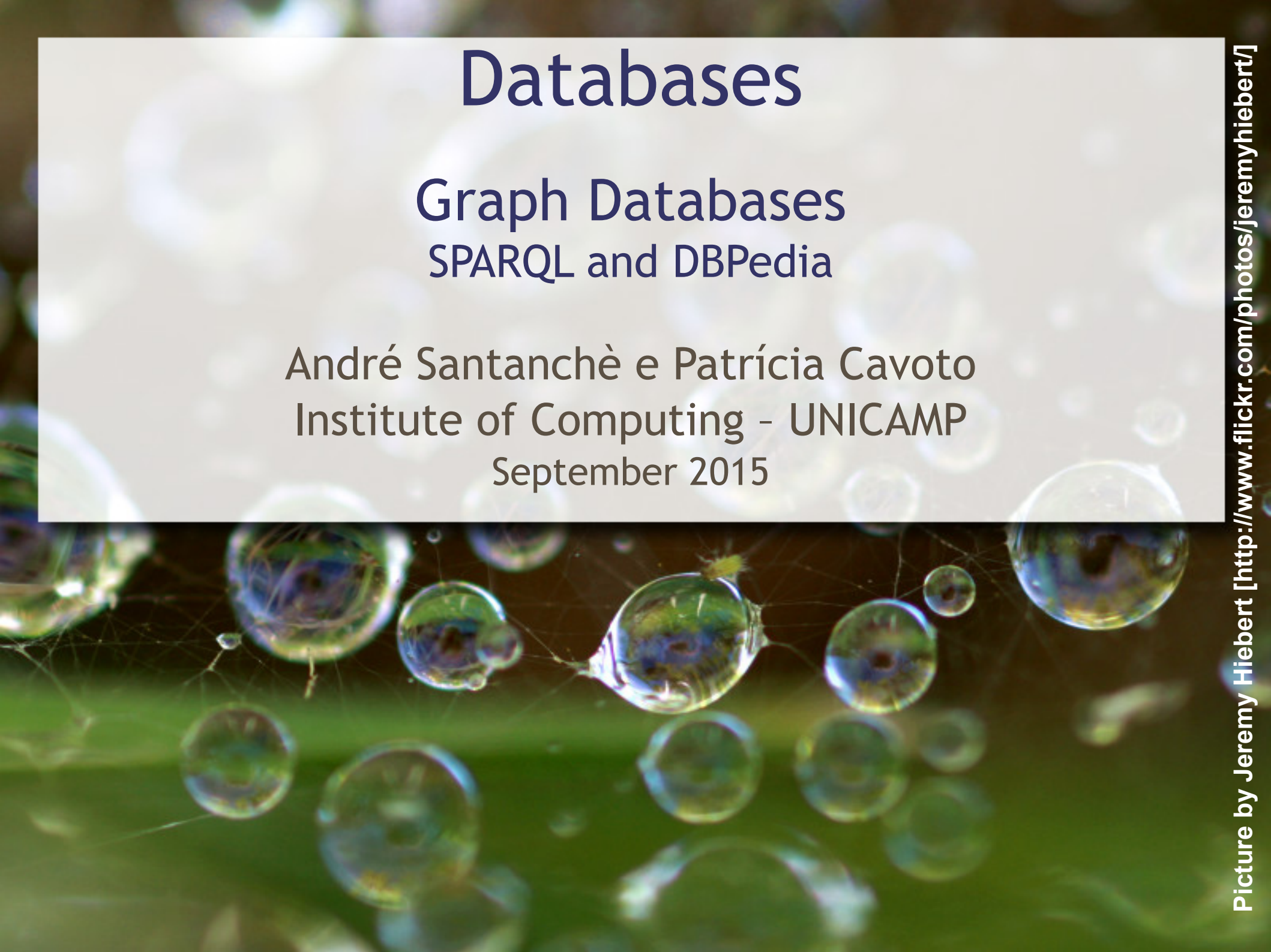
Databases

Graph Databases

SPARQL and DBPedia

André Santanchè e Patrícia Cavoto
Institute of Computing – UNICAMP

September 2015

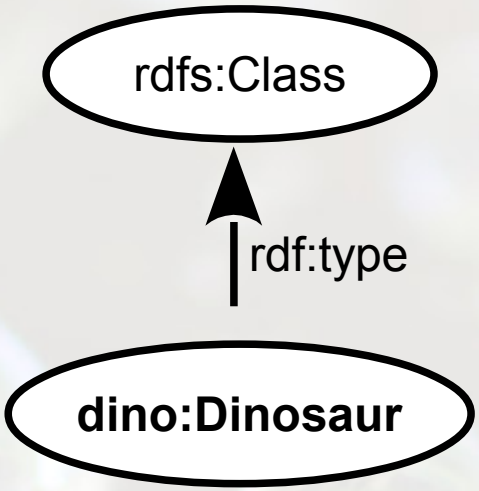


Picture by Jeremy Hiebert [http://www.flickr.com/photos/jeremyhiebert/]

