ontotext

GATE

Module 14

Linked Data Management

Module 14 schedule

9.45-11.00	• Linked Data principles • Vocabularies & datasets
11.00-11.15	Coffee break
11.15-12.30	• Open Government Data • Tools • Open issues & challenges
12.30-14.00	Lunch break
14.00-16.00	• Introduction to *FactForge* and *LinkedLifeData* • The “Modigliani test” for the Semantic Web
16.00-16.30	Coffee

ontotext

GATE

Module 14 outline

- Linked Data principles
- Vocabularies & datasets
- Open Government Data
- Tools
- Open issues & challenges
- Introduction to *FactForge* and *LinkedLifeData*
- The “Modigliani test” for the Semantic Web

ontotext

GATE

Linked Data principles

Linked Data

"To make the Semantic Web a reality, it is necessary to have a large volume of data available on the Web in a standard, reachable and manageable format. In addition the relationships among data also need to be made available. This collection of interrelated data on the Web can also be referred to as ***Linked Data****. Linked Data lies at the heart of the Semantic Web: large scale integration of, and reasoning on, data on the Web."* (W3C)

- *Linked Data* is a set of principles that allows publishing, querying and browsing of RDF data, distributed across different servers
  - similar to the way HTML is currently published & consumed

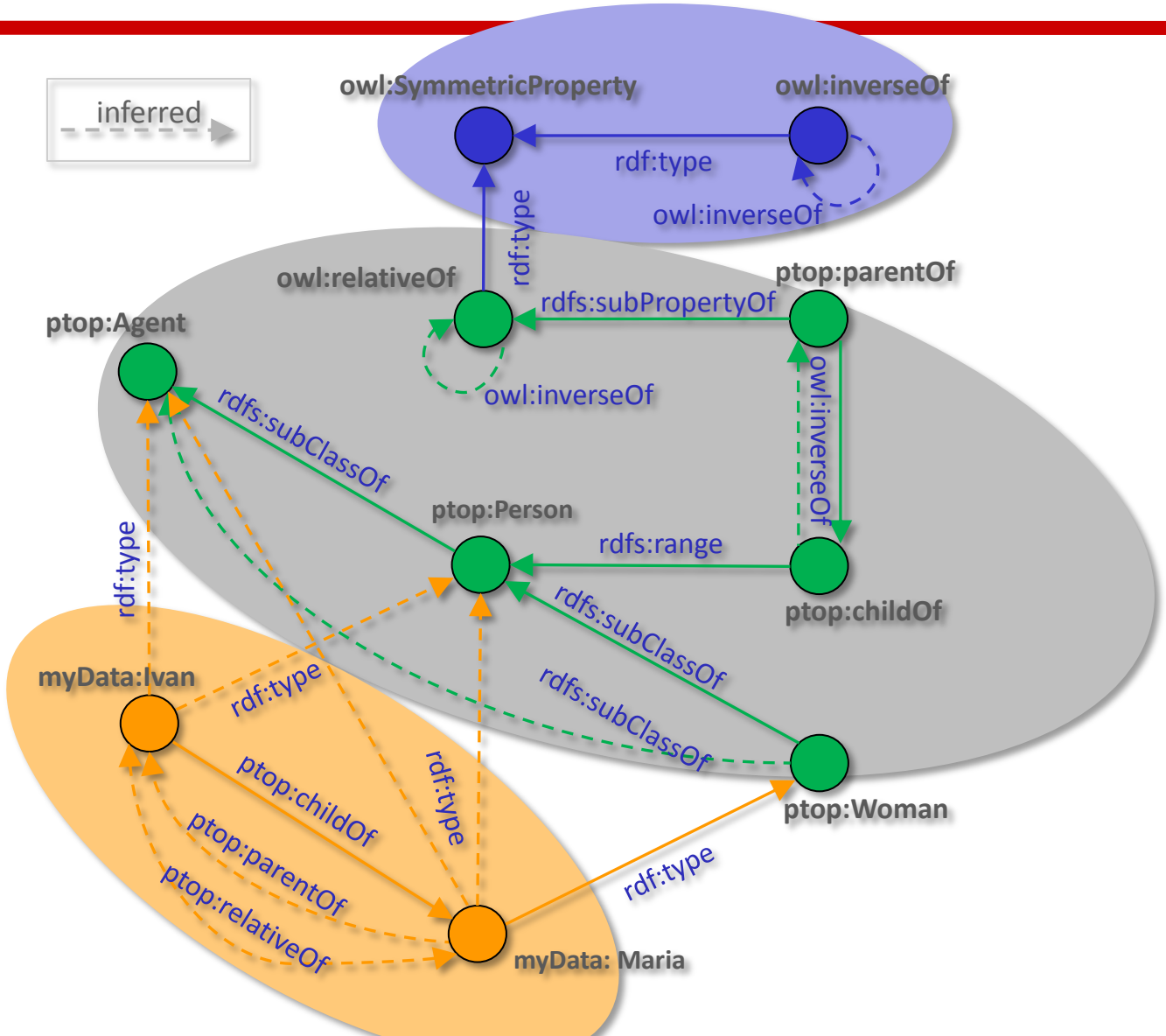
ontotext

GATE

Linked Data design principles

1. Unambiguous identifiers for objects (resources)
   - Use URIs as names for things
2. Use the structure of the web
   - Use HTTP URIs so that people can look up the names
3. Make is easy to discover information about an object (resource)
   - When someone lookups a URI, provide useful information
4. Link the object (resource) to related objects
   - Include links to other URIs

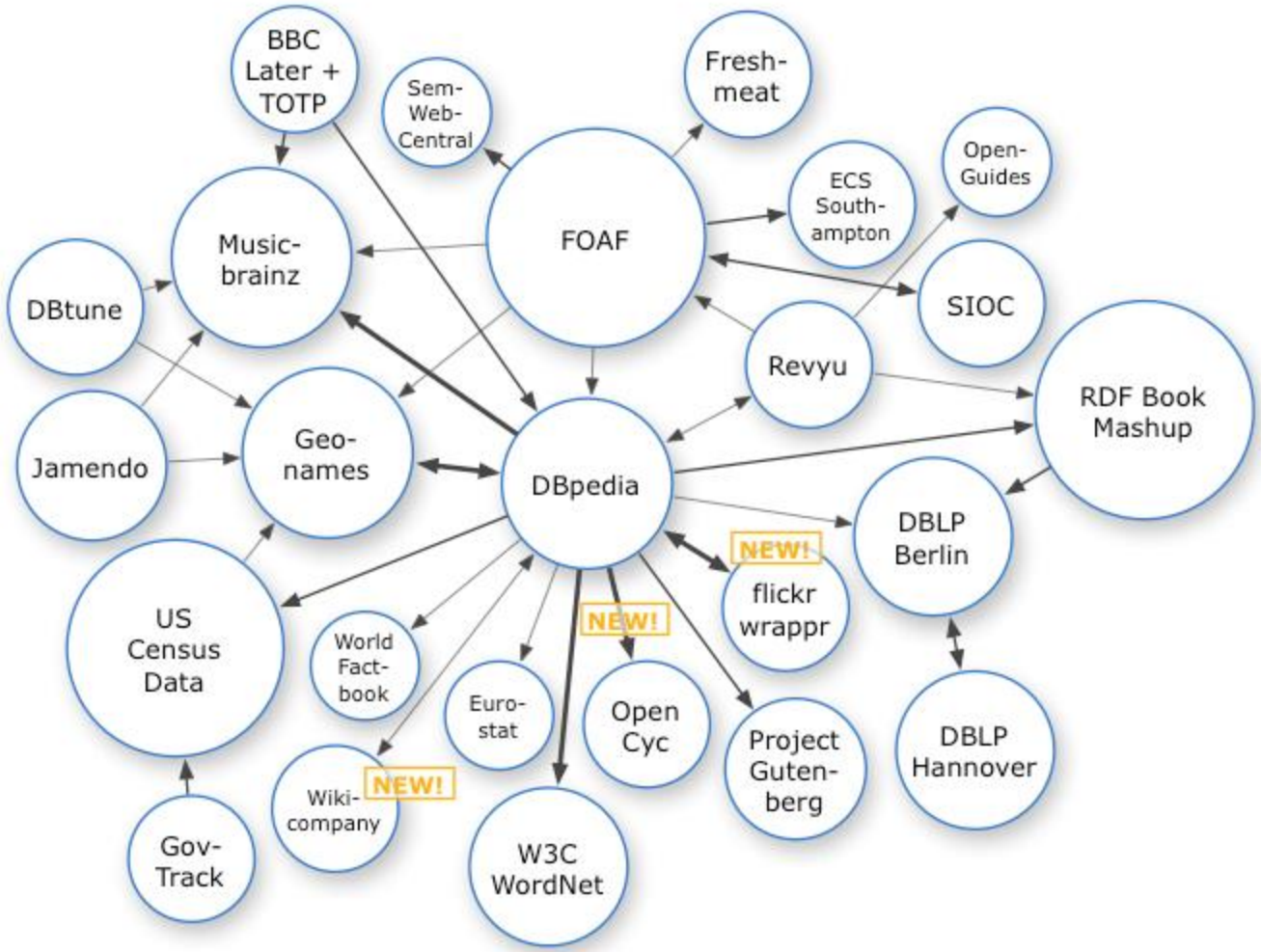
Linked Data (2)



