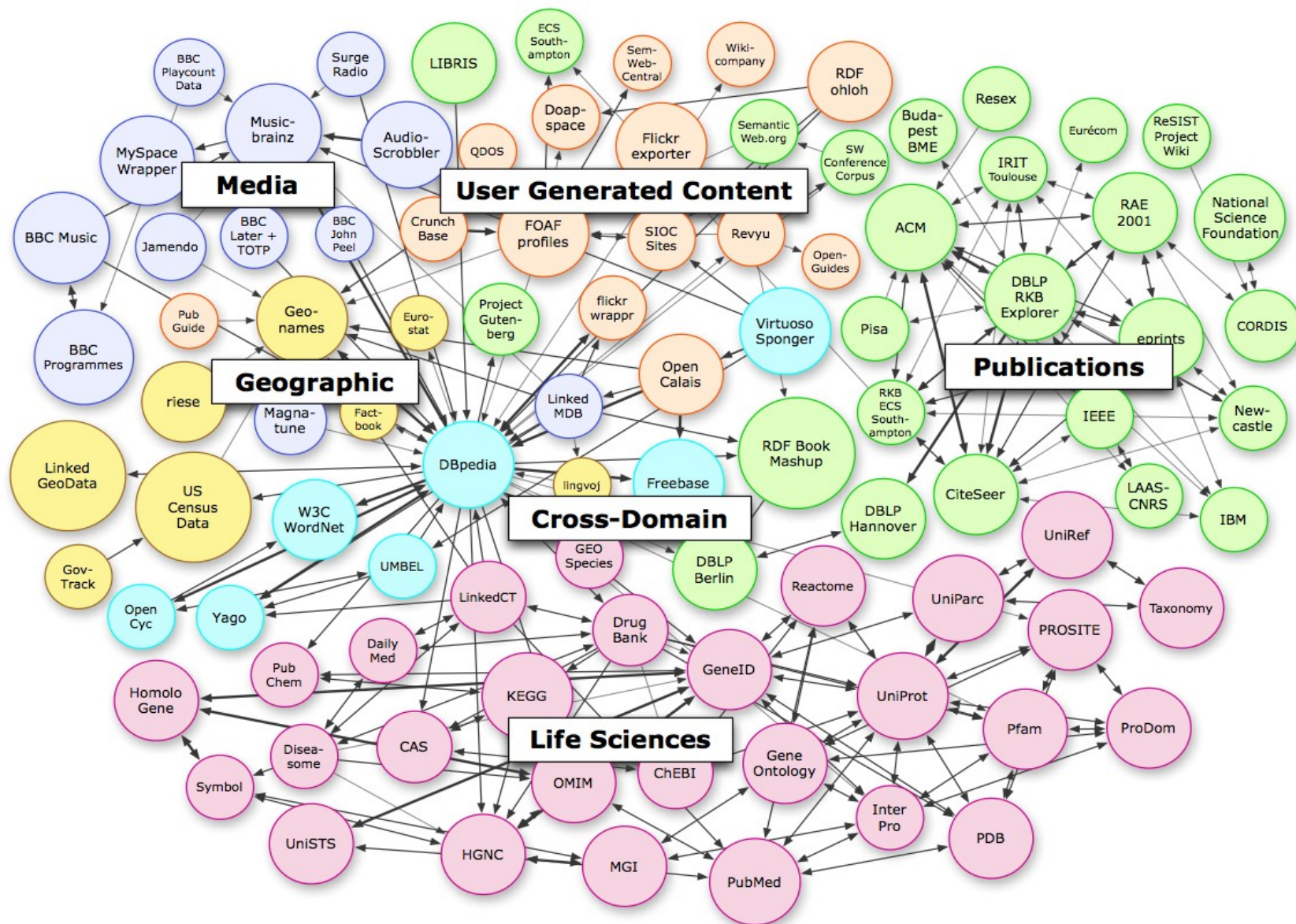
Primer RDF povezave

- RDF povezave iz DBpedia na druge podatkovne vire