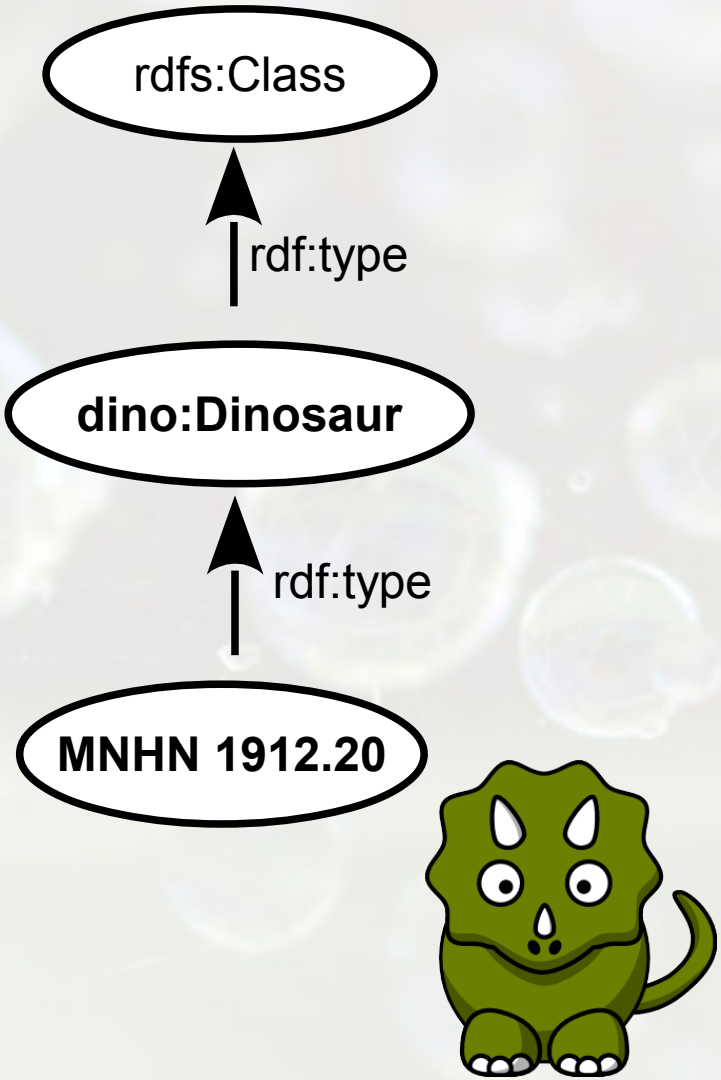
Dinosaur Example

Namespace Dino
dino: <http://purl.org/dinos/>

Dinosaur Class

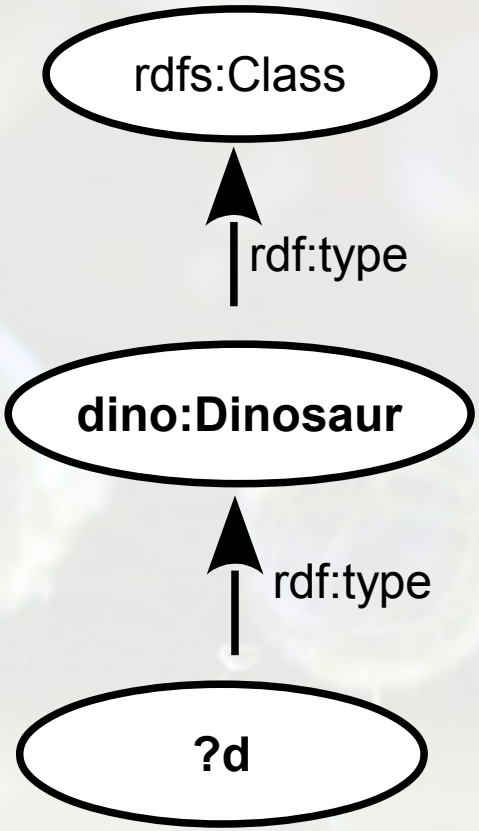


Dinosaur Instance