ontotext

GATE

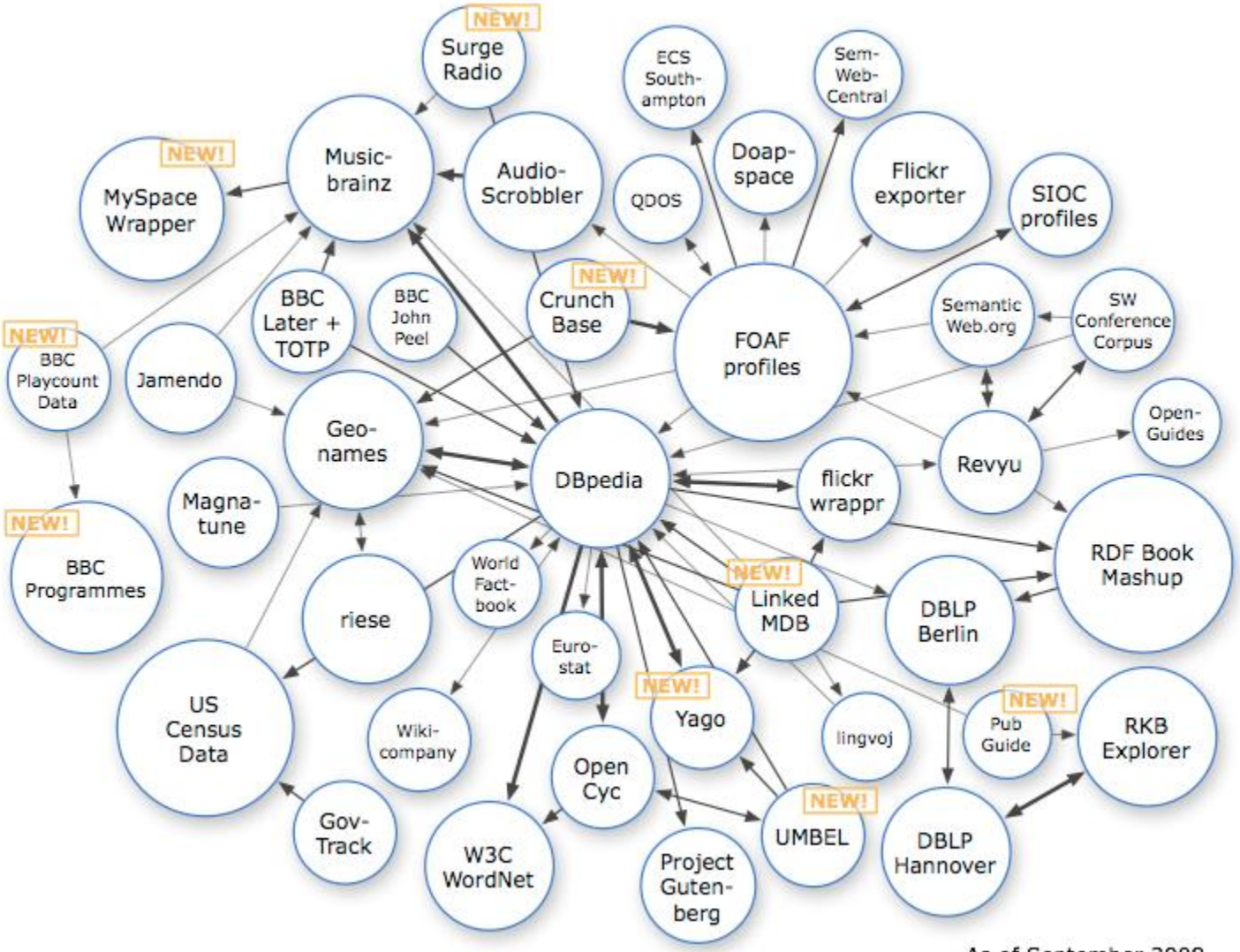
Linked Data evolution – Oct 2007



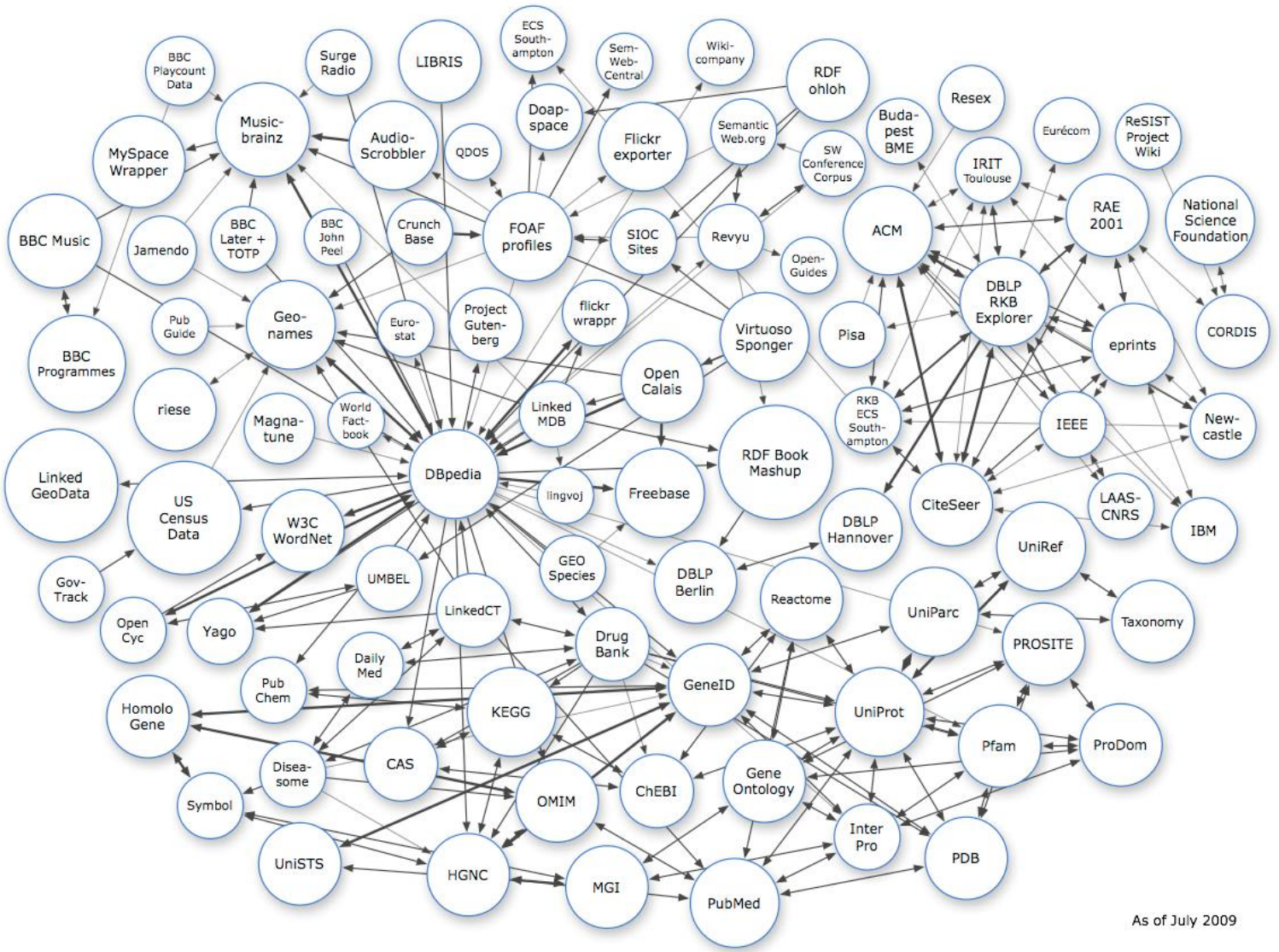
ontotext

GATE

Linked Data evolution – Sep 2008



Linked Data evolution – Jul 2009



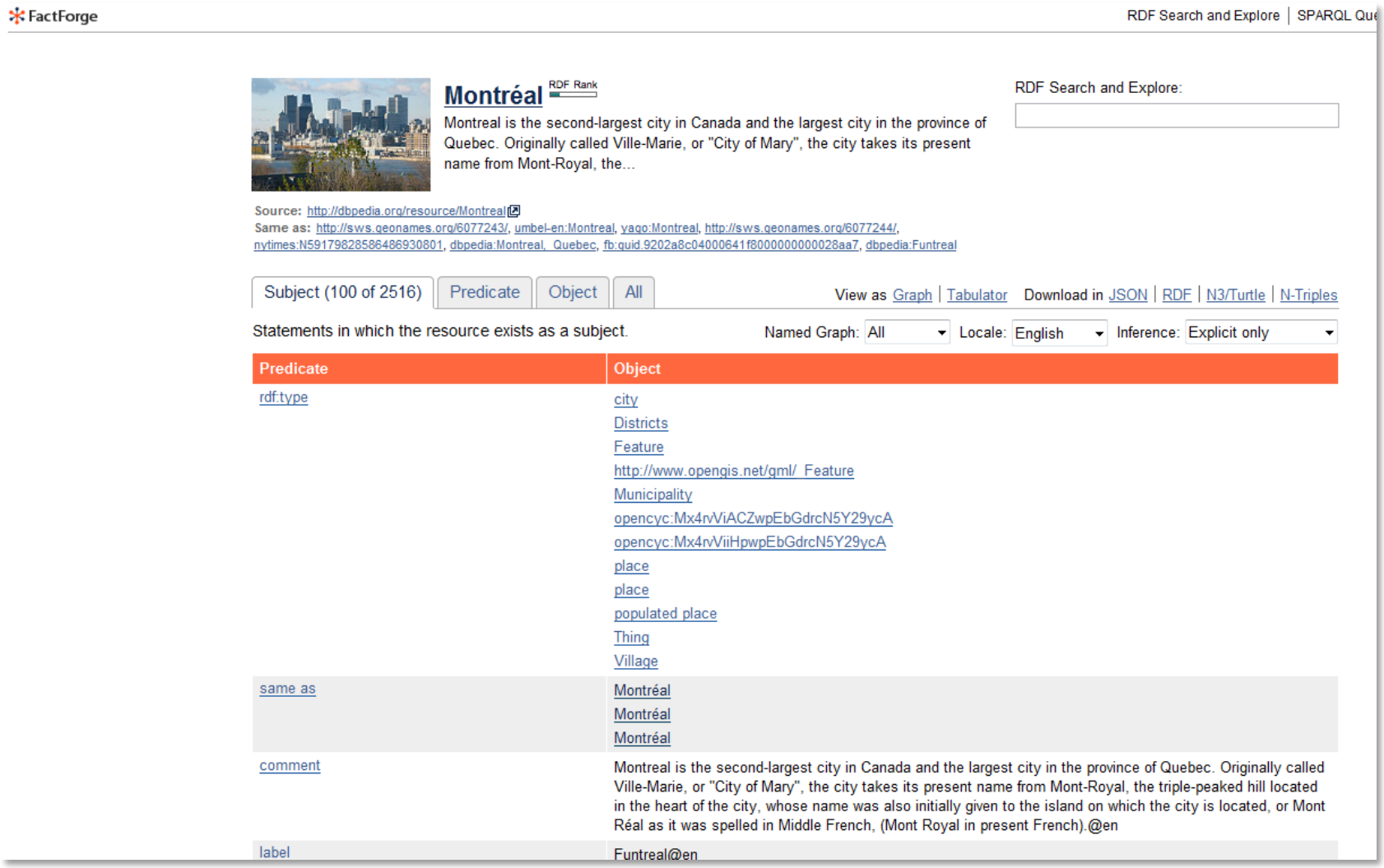
ontotext

GATE

Linked Data evolution – Jul 2010

- 50+ interlinked datasets
- 20 billion RDF triples
  - Data.gov + data.gov wiki – 11.5 billion
  - LinkedGeoData – 3 billion
  - UniProt – 1.1 billion
  - DBpedia – 1 billion
  - US Census Data – 1 billion
  - PubMed – 0.8 billion
  - AudioScrobbler – 0.6 billion
  - Freebase – 0.1 billion

Linked Data example – Montreal



FactForge

RDF Search and Explore | SPARQL Qu

Montréal RDF Rank

Montreal is the second-largest city in Canada and the largest city in the province of Quebec. Originally called Ville-Marie, or "City of Mary", the city takes its present name from Mont-Royal, the...

RDF Search and Explore:

Source: http://dbpedia.org/resource/Montreal
Same as: http://sws.geonames.org/6077243/, umbel-en:Montreal, yago:Montreal, http://sws.geonames.org/6077244/, nytimes:N59179828586486930801, dbpedia:Montreal,_Quebec, fb:guid.9202a8c04000641f8000000000028aa7, dbpedia:Funtreal

Subject (100 of 2516) | Predicate | Object | All

View as Graph | Tabulator Download in JSON | RDF | N3/Turtle | N-Triples

Statements in which the resource exists as a subject.

Named Graph: All Locale: English Inference: Explicit only

Predicate	Object
rdf:type	city Districts Feature http://www.opengis.net/gml/_Feature Municipality opencyc:Mx4rvViACZwpEbGdrcN5Y29ycA opencyc:Mx4rvViiHpwpEbGdrcN5Y29ycA place place populated place Thing Village
same as	Montréal Montréal Montréal
comment	Montreal is the second-largest city in Canada and the largest city in the province of Quebec. Originally called Ville-Marie, or "City of Mary", the city takes its present name from Mont-Royal, the triple-peaked hill located in the heart of the city, whose name was also initially given to the island on which the city is located, or Mont Réal as it was spelled in Middle French, (Mont Royal in present French).@en
label	Funtreal@en

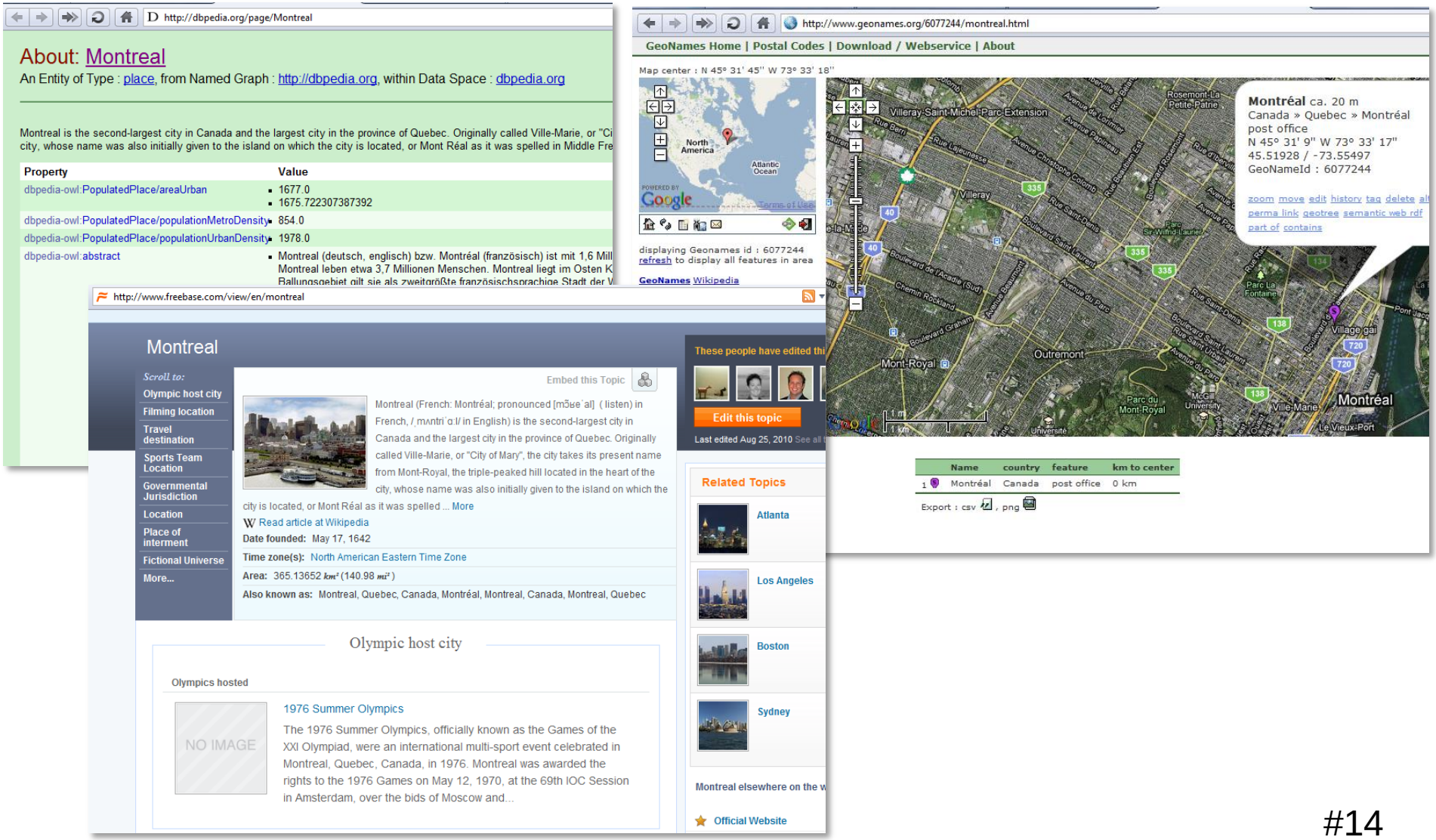
ontotext

GATE

Linked Data example (2)

- The description for Montreal on *FactForge* aggregates data from
  - DBPedia
  - GeoNames
  - Freebase
  - NY Times

Linked Data example (3)



ontotext

GATE

Why use Linked Data?