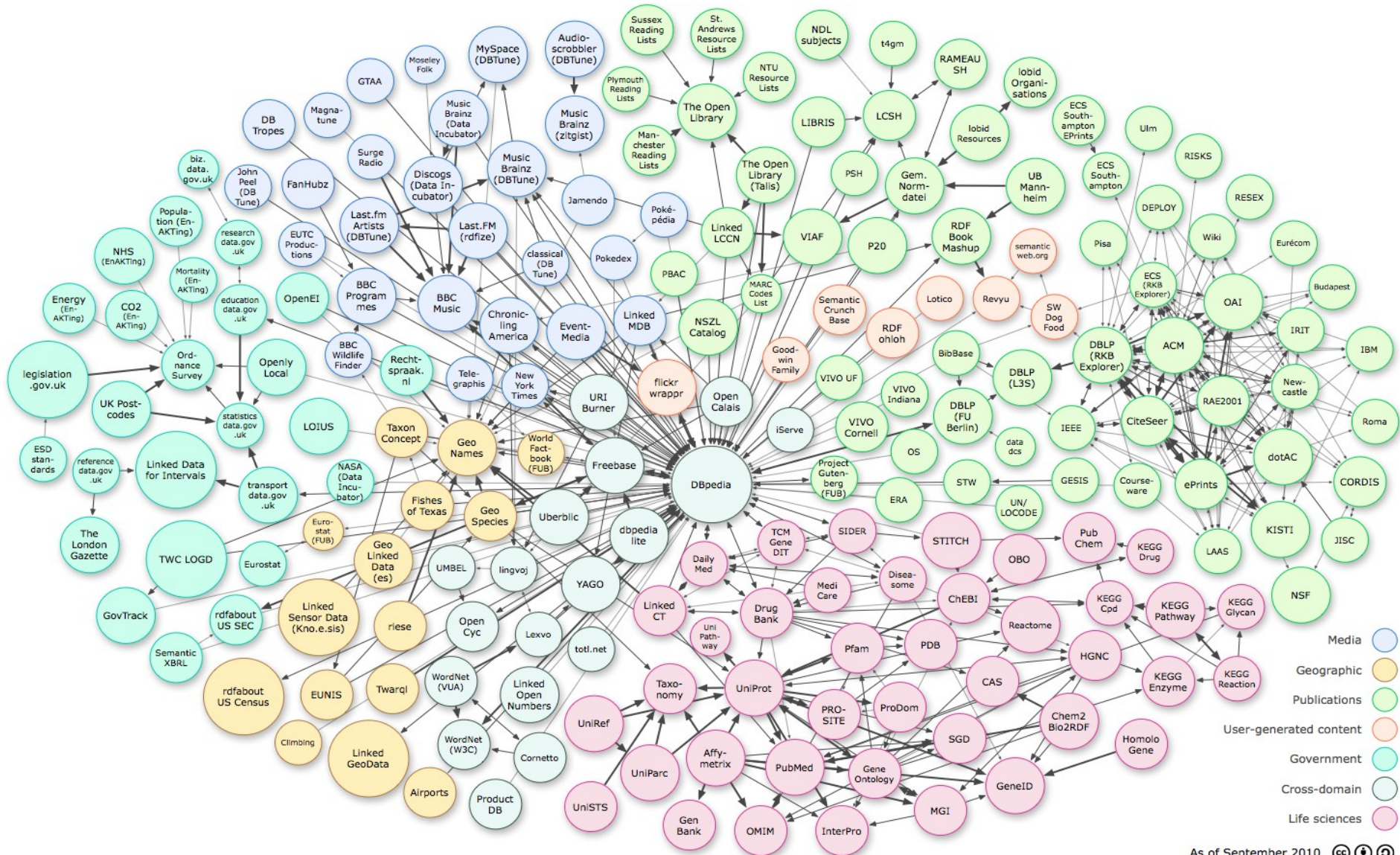
```
<http://dbpedia.org/resource/Berlin> owl:sameAs  
<http://sws.geonames.org/2950159> .
```