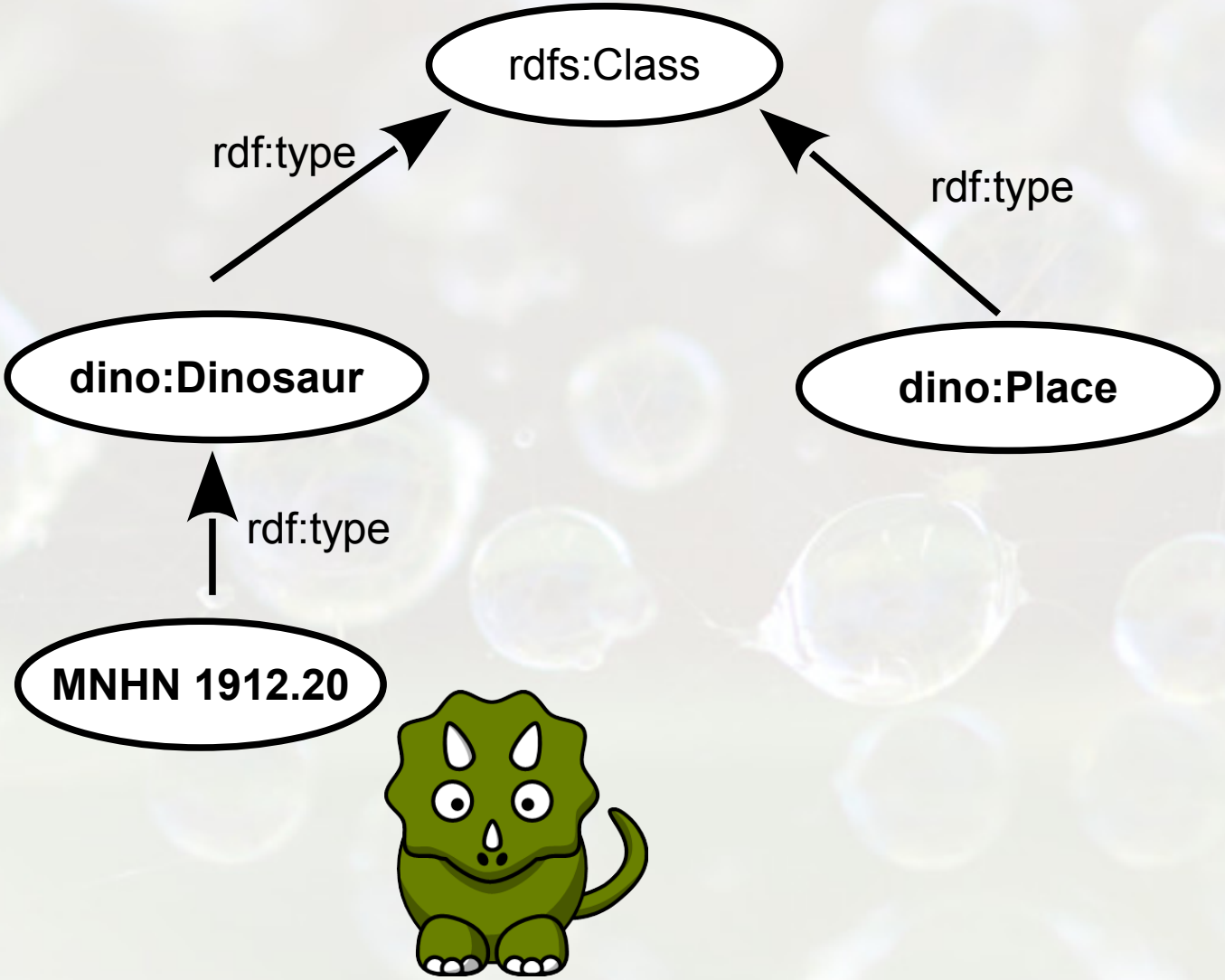
SPARQL
Query Language for RDF

Instances of the Dinosaur class (SPARQL)

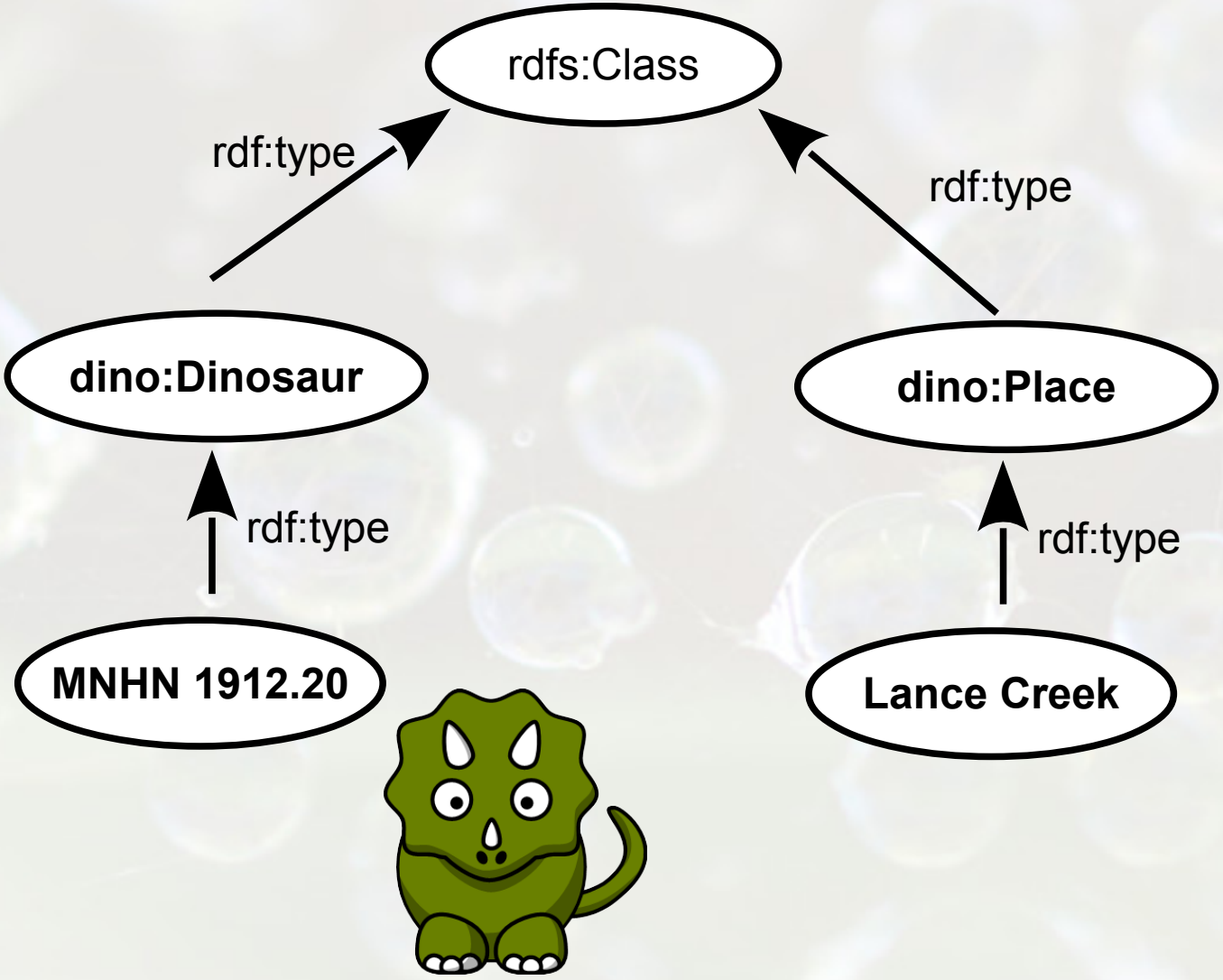



```
SELECT ?d
WHERE { ?d rdf:type dino:Dinosaur }
```