- Facilitate data integration
  - Use LOD as an “interlingua” for EDI
    - Additional public information can help alignment and linking
- Add value to proprietary data
  - Public data can allow enhanced content and more analytics on top of proprietary data
    - E.g. linking to spatial data from GeoNames, search for images
  - Better description and access to content
- Make enterprise data more open & accessible
  - Public identifiers and vocabularies can be used to access them

ontotext

GATE

Success Stories

- BBC Music
  - Integrates information from MusicBrainz and Wikipedia for artist/band infopages
  - Information also available in RDF (in addition to web pages)
  - 3rd party applications built on top of the BBC data
  - BBC also contributes data back to the MusicBrainz
- NY times
  - Maps its thesaurus of 1 million entity descriptions (people, organisations, places, etc) to DBpedia and Freebase

ontotext

GATE

Vocabularies & Datasets

ontotext

GATE

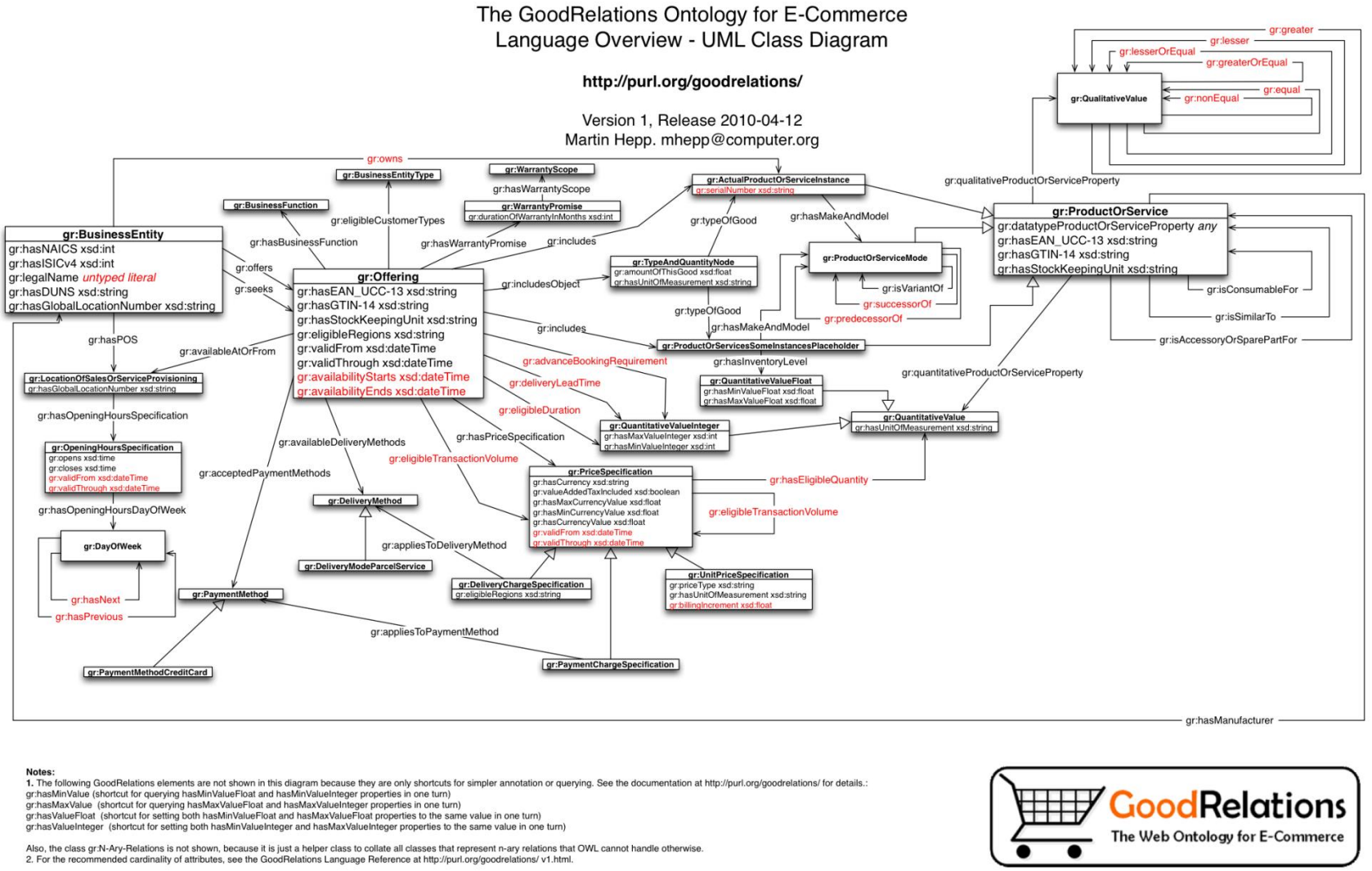
Vocabularies

- Existing vocabularies make publishing & integrating Linked Data easier
  - Friend-of-a-Friend (FOAF)
    - http://xmlns.com/foaf/0.1/
    - Vocabulary for describing people (names, contact info, …)
  - Dublin Core (DC)
    - http://dublincore.org/documents/dcmes-xml/
    - Vocabulary for general metadata attributes (author, topic, …)
  - Semantically-Interlinked Online Communities (SIOC)
    - http://sioc-project.org/
    - Social Web data

Vocabularies (2)

- Existing vocabularies (contd.)
  - SKOS
    - http://www.w3.org/2004/02/skos/
  - GoodRelations
    - Vocabulary for describing products and business entities
    - http://www.heppnetz.de/ontologies/goodrelations/v1
  - Music Ontology – http://musicontology.com/
  - Lined Open Description of Events (LODE) – http://linkedevents.org/ontology/
  - Creative Commons – http://creativecommons.org/ns

Vocabularies (3)



Notes:
1. The following GoodRelations elements are not shown in this diagram because they are only shortcuts for simpler annotation or querying. See the documentation at http://purl.org/goodrelations/ for details.:
gr:hasMinValue (shortcut for querying hasMinValueFloat and hasMinValueInteger properties in one turn)
gr:hasMaxValue (shortcut for querying hasMaxValueFloat and hasMaxValueInteger properties in one turn)
gr:hasValueFloat (shortcut for setting both hasMinValueFloat and hasMaxValueFloat properties to the same value in one turn)
gr:hasValueInteger (shortcut for setting both hasMinValueInteger and hasMaxValueInteger properties to the same value in one turn)

Also, the class gr:N-Ary-Relations is not shown, because it is just a helper class to collate all classes that represent n-ary relations that OWL cannot handle otherwise.
2. For the recommended cardinality of attributes, see the GoodRelations Language Reference at http://purl.org/goodrelations/ v1.html.

Red highlighting indicates elements added or changed in this release.

GoodRelations
The Web Ontology for E-Commerce

ontotext

GATE

Datasets

- DBpedia
  - Linked Data version of Wikipedia
  - 3.5 million entities, incl. 410K places, 310K persons, 146K species, 140K organisations, 95K music albums, 50K films, 33K buildings, 15K videogames, 5K diseases
  - Descriptions available in 90 languages
  - **1 billion** triples, 10 million links to external RDF datasets
  - Ontology – 260 classes, 1200 properties, 1.5 million instances
    - http://www4.wiwiss.fu-berlin.de/dbpedia/dev/ontology.htm

ontotext

GATE

Datasets (2)

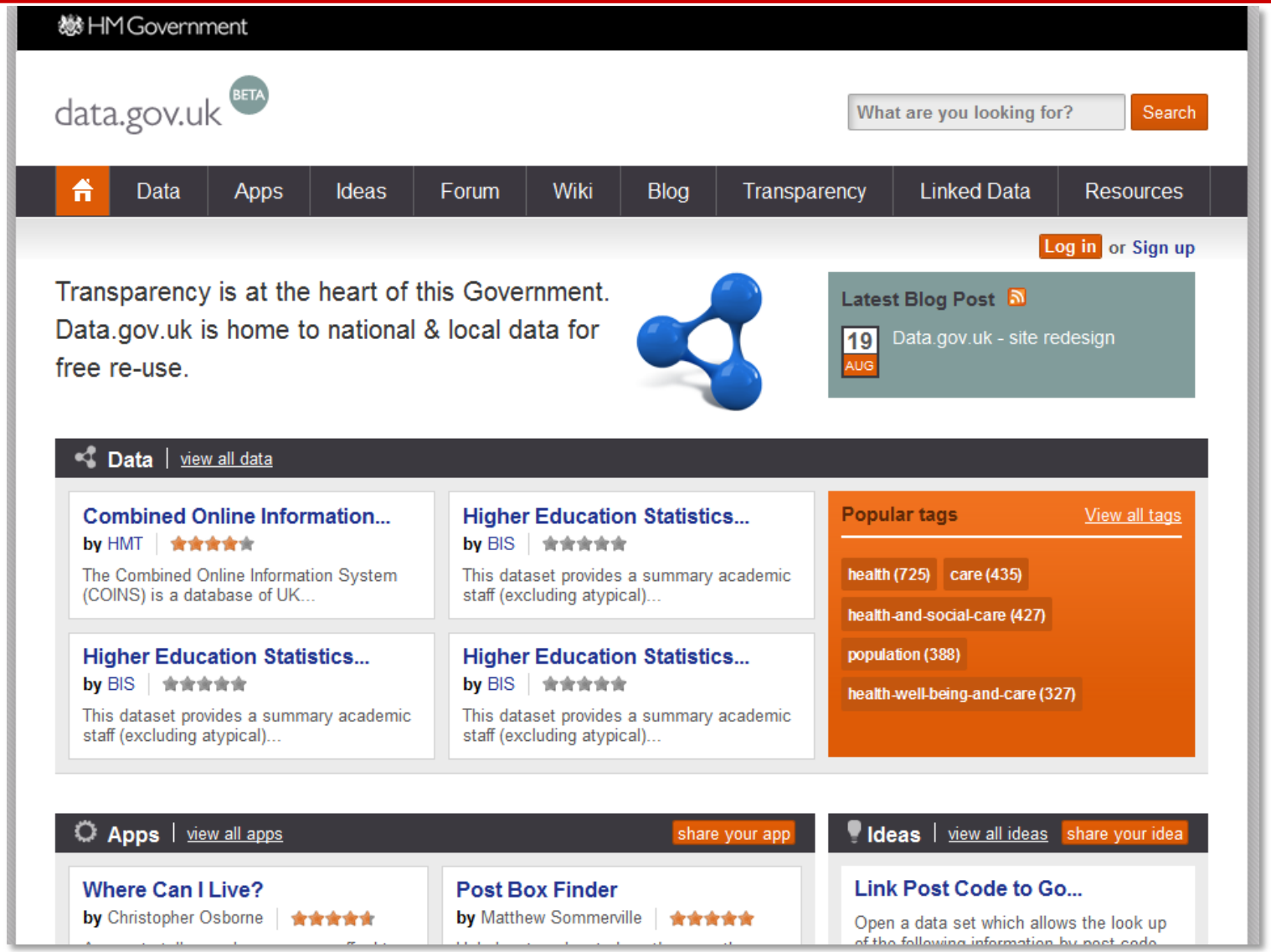
- Freebase
  - Similar to DBpedia
  - Higher data quality but *ten times* less data
- GeoNames
  - Information about 6 million places
  - Ontology: http://www.geonames.org/ontology/ontology_v2.1.rdf
- MusicBrainz
  - 55K artists, 22K albums, 36 million triples

Open Government Data

Data.gov


An Official Web Site of the United States Government
Thursday, August 19, 2010
Text: A+ A- A
Share
DATA.GOV
EMPOWERING PEOPLE
HOME
DATA
TOOLS
COMMUNITY
METRICS
DIALOGUE
GALLERY
WHAT'S NEW
DEEPWATER HORIZON RESPONSE
VIEW MORE
Most Popular Datasets
1. Food and Drug Administration--Recalls
2. Worldwide M1+ Earthquakes, Past 7 Days
3. AVAILABLE TECHNOLOGIES
4. TSCA Inventory
5. 2000 Federal Register in XML
SEARCH OUR CATALOGS
Search our catalogs..
SEARCH
APPS
With so much government data to work with, developers are creating a wide variety of
COMMUNITY
Data.gov is leading the way in democratizing public sector data and driving innovation. The data is being surfaced from many locations making the Government data stores available to researchers to perform their own analysis. Developers are finding good uses for the datasets, providing interesting and useful applications that allow for new views and public analysis. This is a work in progress, but this movement is spreading to cities, states, and
SEMANTIC WEB
As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to realize the promise of Linked Open Government Data.

Data.gov.uk


HM Government
data.gov.uk BETA
What are you looking for?
Search
Data
Apps
Ideas
Forum
Wiki
Blog
Transparency
Linked Data
Resources
Log in or Sign up
Transparency is at the heart of this Government. Data.gov.uk is home to national & local data for free re-use.
Latest Blog Post
19 AUG
Data.gov.uk - site redesign
Data | view all data
Combined Online Information...
by HMT
The Combined Online Information System (COINS) is a database of UK...
Higher Education Statistics...
by BIS
This dataset provides a summary academic staff (excluding atypical)...
Higher Education Statistics...
by BIS
This dataset provides a summary academic staff (excluding atypical)...
Higher Education Statistics...
by BIS
This dataset provides a summary academic staff (excluding atypical)...
Popular tags
View all tags
health (725)
care (435)
health-and-social-care (427)
population (388)
health-well-being-and-care (327)
Apps | view all apps
share your app
Ideas | view all ideas
share your idea
Where Can I Live?
by Christopher Osborne
Post Box Finder
by Matthew Sommerville
Link Post Code to Go...
Open a data set which allows the look up

Data.gov.uk (2)

- *"…we will aim for the **majority of government-published information to be reusable, linked data** by June 2011; and we will establish a common licence to reuse data which is interoperable with the internationally recognised Creative Commons model."* (UK Government, Dec 2009)

ontotext

GATE

Some statistics

- Data.gov
  - **2,400 datasets**
  - Contributed by 90 agencies
  - … but only 400 datasets RDFized at present
  - **6.5 billion triples** / 0.5 billion entities
- Data.gov.uk
  - **3000** datasets

ontotext

GATE

ThisWeKnow

This We Know:

About : Team : New

Explore U.S. Government Data About Your Community.