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


LOD zbirke na spletu: Julij 2009

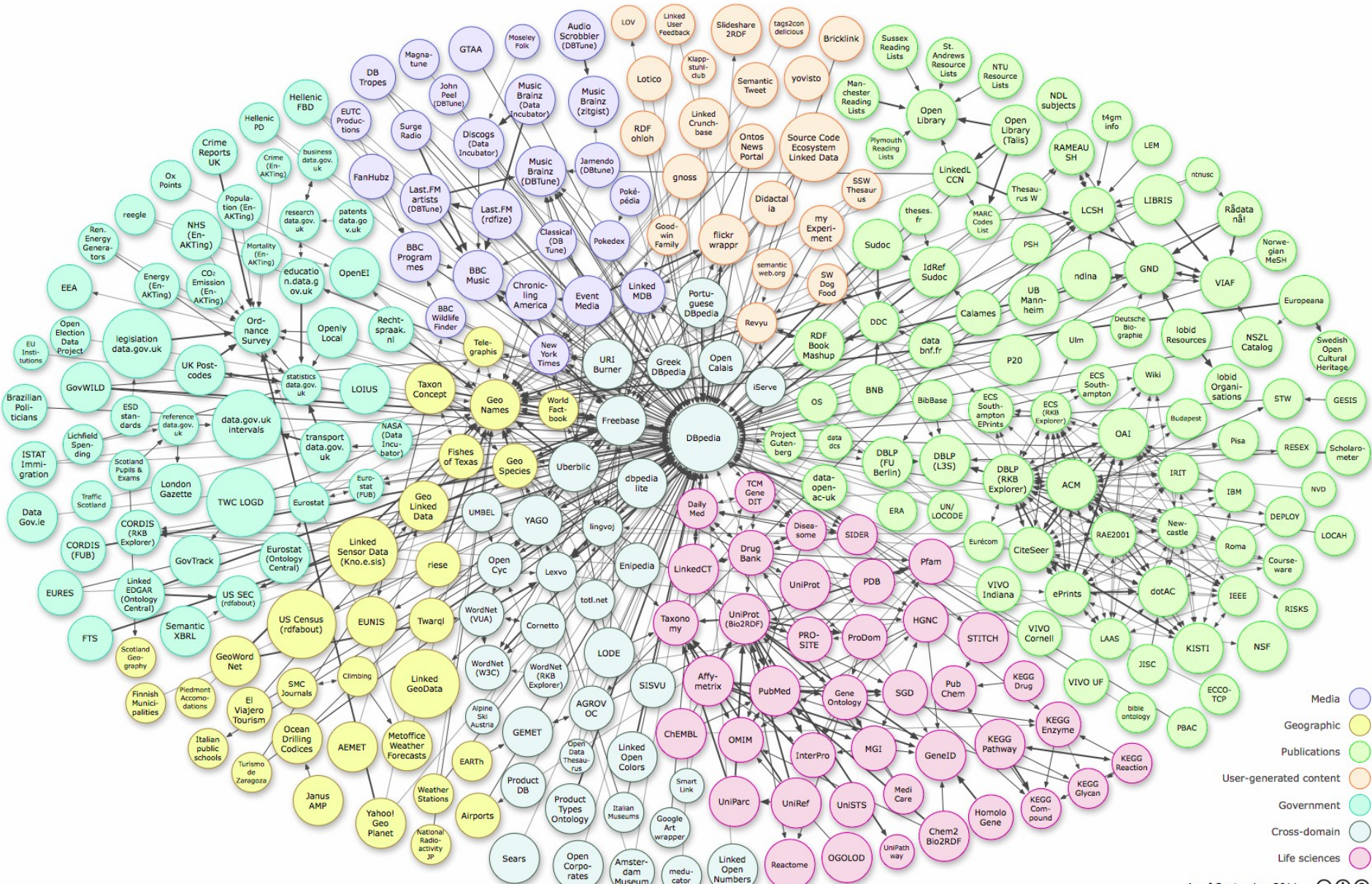


LOD zbirke na spletu: Sept 2010



As of September 2010

LOD zbirke na spletu: Sept 2011

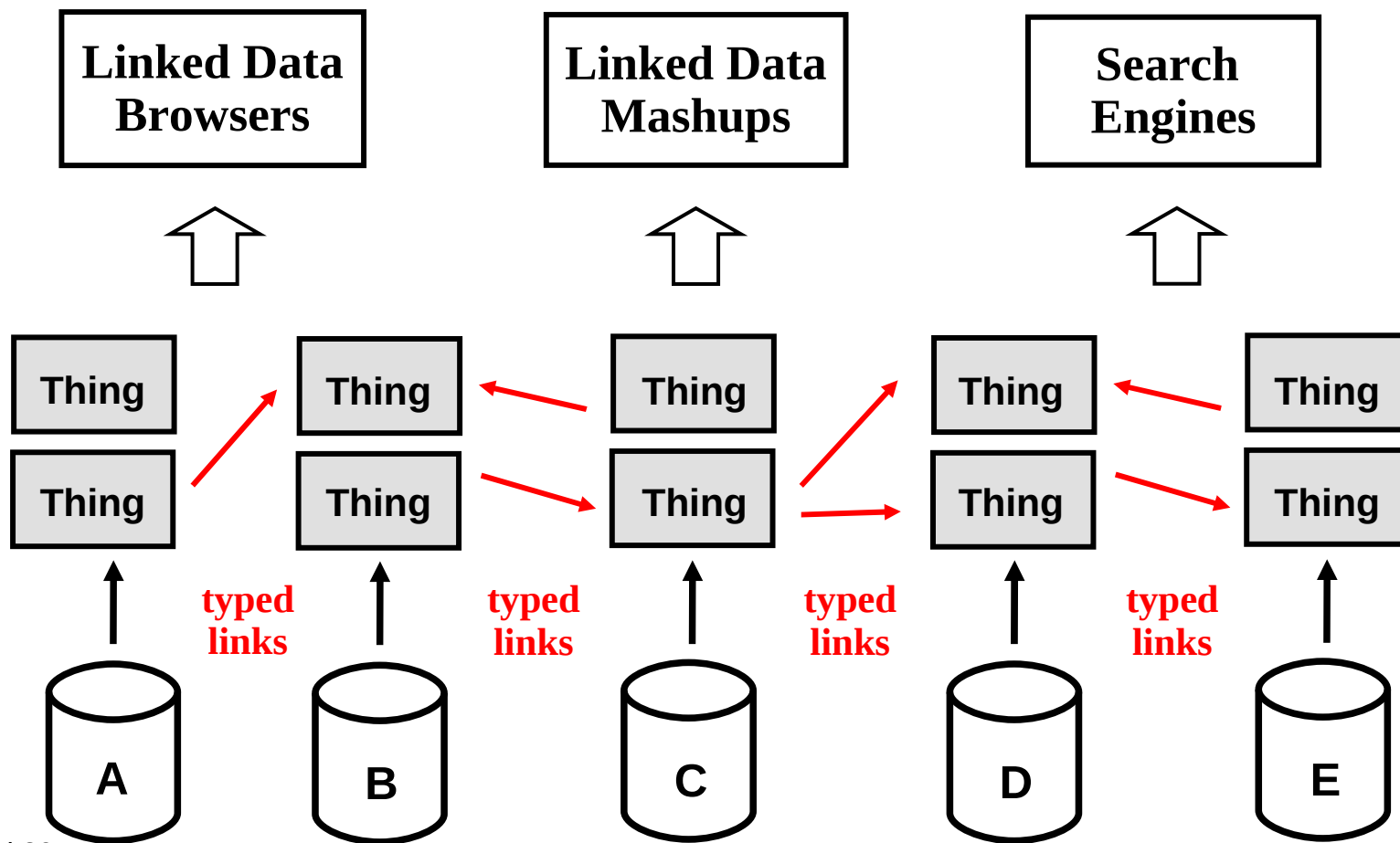


LOD statistika: Sept 2011

Domain	Number of datasets	Triples
Media	25	1,841,852,061
Geographic	31	6,145,532,484
Government	49	13,315,009,400
Publications	87	2,950,720,693
Cross-domain	41	4,184,635,715
Life sciences	41	3,036,336,004
User-generated content	20	134,127,413
	295	31,634,213,770

Aplikacije

- Kaj lahko naredimo s tem?



Brskalnik za povezane podatke

- **Tabulator Browser (MIT, USA)**
- **Marbles (FU Berlin, DE)**
- **OpenLink RDF Browser (OpenLink, UK)**
- **Zitgist RDF Browser (Zitgist, USA)**
- **Humboldt (HP Labs, UK)**
- **Disco Hyperdata Browser (FU Berlin, DE)**
- **Fenfire (DERI, Irland)**

Povezani podatki v prepletenih storitvah (meshups)