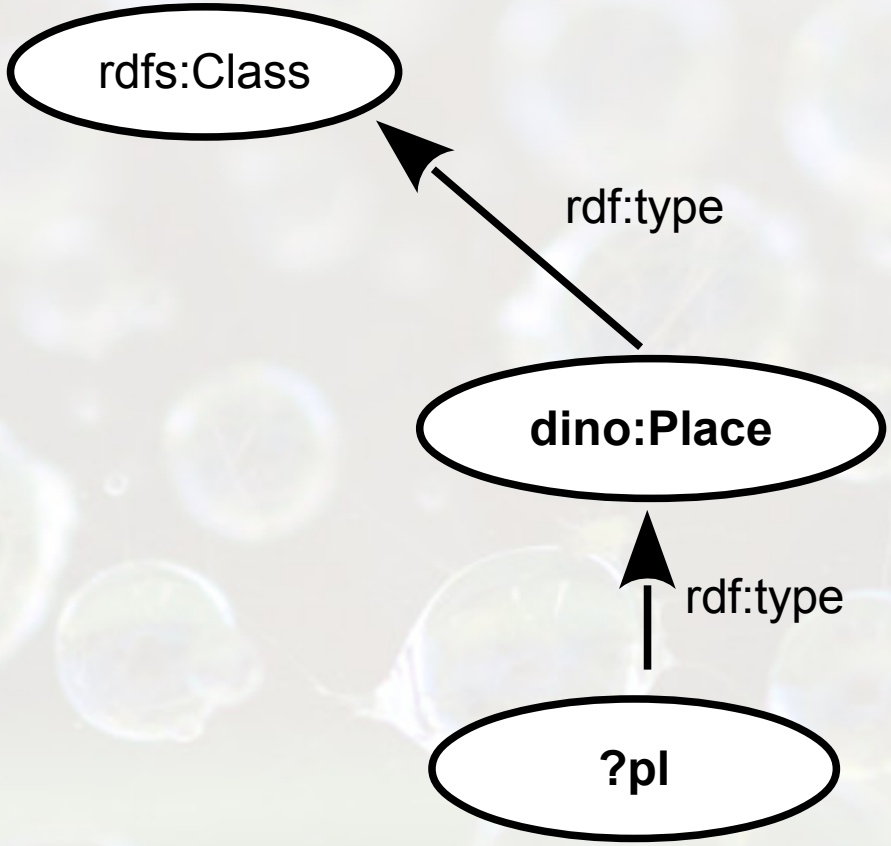
Place Class


rdfs:Class
rdf:type
rdf:type
dino:Dinosaur
dino:Place
rdf:type
MNHN 1912.20

Place Instance

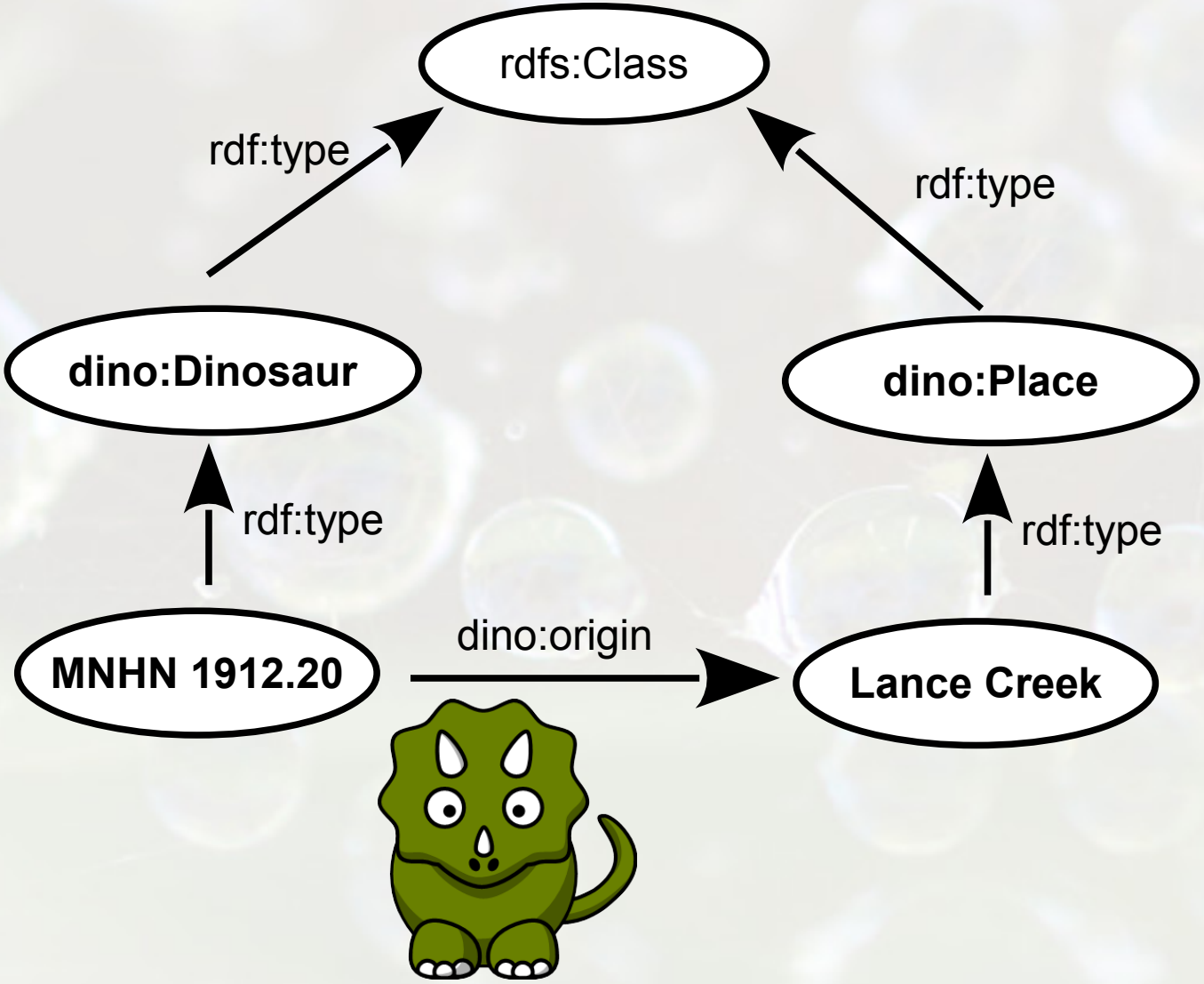


Instances of the Place class (SPARQL)