Enter a Location... SEARCH

e.g. "Bridgeport, CT", "90210", "Miami, FL", "Los Angeles, CA"

Least Toxins:

1. Altamonte Springs, FL
2. Dahlonega, GA
3. Bladensburg, MD
4. Republic, MO
5. Beaufort, NC

Explore More »

Source: 2005 Toxics Release Inventory

Most Nomadic:

1. Emlenton, PA
2. Fort Benning, GA
3. Fort Campbell, KY
4. Camp Lejeune, NC
5. Camp Pendleton, CA

Explore More »

Source: US Census 2000

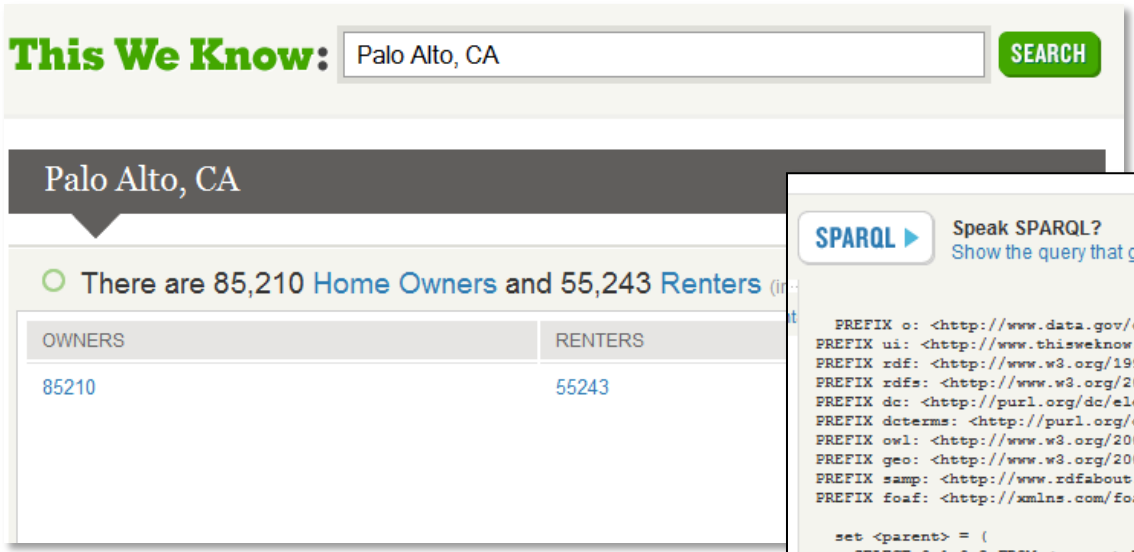
Lowest Unemployment:

1. Dickinson, ND
2. Miller, SD
3. Los Alamos, NM
4. Selden, KS
5. Fort Pierre, SD

Explore More »

Source: Local Area Unemployment Statisti

ThisWeKnow (2)



SPARQL ▶ Speak SPARQL?
Show the query that generated this page »

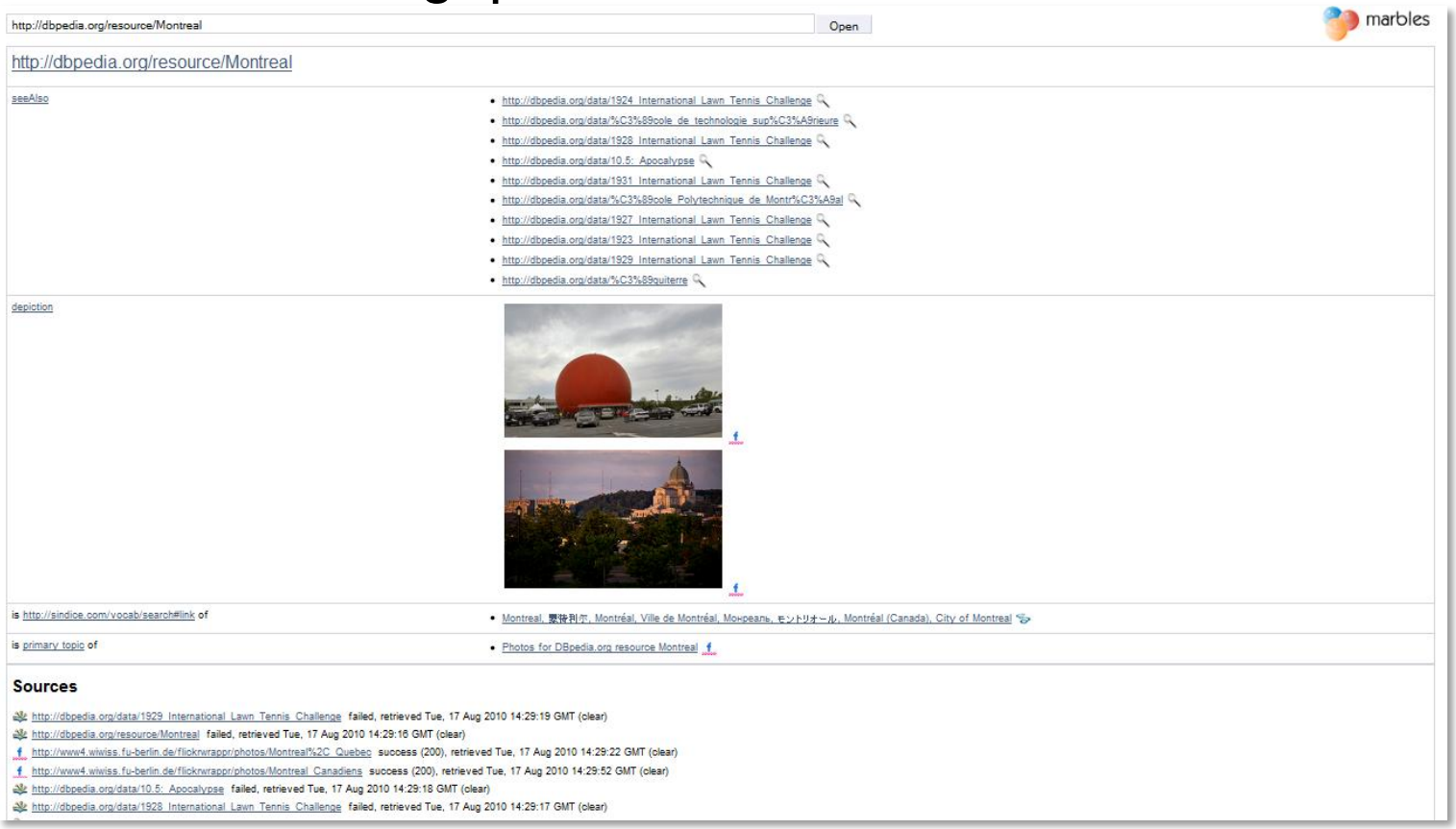
```
  PREFIX o: <http://www.data.gov/ontology#>
PREFIX ui: <http://www.thisweknow.org/ui#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX samp: <http://www.rdfabout.com/rdf/schema/uscensus/details/samp/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

  set <parent> = (
    SELECT ?p1 ?p2 FROM <census> WHERE {
        ?p1 dcterms:isPartOf ?p2 .
FILTER(?p1=<http://www.rdfabout.com/rdf/usgov/geo/us/ca/counties/san_mateo_county/san_mateo/east_palo_alto>)
    }
)

select ?owners ?town ?renters ?town
from <census_data2>
from <census_data>
from <census>
where {

?town <tag:govshare.info,2005:rdf/census/details> [
   samp:occupiedHousingUnits [
     samp:renterOccupied [rdf:value ?renters ];
     samp:ownerOccupied [rdf:value ?owners]
     ];
   samp:population5YearsAndOver[
       samp:differentHouseIn1995 [rdf:value ?different];
       samp:sameHouseIn1995 ?same;
       rdf:value ?pop
       ]
   ].
<parent>(?p1, ?p2) .
filter(?town=?p1||?town=?p2)
}
```

SPARQL is an RDF query language ThisWeKnow uses to formulate the information displayed on these pages.

Tools

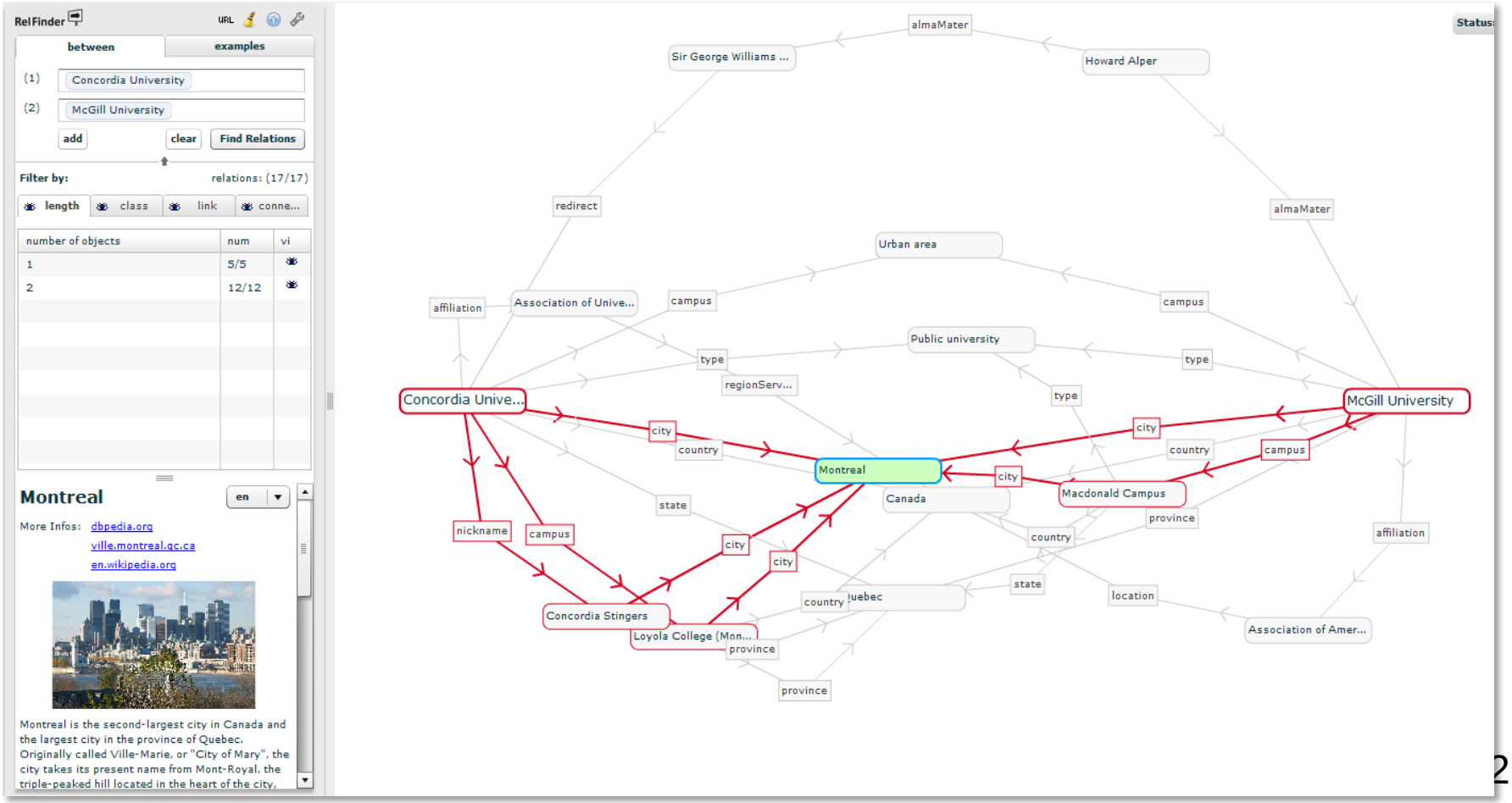
Linked Data browsers – Marbles

- http://marbles.sourceforge.net
- XHTML views of RDF data (SPARQL endpoint), caching, predicate traversal