Aplikacije s specifično domeno, ki uporabljajo povezane podatke iz spleta

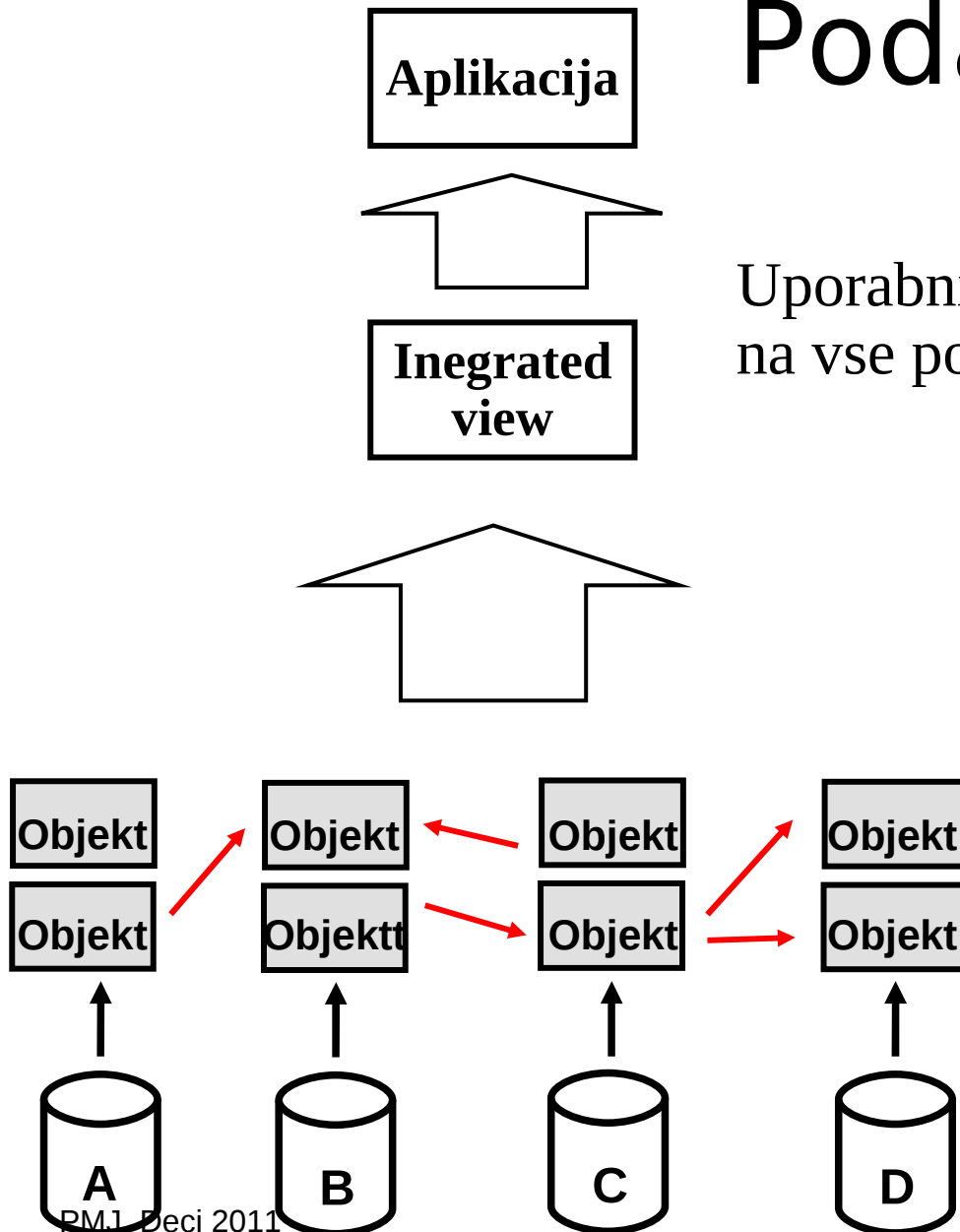
Podatkovna fuzija

Uporabnik želi imeti integriran pogled na vse podatke, ki so dostopni za objekt.

Znani problemi:

Preslikave shem

Reševanje nekonsistentnosti

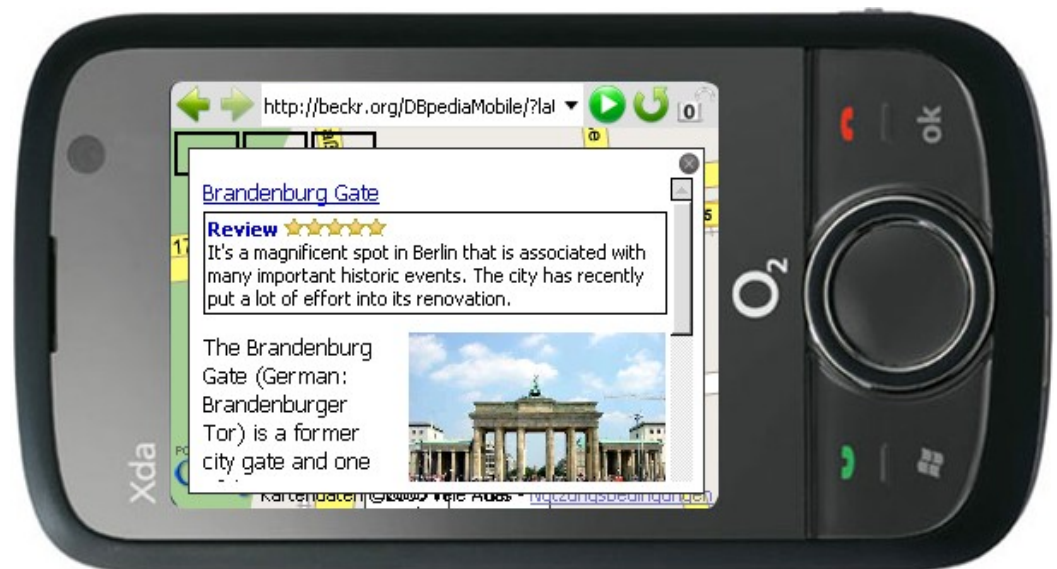


DBpedia Mobile



Geospatial
vhodna točka
na splet
podatkov

Uporablja:
DBpedia,
Revyu in Flickr
podatke



Podatkovni splet in iskalniki

Falcons (IWS, China)

<http://iws.seu.edu.cn/services/falcons/documentsearch/>

Sig.ma (DERI, Ireland)

<http://www.deri.ie/>, <http://sig.ma/>

Swoogle (UMBC, USA)

<http://swoogle.umbc.edu/>

Watson (Open University, UK)

<http://watson.kmi.open.ac.uk/WatsonWUI/>