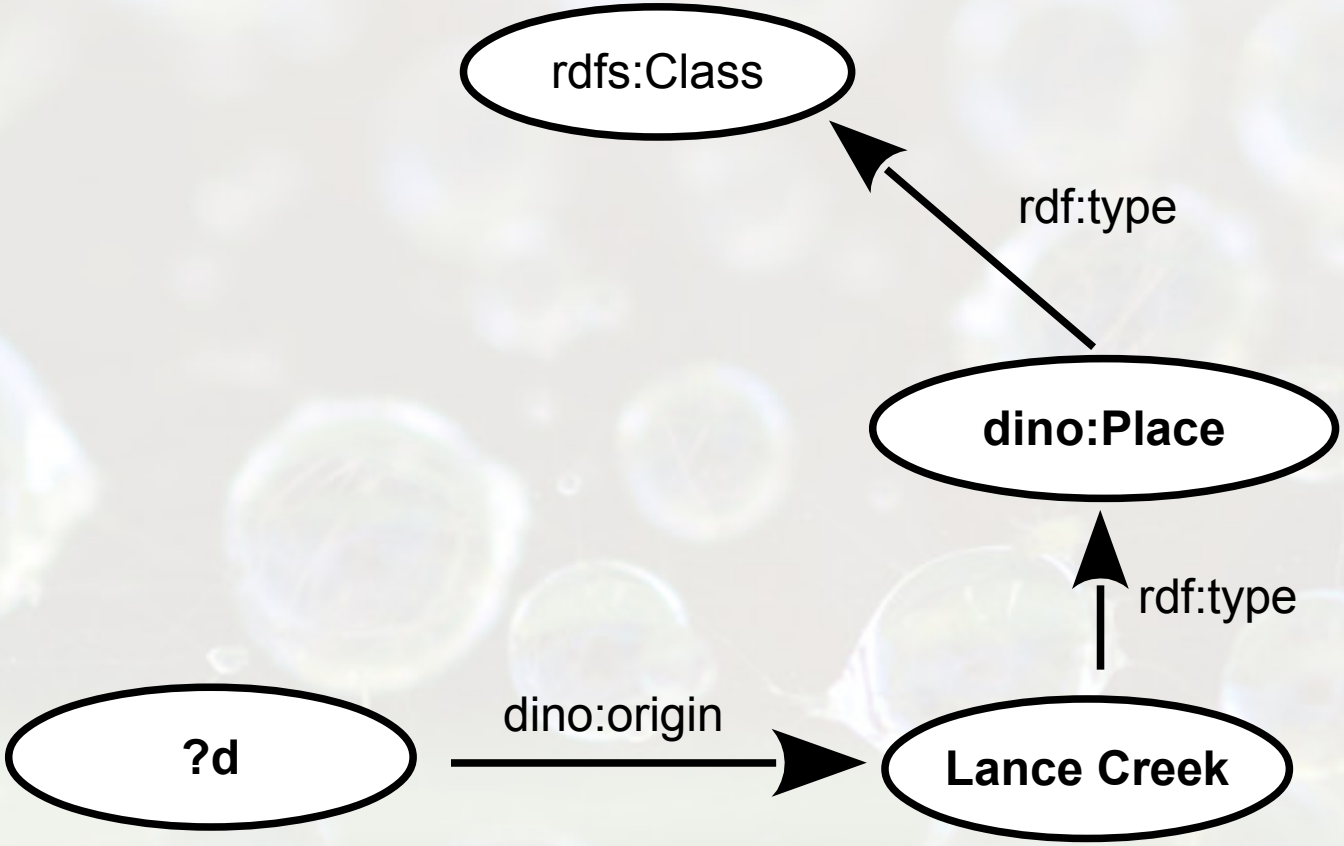



```
SELECT ?pl
WHERE { ?pl rdf:type dino:Place }
```