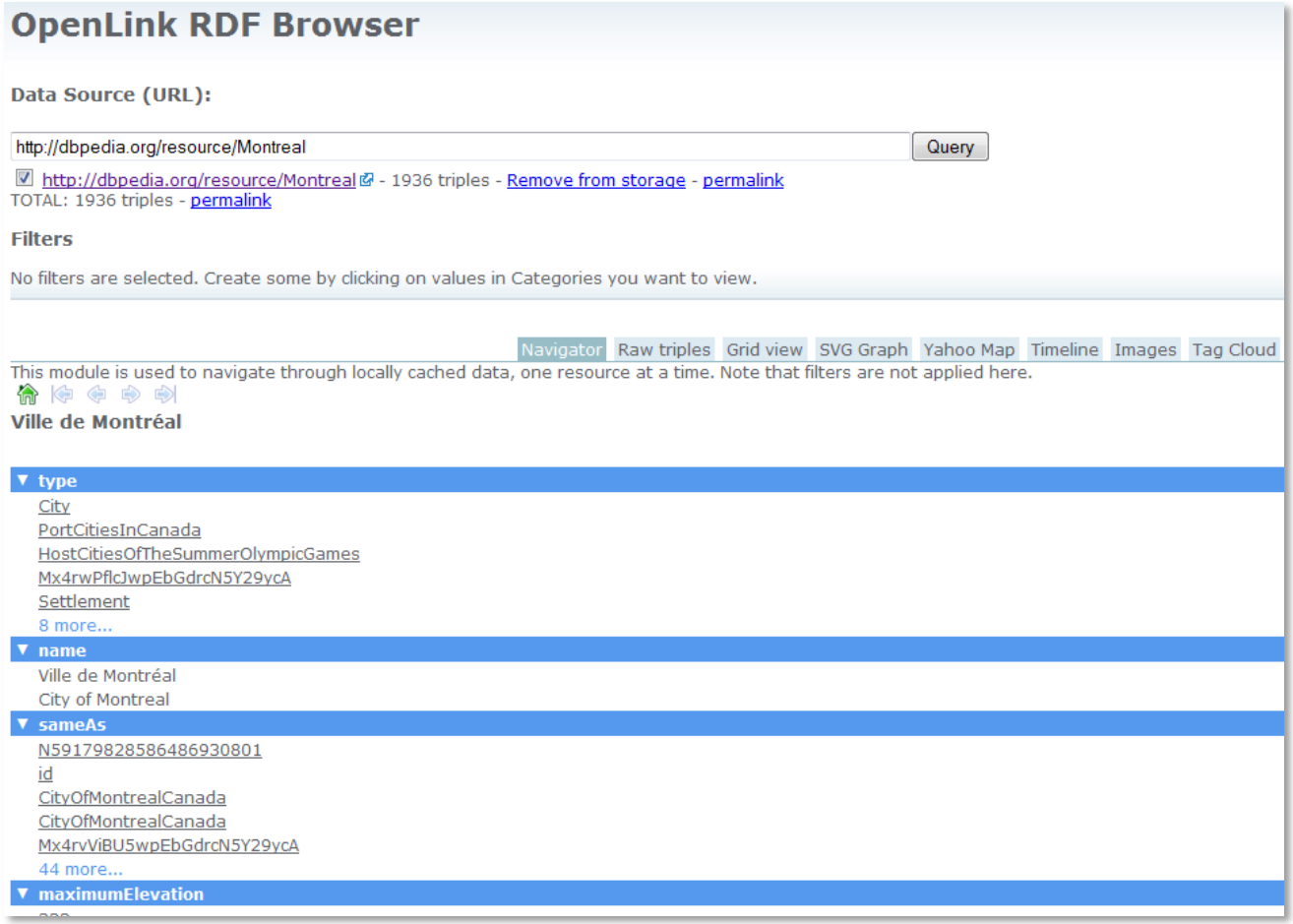
Linked Data browsers – RelFinder

- http://relfinder.dbpedia.org
- Explore & navigate relationships in a RDF graph



Linked Data browsers – OpenLink RDF Browser

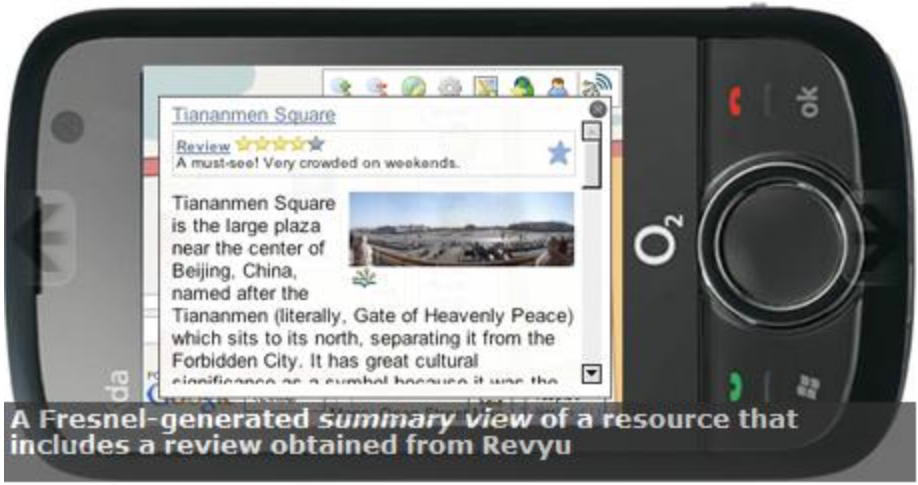
- http://demo.openlinksw.com/DAV/JS/rdfbrowser/index.html
- Explore & navigate relationships in a RDF graph



DBPedia Mobile

- http://wiki.dbpedia.org/DBpediaMobile
- Based on user's GPS position, renders a map with nearby places of interest (from DBpedia)

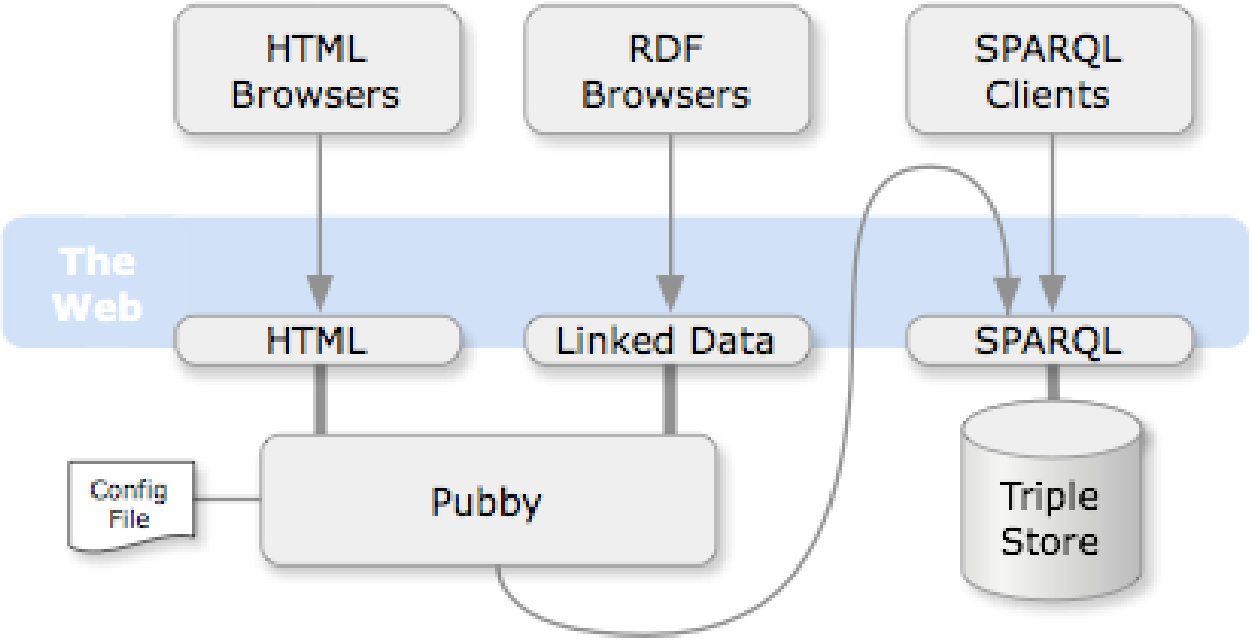
DBpedia Mobile's *map view* of resources in the user's proximity



A Fresnel-generated *summary view* of a resource that includes a review obtained from Revyu

Pubby – A Linked Data Frontend for SPARQL Endpoints

- http://www4.wiwiss.fu-berlin.de/pubby/
- Linked Data interface to local/remote SPARQL endpoints
- URI rewriting of SPARQL resultsets
- Simple HTML interface



ontotext

GATE

Open issues & challenges

ontotext

GATE

Linked Data – open issues

- LOD is **hard to comprehend**
  - Schema diversity & proliferation
- **Quality of data** is poor
  - Many of the datasets are well positioned to serve as “master data” but their quality is very far from the enterprise standards
  - No kind of consistency is guaranteed
- Issues with **reliability** of data end-points
  - High down-time is not unusual
  - There is no any kind of SLA provided

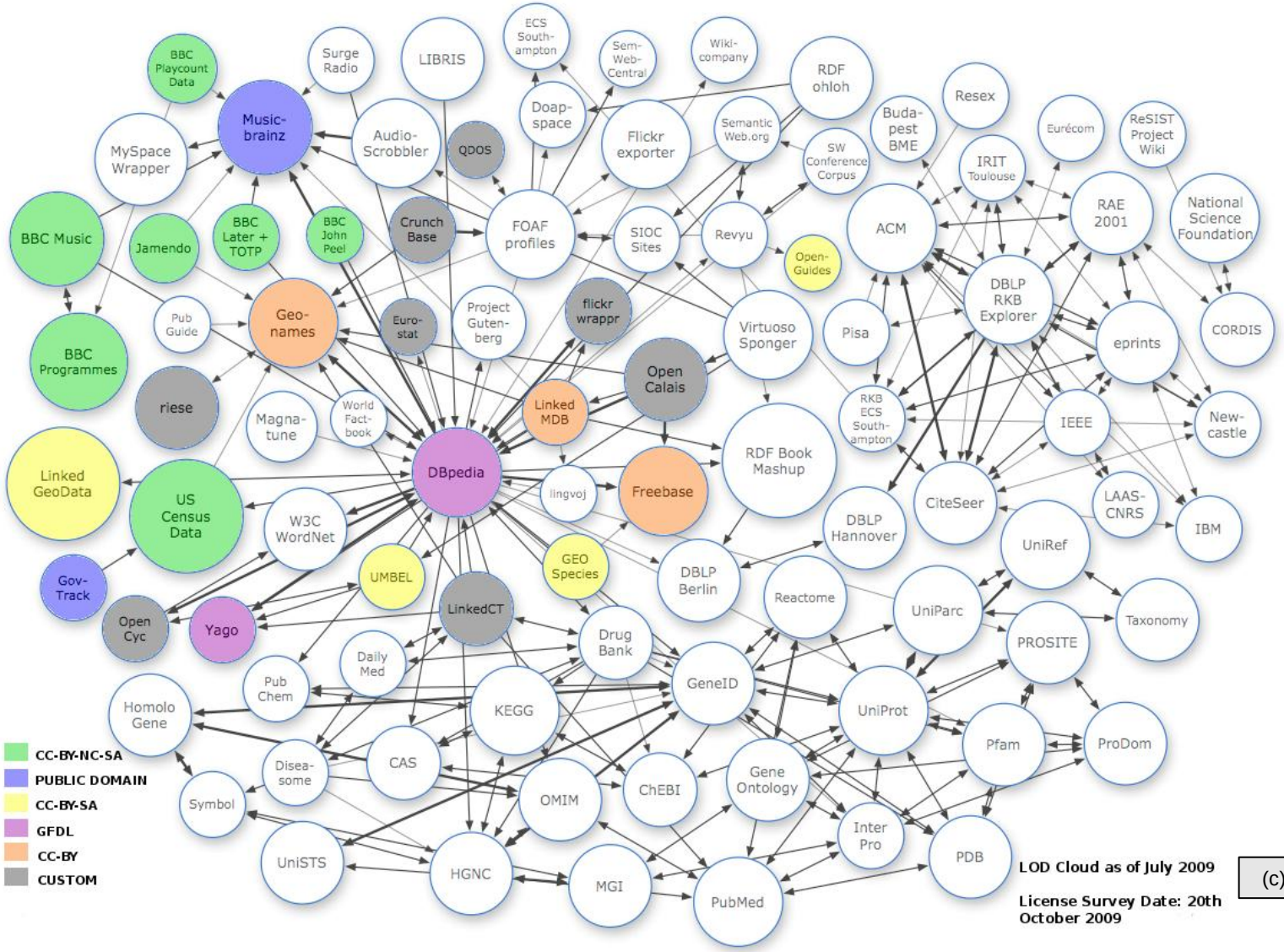
ontotext

GATE

Linked Data – open issues (2)

- **Querying** of linked data is slow
  - Data is distributed on the web
    - Federated SPARQL queries are slow
  - Even single SPARQL endpoints can be slow
    - Most end-points are experimental/research projects with no resources for quality guarantees
- **Licensing** issues
  - majority of datasets carry no explicit open license
  - Copyright-based licenses (CC) are difficult to apply to factual data

Linked Data – licensing issues



(c) Leigh Dodds

Weaving The *Pedantic* Web

- Initiative of DERI / KIT
- http://pedantic-web.org
- Goals
  - Analyse most common errors in RDF publishing
  - Propose possible approaches to avoid (publisher side) or deal with (consumer side) such errors

Weaving The *Pedantic* Web (2)

Category	Problem
Incompleteness	Dereferencability issues
	No structured data available
	Misreported content types
	RDF/XML Syntax Errors
Incoherence	Atypical use of collections, containers and reification
	Use of undefined classes and properties
	Misplaced classes/properties
	Misuse of *owl:DatatypeProperty (ObjectProperty)*
	Members of deprecated classes/properties
	Malformed datatype literals
	Literals incompatible with datatype range
	Ontology hijacking
Hijacking	Bogus *owl:InverseFunctionalProperty* values
	Ontology hijacking
Inconsistencies	Literals incompatible with datatype range
	OWL inconsistencies

Weaving The *Pedantic* Web (3)

- *Dereferencability issues*
  - URI lookup returns an error (violates 3rd LOD principle)
  - Or results in a redirect (with the wrong code)
- *No structured data available*
  - RDF data should be returned
- *Misreported content types*
  - A consumer application needs the correct content type in order to decide if it can consume the content (should be *application/rdf+xml*)

Weaving The *Pedantic* Web (4)

- *RDF/XML Syntax Errors*
- *Atypical use of collections, containers and reification*
- *Use of undefined classes and properties*
  - although not prohibited, ad-hoc/undefined classes and properties lead to more complex data integration and less effective inferences
- *Misplaced classes/properties*
  - Sometimes, a URI defined as a class is used as a property or vice versa (such usage ruins the inference)

Weaving The *Pedantic* Web (5)

- *Members of deprecated classes/properties*
- *Malformed datatype literals / Literals incompatible with datatype range*
- *Bogus owl:InverseFunctionalProperty values*
  - When two resources have the same value for an *inverse-functional property* the reasoner will treat them as equivalent
- *Ontology hijacking*
  - Redefinition by 3rd parties of external classes/properties affects the reasoner results

ontotext

GATE

Introduction to *FactForge* and *LinkedLifeData*

Reason-able Views to the Web of Data

- *Reason-able views* represent an approach for reasoning and management of linked data
  - Integrate selected datasets and ontologies in one dataset
    - Clean up, post-process and enrich the datasets if necessary
  - Load the compound dataset in a single RDF repository
  - Perform inference with respect to tractable OWL dialects
  - Define sample queries against the integrated dataset

ontotext

GATE

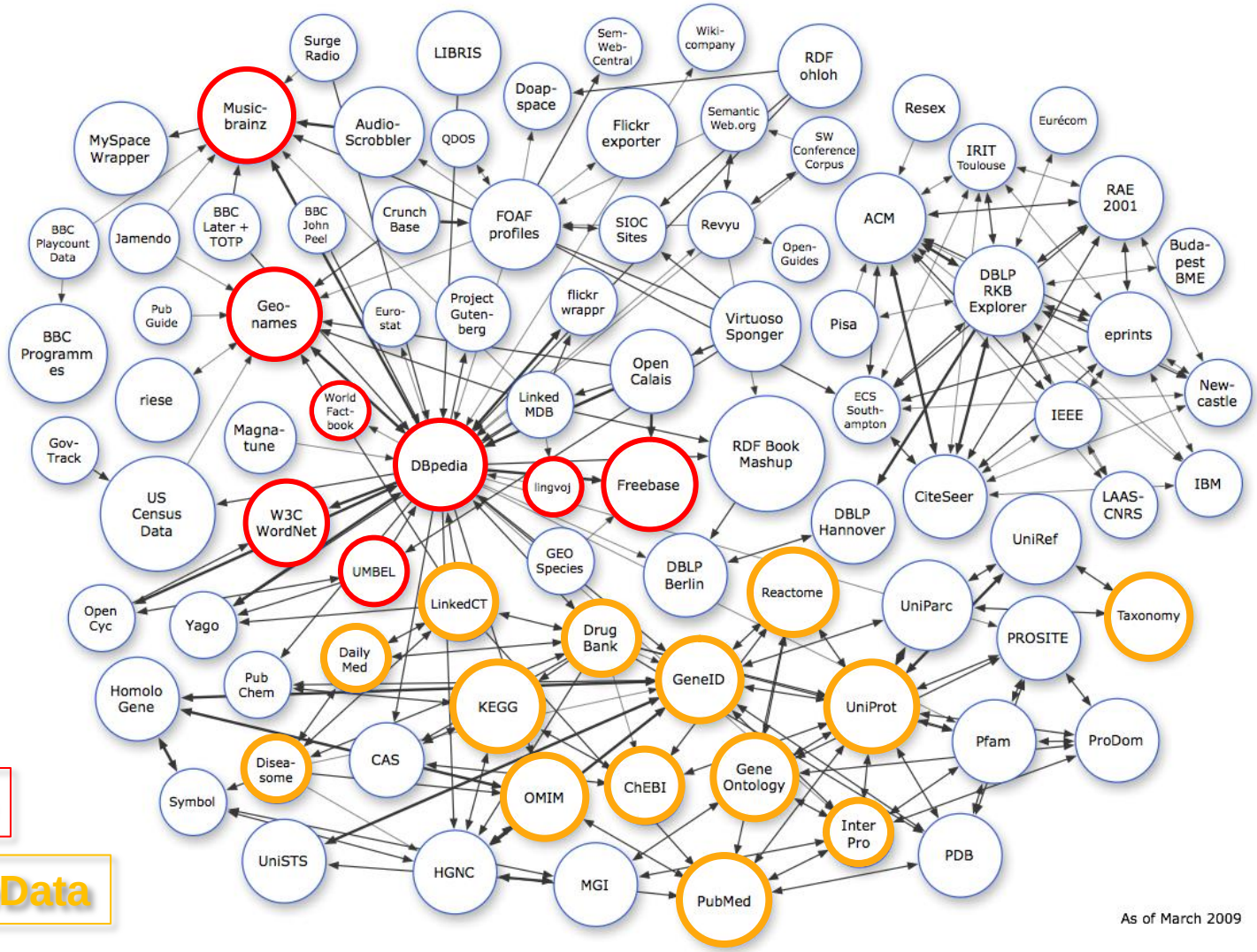
Reason-able Views: Objectives

- Make reasoning and query evaluation feasible
- Guarantee a basic level of consistency
  - The sample queries guarantee provide “regression tests” w.r.t. the consistency of the data
- Guarantee availability
- Better usability for querying and data exploration
  - URI auto-complete and RDF search
  - Sample queries provide re-usable extraction patterns, which reduce the time for learning about new datasets and their inter-relations

Two Reason-able Views to the Web of Linked Data

- *FactForge*
  - Integrates some of the most central LOD datasets
  - General-purpose information (not specific to a domain)
  - 1.2B explicit plus 1B inferred statements (10B retrievable)
  - The largest upper-level knowledge base
  - **http://www.FactForge.net/**
- *Linked Life Data*
  - 25 of the most popular life-science datasets
  - 2.7B explicit and 1.4B inferred triples
  - **http://www.LinkedLifeData.com**

Linking Open Data Datasets and Views



FactForge

LinkedLifeData

FactForge: Fast Track to the Center of the Web of Data

- Datasets
  - DBPedia, Freebase, Geonames, UMBEL, MusicBrainz, Wordnet, CIA World Factbook, Lingvoj
- Ontologies
  - Dublin Core, SKOS, RSS, FOAF
- Inference
  - materialization with respect to OWL 2 RL
  - owl:sameAs optimization in OWLIM allows reduction of the indices without loss of semantics

FactForge: Fast Track to the Center of the Web of Data (2)

- Free public service at **http://www.FactForge.net**
  - Very fast incremental URI auto-completion
  - Querying and exploration through *Forest* and *Tabulator*
  - RDF Search: retrieve ranked list of URIs by keywords
  - SPARQL end-point

FactForge – Loading and Inference Statistics

Dataset	Explicit Indexed Triples ('000)	Inferred Indexed Triples ('000)	Total # of Stored Triples ('000)	Entities ('000 of nodes in the graph)	Inferred closure ratio
Sechmata and ontologies	11	7	18	6	0.6
DBpedia (categories)	2,877	42,587	45,464	1,144	14.8
DBpedia (sameAs)	5,544	566	6,110	8,464	0.1
UMBEL	5,162	42,212	47,374	500	8.2
Lingvoj	20	863	883	18	43.8
CIA Factbook	76	4	80	25	0.1
Wordnet	2,281	9,296	11,577	830	4.1
Geonames	91,908	125,025	216,933	33,382	1.4
DBpedia core	560,096	198,043	758,139	127,931	0.4
Freebase	463,689	40,840	504,529	94,810	0.1
MusicBrainz	45,536	421,093	466,630	15,595	9.2
Total	**1,177,961**	**881,224**	**2,058,185**	**283,253**	**0.7**

ontotext

GATE

Fact Forge – post-processing

- Several kinds of post-processing were performed
  - Goal: to allow easier navigation and browsing
  - E.g. preferred labels, text snippets, RDF Rank for nodes
    - Results available through system predicates
- Final Statistics
  - Number of entities (RDF graph nodes): **405M**
  - Number of inserted statements (NIS): **1.2B**
  - Number of stored statements (NSS): **2.2B**
  - Number of retrievable statements (NRS): **9.8B**
    - 7.6B statements “compressed” through OWLIM’s owl:sameAs optimisation

Guess who is the most popular German entertainer?

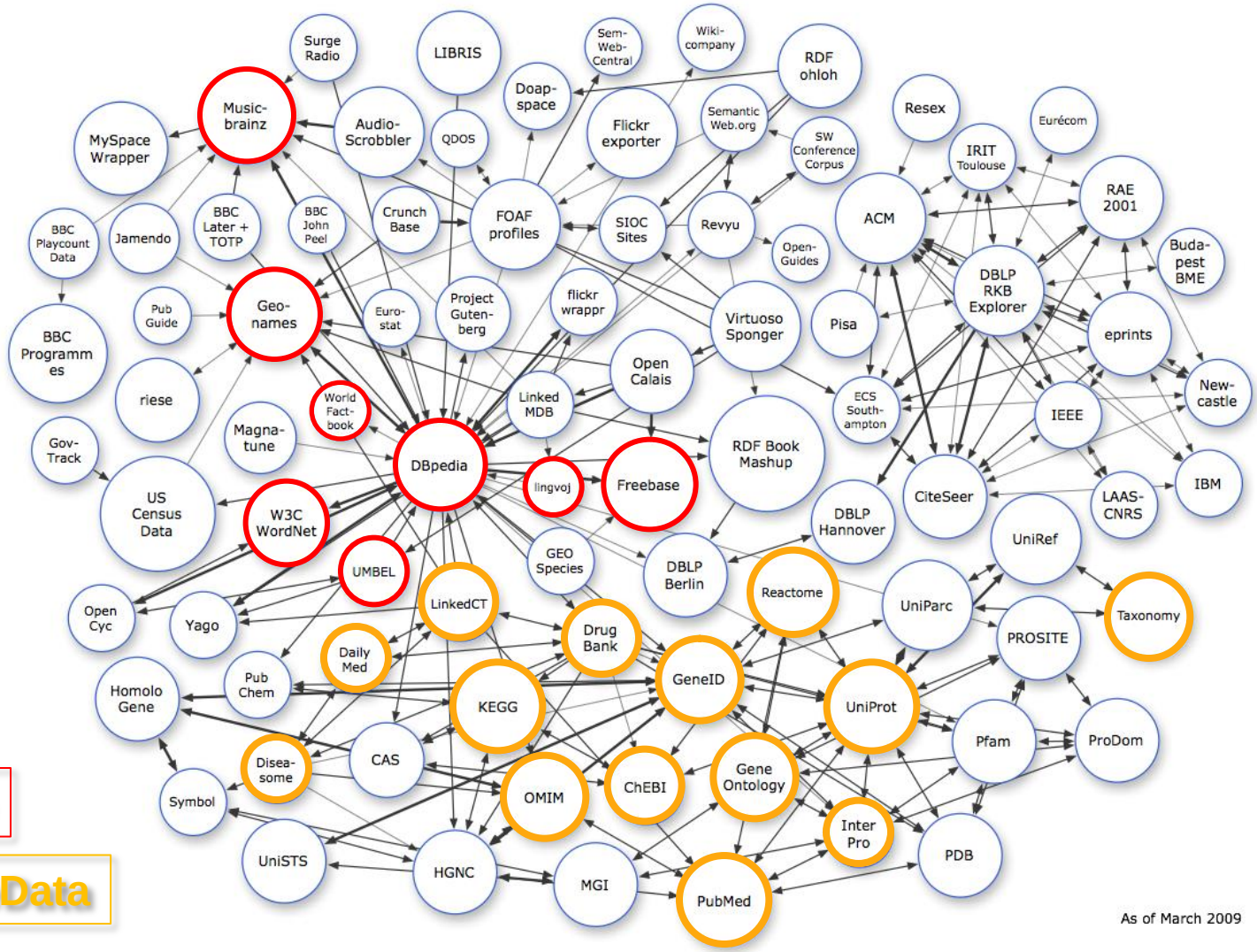
```
PREFIX rdf: …          (run the query at http://factforge.net/sparql)

SELECT * WHERE {
     ?Person dbp-ont:birthPlace ?BirthPlace ;
             rdf:type opencyc:Entertainer ;
             ff:hasRDFRank ?RR .
     ?BirthPlace geo-ont:parentFeature dbpedia:Germany .
} ORDER BY DESC(?RR) LIMIT 100
```

- Without FF, answering such queries in real time is impossible
  - Used data from: DBPedia, Geonames, UMBEL and MusicBrainz
- The most popular entertainer born in Germany is:
  - Asking factual questions to a global KB can bring unexpected and strange results **F. Nietzsche**
  - We ask who is the most popular person, who qualifies as an entertainer
  - It uses a simple notion of popularity: RDFRank

Linked Life Data

- Quick facts
  - Integrates more than 25 popular data sources
  - 5 billion RDF statements, 0.5 billion entities
  - Querying & exploration of integrated data
  - Public SPARQL end point
  - http://linkedlifedata.com/