Place Instance

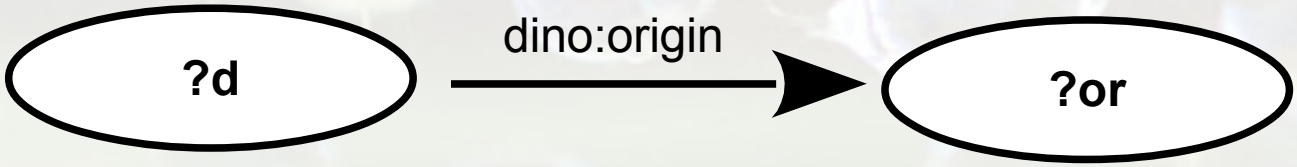


Origin: Lance Creek



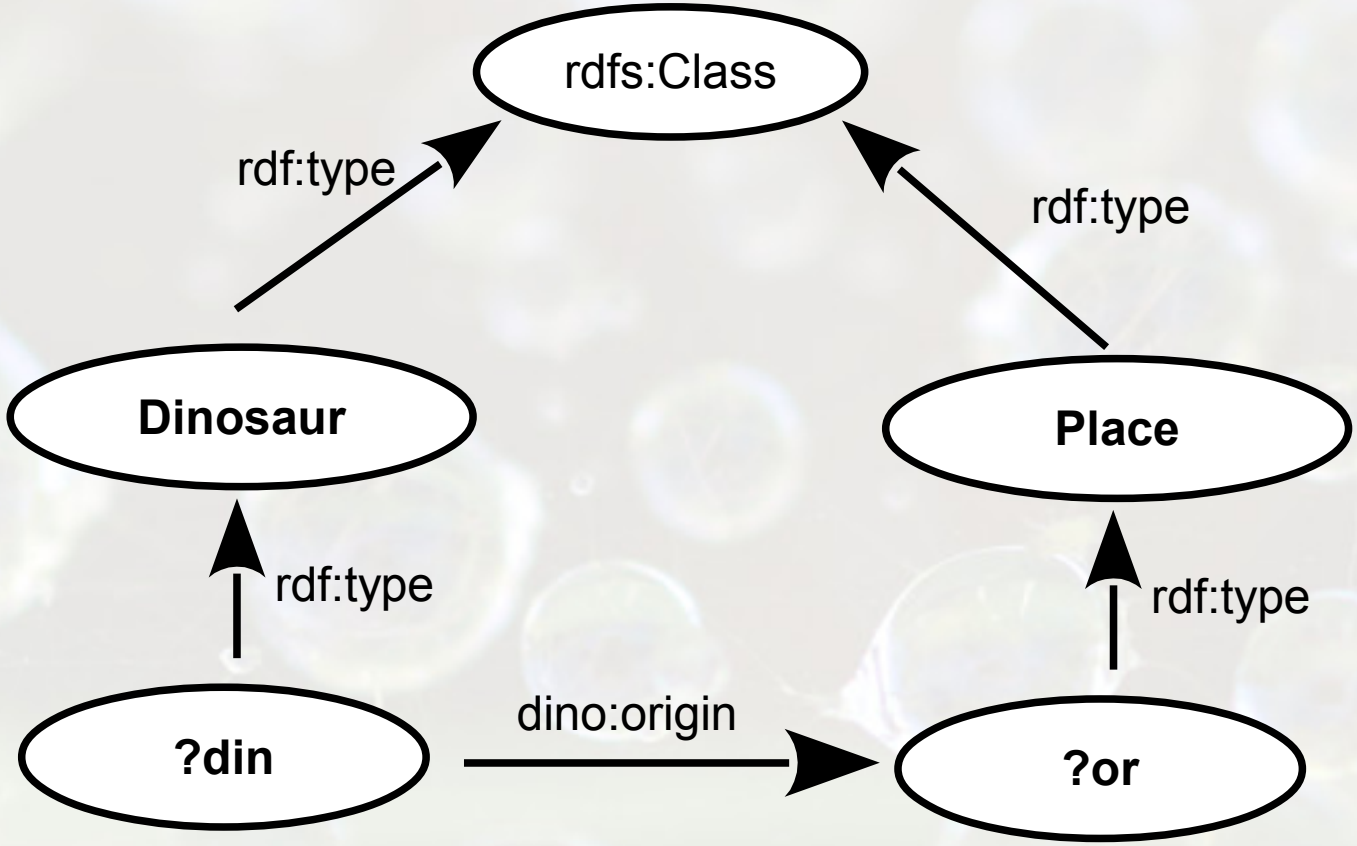

```
SELECT ?d
WHERE { ?d dino:origin dino:Lance_Creek }
```

Resources and their origins



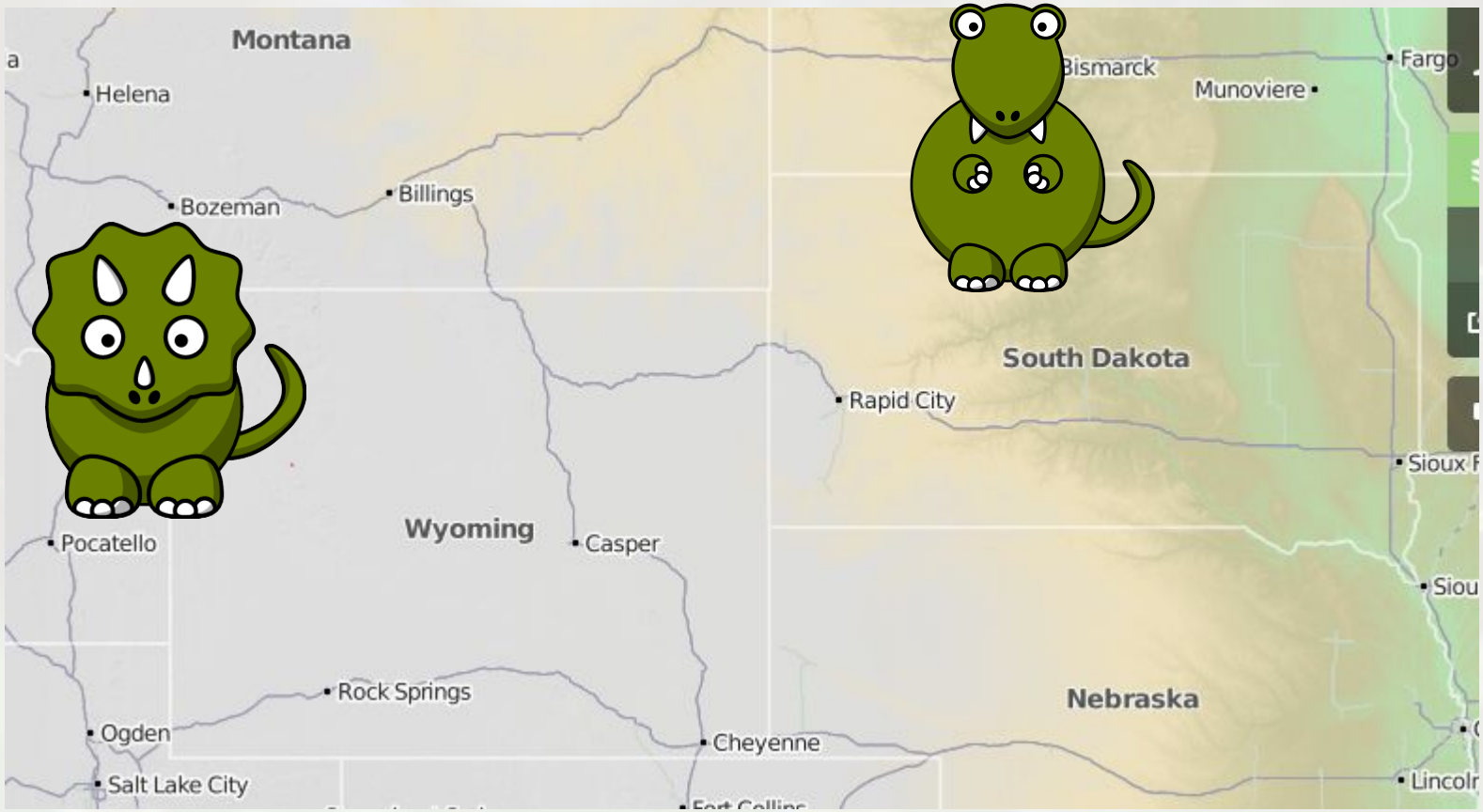

```
SELECT ?d ?or
WHERE { ?d dino:origin ?or }
```