LLD data sources



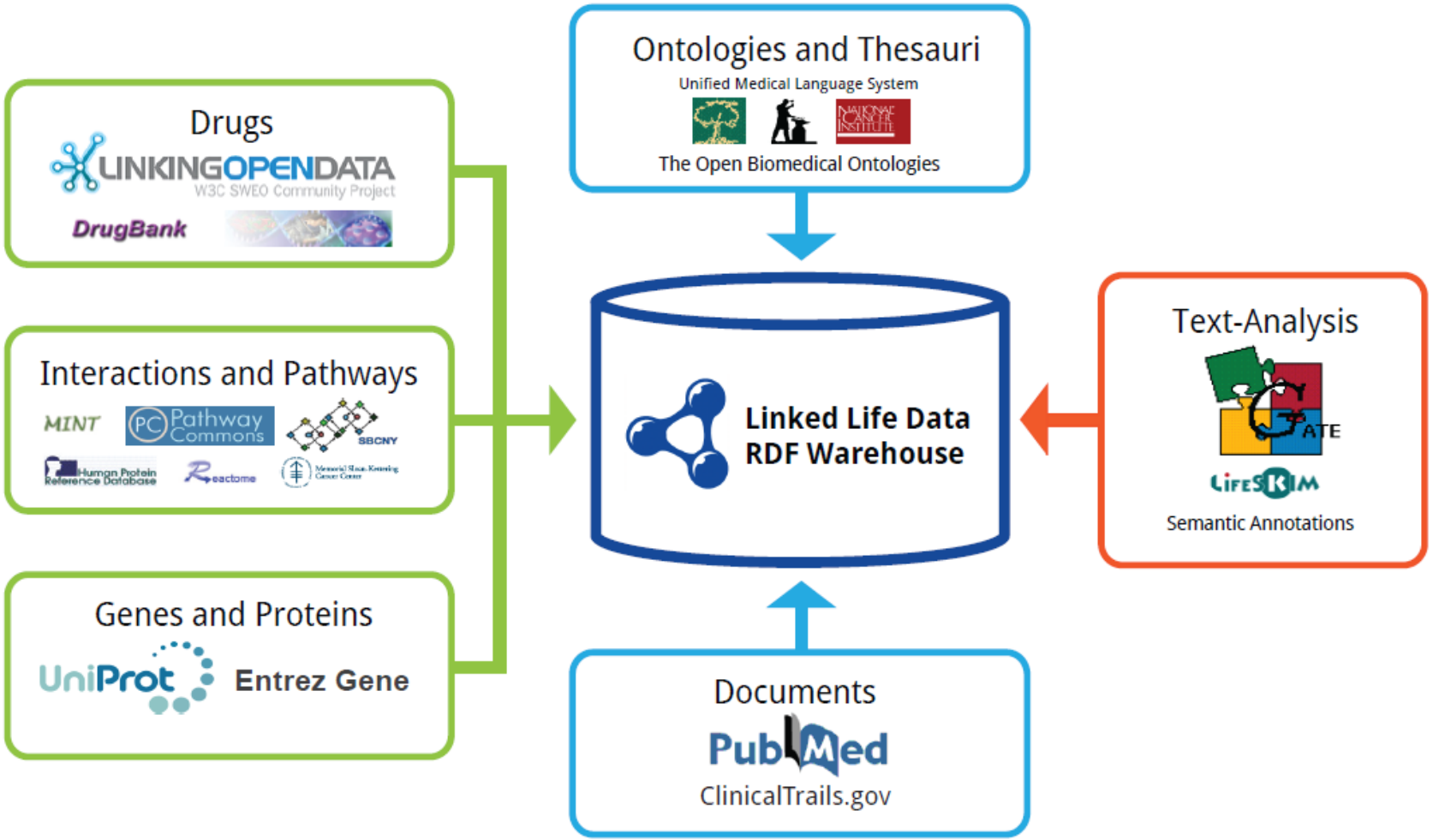
FactForge

LinkedLifeData

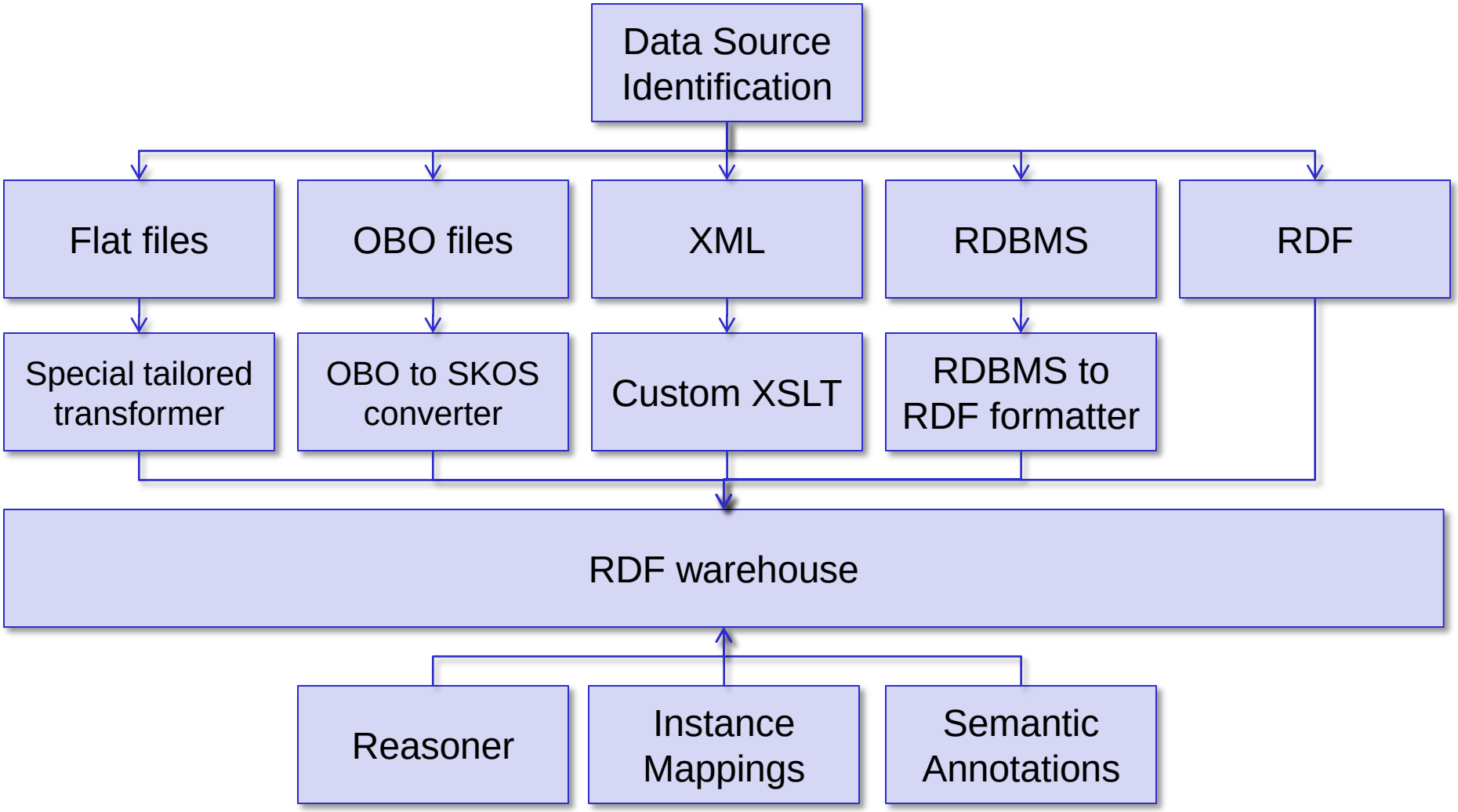
ontotext

GATE

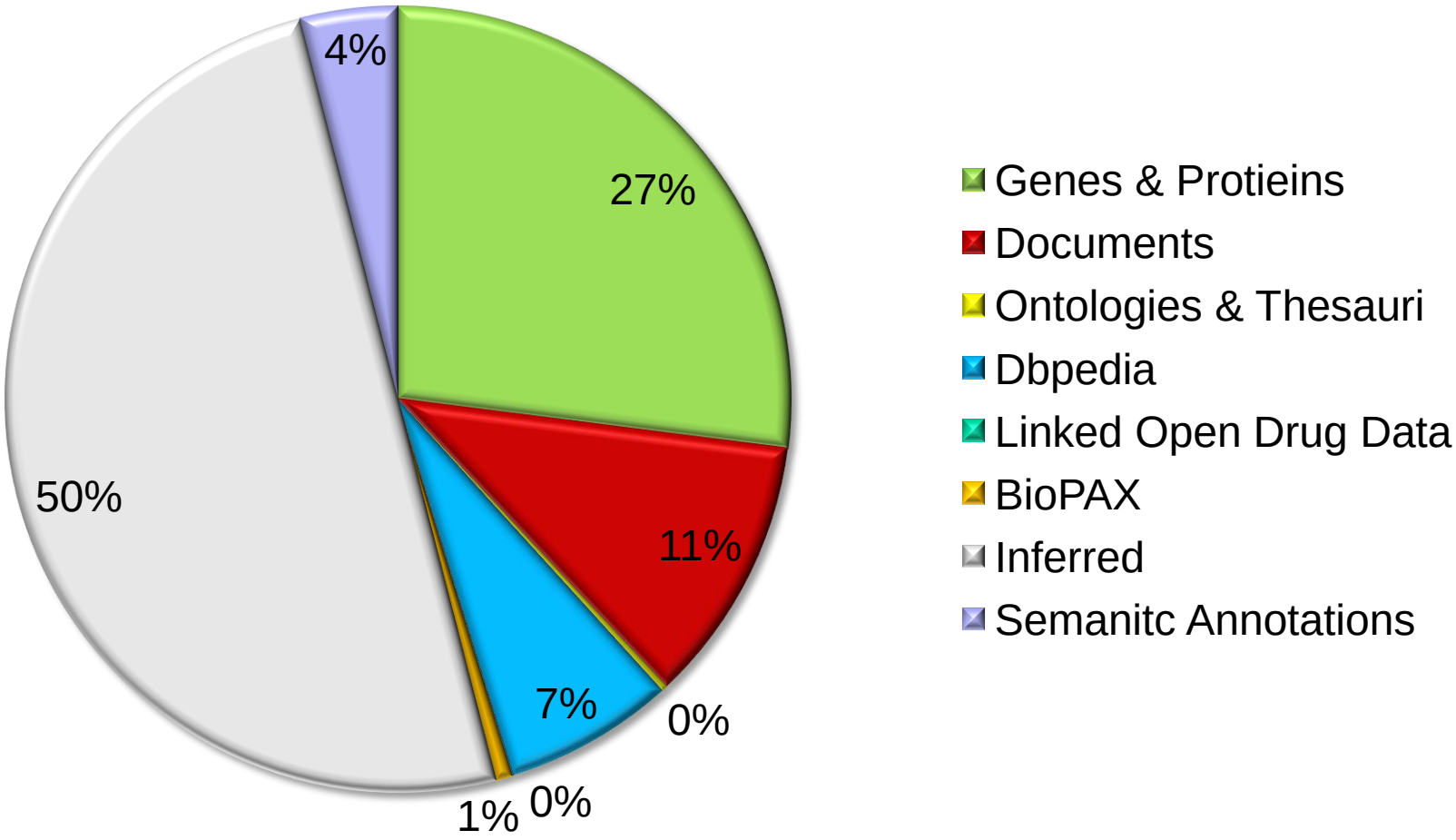
Linked Life Data – ETL process



Linked Life Data – ETL process (2)



Linked Life Data – triple distribution



ontotext

GATE

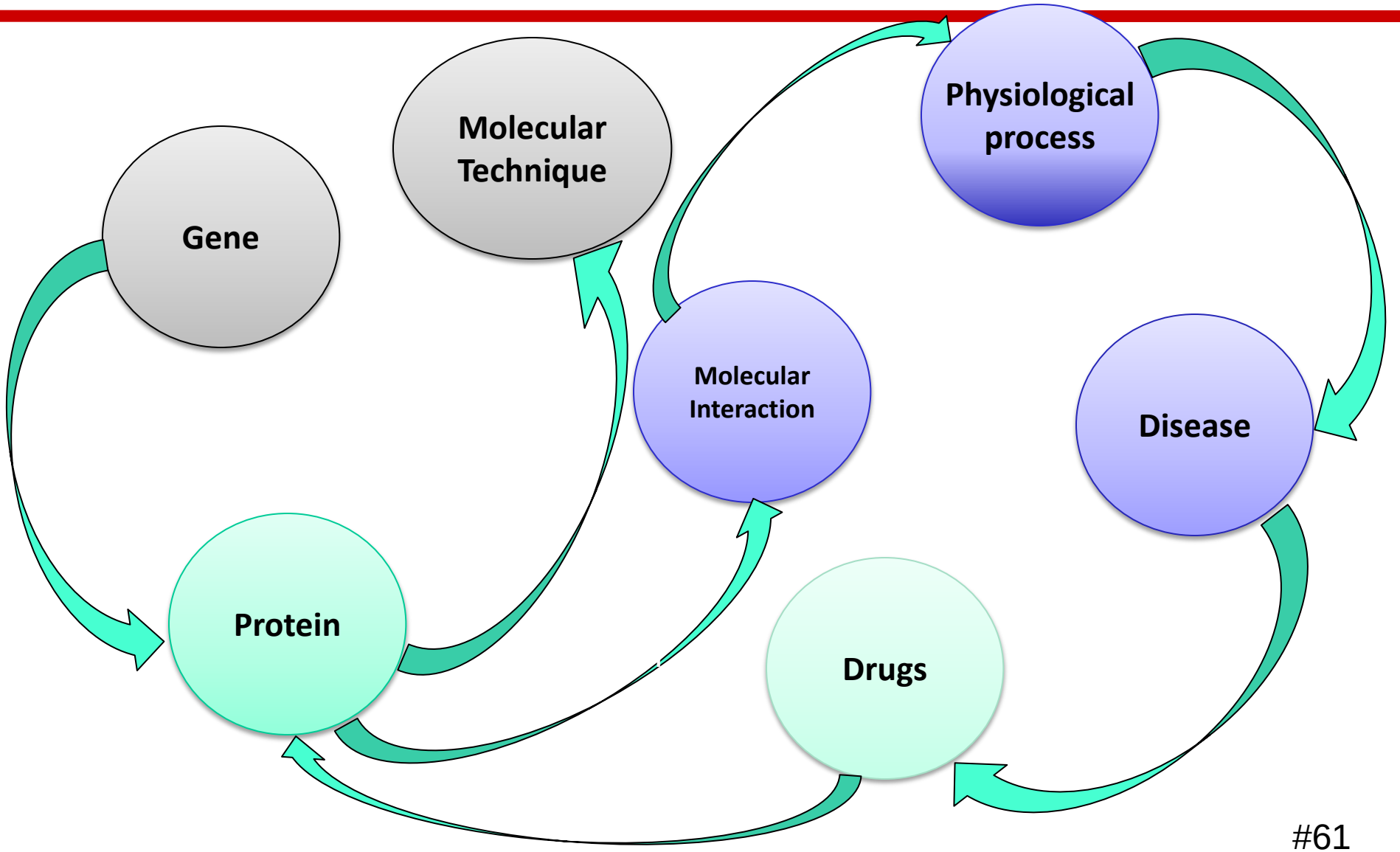
Different Types of Semantic Search

Search Type	Example Queries
Simple Keyword Search	Documents that includes given regular expressions
Text Analysis	Documents that refer to a gene and a disease
Semantic Search	Documents that contain a human gene and a respiratory disease
The Impossible Query	Tell me what is it the best drug to take

ontotext

GATE

Complex Cross-domain Query



ontotext

GATE

New Type of Possible Query #1

Select drugs related to asthma that are linked to a curated molecular interaction in the literature where the protein is known to cause inflammatory response