Dinosaurs and their origins



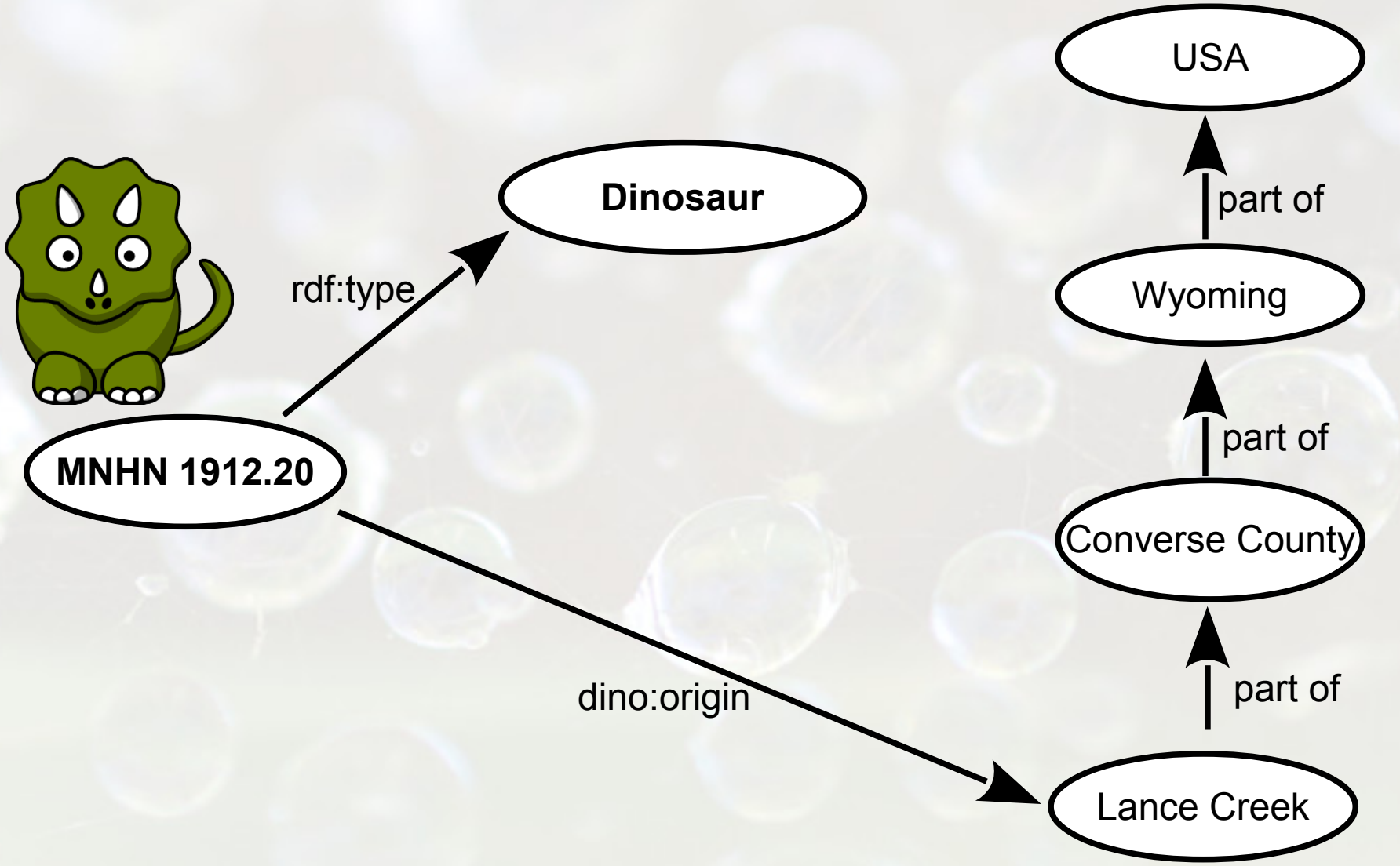

```
SELECT ?din ?or
WHERE { ?din rdf:type dino:Dinosaur .
        ?din dino:origin ?or }
```