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX biopax2: <http://www.biopax.org/release/biopax-
        level2.owl#>
PREFIX uniprot: <http://purl.uniprot.org/core/>
PREFIX drugbank: <http://www4.wiwiss.fu-
        berlin.de/drugbank/resource/drugbank/>

SELECT DISTINCT ?fullname ?drugname
WHERE {
   ?interaction rdf:type biopax2:physicalInteraction .
   ?interaction biopax2:PARTICIPANTS ?participant .
   ?participant biopax2:PHYSICAL-ENTITY ?physicalEntity .
   ?physicalEntity skos:exactMatch ?protein .
   ?protein uniprot:classifiedWith
        <http://purl.uniprot.org/go/0006954>.
   ?protein uniprot:recommendedName ?name.
   ?name uniprot:fullName ?fullname .
   ?target skos:exactMatch ?protein .
   ?drug drugbank:target ?target .
   ?drug drugbank:genericName ?drugname .
   ?drug drugbank:indication ?indication .
}
```

The red graph patterns indicate the usage of mapping rules.

New Type of Possible Query #2

Select all located in Y-chromosome, human genes with known molecular interactions, which are analysed with 'Transfection'

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX gene:
<http://linkedlifedata.com/resource/entrezgene/>
PREFIX core: <http://purl.uniprot.org/core/>
PREFIX biopax2: <http://www.biopax.org/release/biopax-
level2.owl#>
PREFIX lifeskim:
<http://linkedlifedata.com/resource/lifeskim/>
PREFIX umls: <http://linkedlifedata.com/resource/umls/>
PREFIX pubmed:
<http://linkedlifedata.com/resource/pubmed/>

SELECT distinct ?genedescription ?prefLabel ?pmid
WHERE {
   ?interaction rdf:type biopax2:interaction .
   ?interaction biopax2:PARTICIPANTS ?p .
   ?p biopax2:PHYSICAL-ENTITY ?protein .
   ?protein skos:exactMatch ?uniprotaccession .
   ?uniprotaccession core:organism
<http://purl.uniprot.org/taxonomy/9606> .
   ?geneid gene:uniprotAccession ?uniprotaccession .
   ?geneid gene:description ?genedescription .
   ?geneid gene:pubmed ?pmid .
   ?geneid gene:chromosome 'Y' .
   ?pmid lifeskim:mentions ?umlsid .
   ?umlsid skos:prefLabel 'Transfection' .
   ?umlsid skos:prefLabel ?prefLabel .
}
```

ontotext

GATE

Query Results

Results for PREFIX rdf: <http://www.w3.... (14) View as Exhibit Download in JSON | SPARQL Results in XML | SPARQL Results in JSON

genedescription	prefLabel	pmid
interleukin 9 receptor	Transfection	pubmed-citation:1376929
RNA binding motif protein, Y-linked, family 1, member A1	Transfection	pubmed-citation:11149922
vesicle-associated membrane protein 7	Transfection	pubmed-citation:18362137
CD99 molecule	Transfection	pubmed-citation:16421247
colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)	Transfection	pubmed-citation:12504125
interleukin 3 receptor, alpha (low affinity)	Transfection	pubmed-citation:12504125
sex determining region Y	Transfection	pubmed-citation:18454134
jumonji, AT rich interactive domain 1D	Transfection	pubmed-citation:9143681
zinc finger, BED-type containing 1	Transfection	pubmed-citation:12663651
sex determining region Y	Transfection	pubmed-citation:9346931
short stature homeobox	Transfection	pubmed-citation:12960152
CD99 molecule	Transfection	pubmed-citation:16984917
thymosin beta 4, Y-linked	Transfection	pubmed-citation:15557202
solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	Transfection	pubmed-citation:14746803

New Type of Possible Query #3

Select all participating in interactions human genes which are drug target and are analysed with 'Transfection'

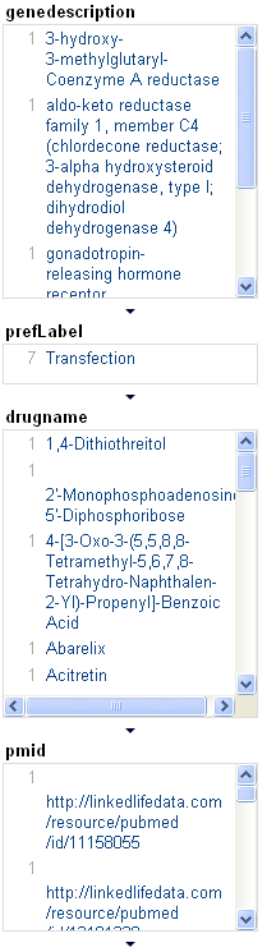
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX gene: <http://linkedlifedata.com/resource/entrezgene/>
PREFIX core: <http://purl.uniprot.org/core/>
PREFIX biopax2: <http://www.biopax.org/release/biopax-
level2.owl#>
PREFIX lifeskim: <http://linkedlifedata.com/resource/lifeskim/>
PREFIX umls: <http://linkedlifedata.com/resource/umls/>
PREFIX pubmed: <http://linkedlifedata.com/resource/pubmed/>
PREFIX drugbank: <http://www4.wiwiss.fu-
berlin.de/drugbank/resource/drugbank/>

SELECT distinct ?genedescription ?prefLabel ?drugname ?pmid
    WHERE {
        ?interaction rdf:type biopax2:interaction .
        ?interaction biopax2:PARTICIPANTS ?p .
        ?p biopax2:PHYSICAL-ENTITY ?protein .
        ?protein skos:exactMatch ?uniprotaccession .
        ?uniprotaccession core:organism
<http://purl.uniprot.org/taxonomy/9606> .
        ?geneid gene:uniprotAccession ?uniprotaccession .
        ?geneid gene:description ?genedescription .
        ?geneid gene:pubmed ?pmid .
        ?pmid lifeskim:mentions ?umlsid .
        ?umlsid skos:prefLabel 'Transfection' .
        ?umlsid skos:prefLabel ?prefLabel .
        ?target skos:closeMatch ?geneid.
        ?drug drugbank:target ?target .
        ?drug rdfs:label ?drugname .
    }
```

Query Results

SPARQL Query

7 record(s)

genedescription	prefLabel	drugname	pmid
S100 calcium binding protein P	Transfection	Cromoglicate	http://linkedlifedata.com/resource/pubmed/id/16061848
retinoic acid receptor, gamma	Transfection	SR11254, Tazarotene, Alitretinoin, Dodecyl-Alpha-D-Maltoside, CD564, Tretinoin, Etretinate, Adapalene, BMS184394, Acitretin, and 4-[3-Oxo-3-(5,5,8,8-Tetramethyl-5,6,7,8-Tetrahydro-Naphthalen-2-Yl)-Propenyl]-Benzoic Acid	http://linkedlifedata.com/resource/pubmed/id/1318502
progesterone receptor	Transfection	Progesterone, Mifepristone, Methyltrienolone, Megestrol, Dydrogesterone, Norgestimate, Norgestrel, Tanaproget, Desogestrel, Norethindrone, Levonorgestrel, Drospirenone, Etonogestrel, Medroxyprogesterone, and Ethynodiol Diacetate	http://linkedlifedata.com/resource/pubmed/id/12101239, http://linkedlifedata.com/resource/pubmed/id/16647340, http://linkedlifedata.com/resource/pubmed/id/15084343, and http://linkedlifedata.com/resource/pubmed/id/15084345
interleukin 8	Transfection	Ketoprofen, Simvastatin, Zileuton, and Salbutamol	http://linkedlifedata.com/resource/pubmed/id/14645117, http://linkedlifedata.com/resource/pubmed/id/17035306, and http://linkedlifedata.com/resource/pubmed/id/15039334
gonadotropin-releasing hormone receptor	Transfection	Nafarelin, Leuprolide, Danazol, Cetrorelix, Abarelix, and Gonadorelin	http://linkedlifedata.com/resource/pubmed/id/16613990
aldo-keto reductase family 1, member C4 (chlordecone reductase; 3-alpha hydroxysteroid dehydrogenase, type I; dihydrodiol dehydrogenase 4)	Transfection	NADH	http://linkedlifedata.com/resource/pubmed/id/11158055
3-hydroxy-3-methylglutaryl-Coenzyme A reductase	Transfection	1,4-Dithiothreitol, Pravastatin, Rosuvastatin, Adenosine-5'-Diphosphate, Simvastatin, NADH, 2'-Monophosphoadenosine 5'-Diphosphoribose, Atorvastatin, and Lovastatin	http://linkedlifedata.com/resource/pubmed/id/14697242



ontotext

GATE

The “Modigliani test” for the Semantic Web

The tipping point for the Semantic Web

- http://www.readwriteweb.com/archives/the_modigliani_test_semantic_web_tipping_point.php

Key Trends ▾ Top Topics ▾ Channels ▾ Reports ▾ International ▾

ReadWriteWeb

Home | Archives | Features | Tags | Best of RWW | Featured: **ReadWriteCloud** | **VS 2010 Competition**

The Modigliani Test: The Semantic Web's Tipping Point

Written by Richard MacManus / April 16, 2010 12:06 AM / 21 Comments

« Prior Post Next Post »

In our recent posts about Structured Data, we've emphasized that most of the current initiatives have been around uploading new data to the Web - whatever the format. The U.S. and U.K. governments have led the way with their 'open data' websites, but much of that data isn't 'linked' yet. In other words, it's online - but siloed. So how do we get to the next stage of the Semantic Web, linking disparate data sets together so that people can begin to ***use*** that data?

The tipping point for the long-awaited Semantic Web may be when you can query a set of data about someone not too famous, and get a long list of structured results in return. I've decided to term this 'The Modigliani Test.'

Amedeo Modigliani is one of my favorite artists. He was moderately famous during the early 20th century and has something of a cult following nowadays. But he's not Da Vinci or Picasso famous. What I'd like to do in a Semantic Web is type the following query into a search engine and get

The tipping point for the Semantic Web (2)

- Richard McManus (ReadWriteWeb)
  - *"…the tipping point for the Semantic Web may be when one can … deliver – using Linked Data – a comprehensive list of locations of original Modigliani art works …" (Apr, 2010)*
- *FactForge* was the first system to pass the Modigliani test
  - Using data from 3 different datasets
  - Neither DBPedia, not Freebase alone can pass the test

ontotext

GATE

Passing the test with FactForge

```
PREFIX fb: <http://rdf.freebase.com/ns/>
PREFIX dbpedia: <http://dbpedia.org/resource/>
PREFIX dbp-prop: <http://dbpedia.org/property/>
PREFIX dbp-ont: <http://dbpedia.org/ontology/>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
PREFIX ff: <http://factforge.net/>

SELECT DISTINCT ?painting_l ?owner_l ?city_fb_con ?city_db_loc ?city_db_cit
WHERE {
     ?painting fb:visual_art.artwork.artist dbpedia:Amedeo_Modigliani ;
               fb:visual_art.artwork.owners [fb:visual_art.artwork_owner_relationship.owner ?ow];
               ff:preferredLabel ?painting_l .
     ?ow ff:preferredLabel ?owner_l .
     OPTIONAL { ?ow fb:location.location.containedby [ff:preferredLabel ?city_fb_con] } .
     OPTIONAL { ?ow dbp-ont:city [ff:preferredLabel ?city_db_cit] } .
     OPTIONAL { ?ow dbp-prop:location ?loc .
                ?loc rdf:type umbel-sc:City ;
                     ff:preferredLabel ?city_db_loc }
}
```

Passing the test with FactForge (2)

FactForge

RDF Search and Explore | SPARQL Query | Relfinder | About | Contact

SPARQL Query

Results for PREFIX fb: <http://rdf.fr… (8)

View as Exhibit Download in JSON | SPARQL Results in XML | SPARQL Results in JSON

painting_l	owner_l	city_fb_con	city_db_loc	city_db_cit
Head@en	Museum of Modern Art	New York City		
Anna Zborowska@en	Museum of Modern Art	New York City		
Portrait of Diego Rivera@en	The São Paulo Museum of Art@en		São Paulo	
Woman with a Necklace@en	School of the Art Institute of Chicago@en			Chicago
Portrait of a Woman@en	School of the Art Institute of Chicago@en			Chicago
Reclining Nude@en	Museum of Modern Art	New York City		
Madam Pompadour@en	School of the Art Institute of Chicago@en			Chicago
Jeanne Hébuterne@en	Barnes Foundation@en	Philadelphia