Analyzing in the Space

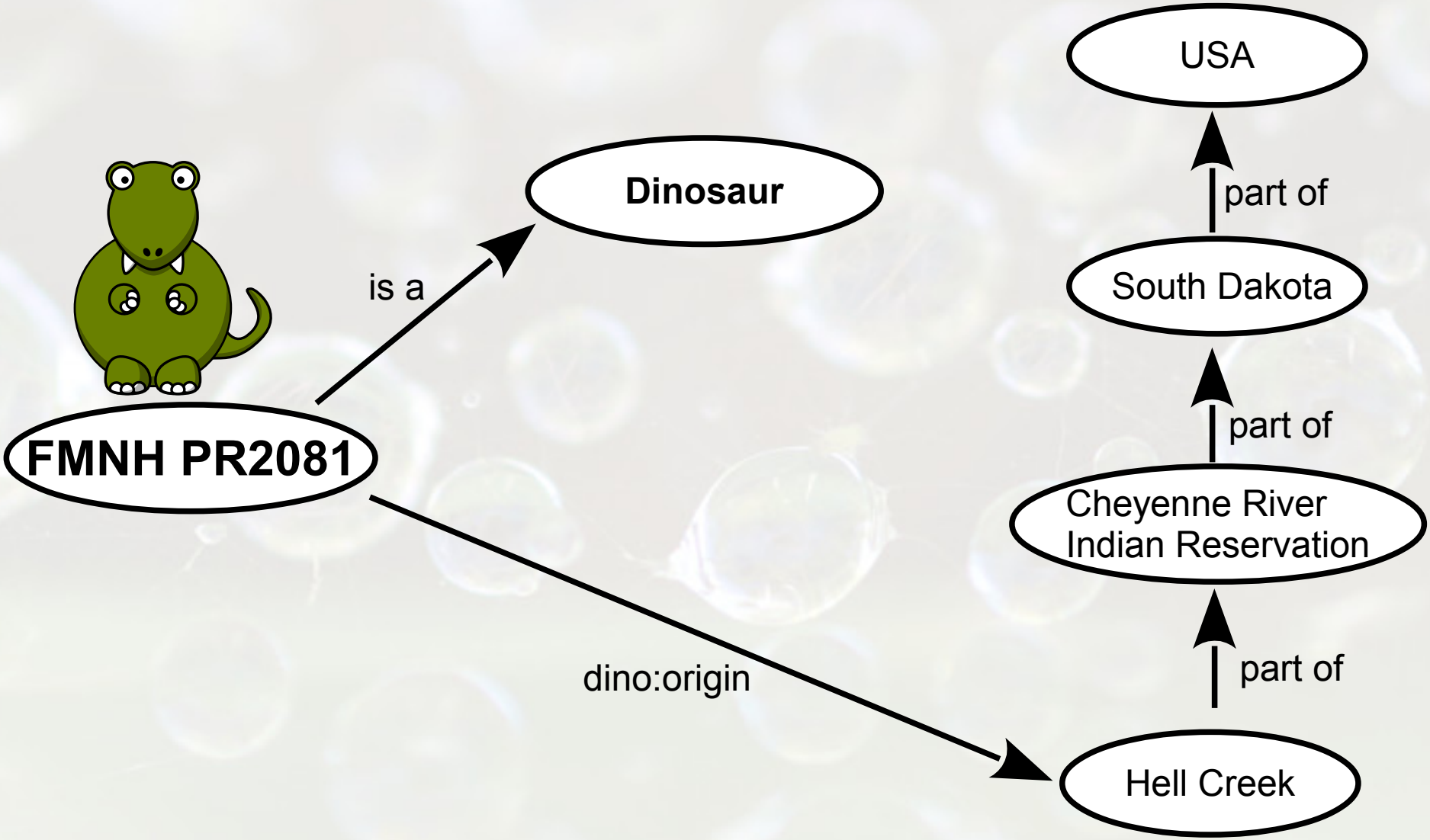


OpenStreetMap

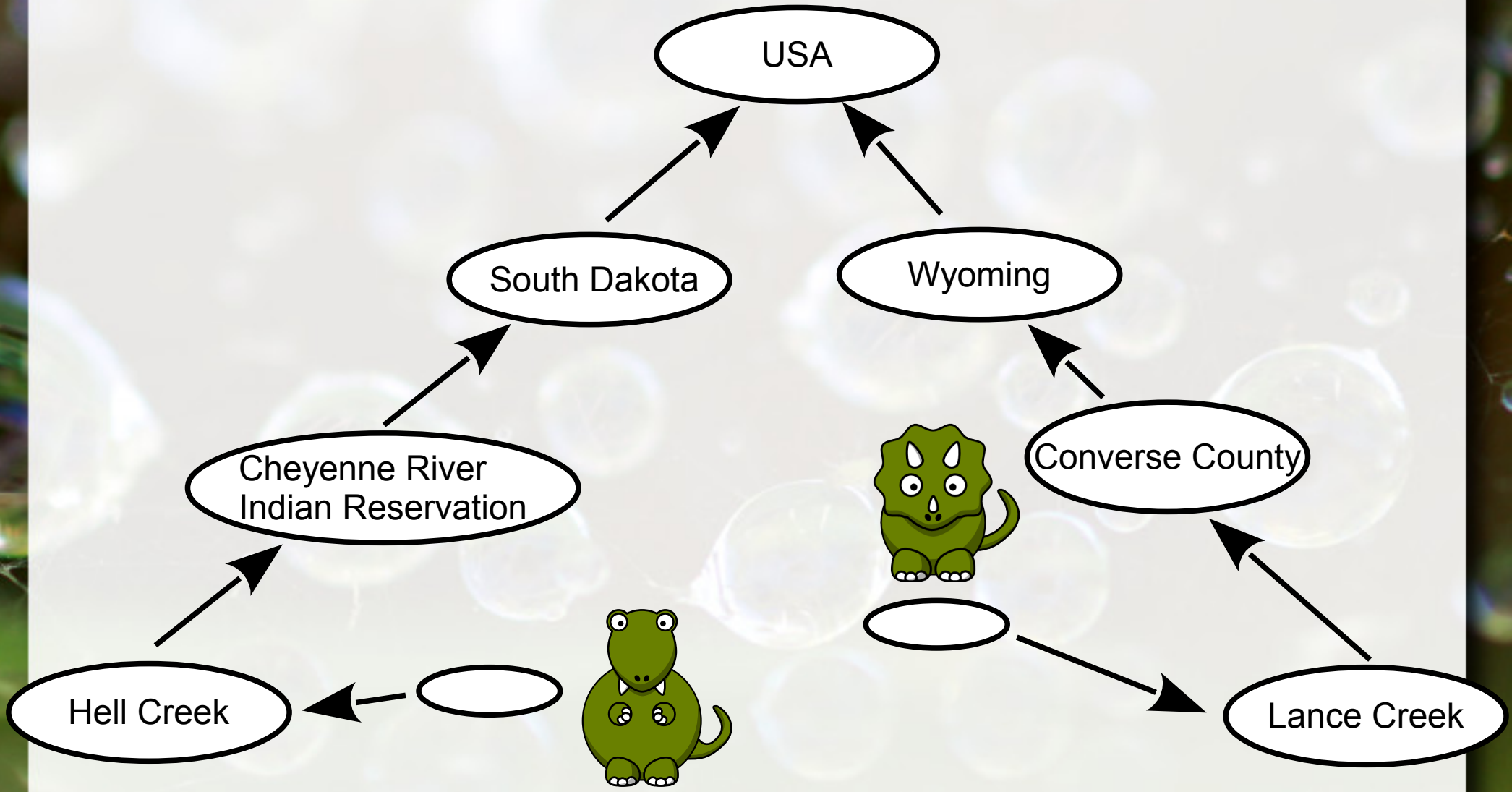
Triceratops in a Graph



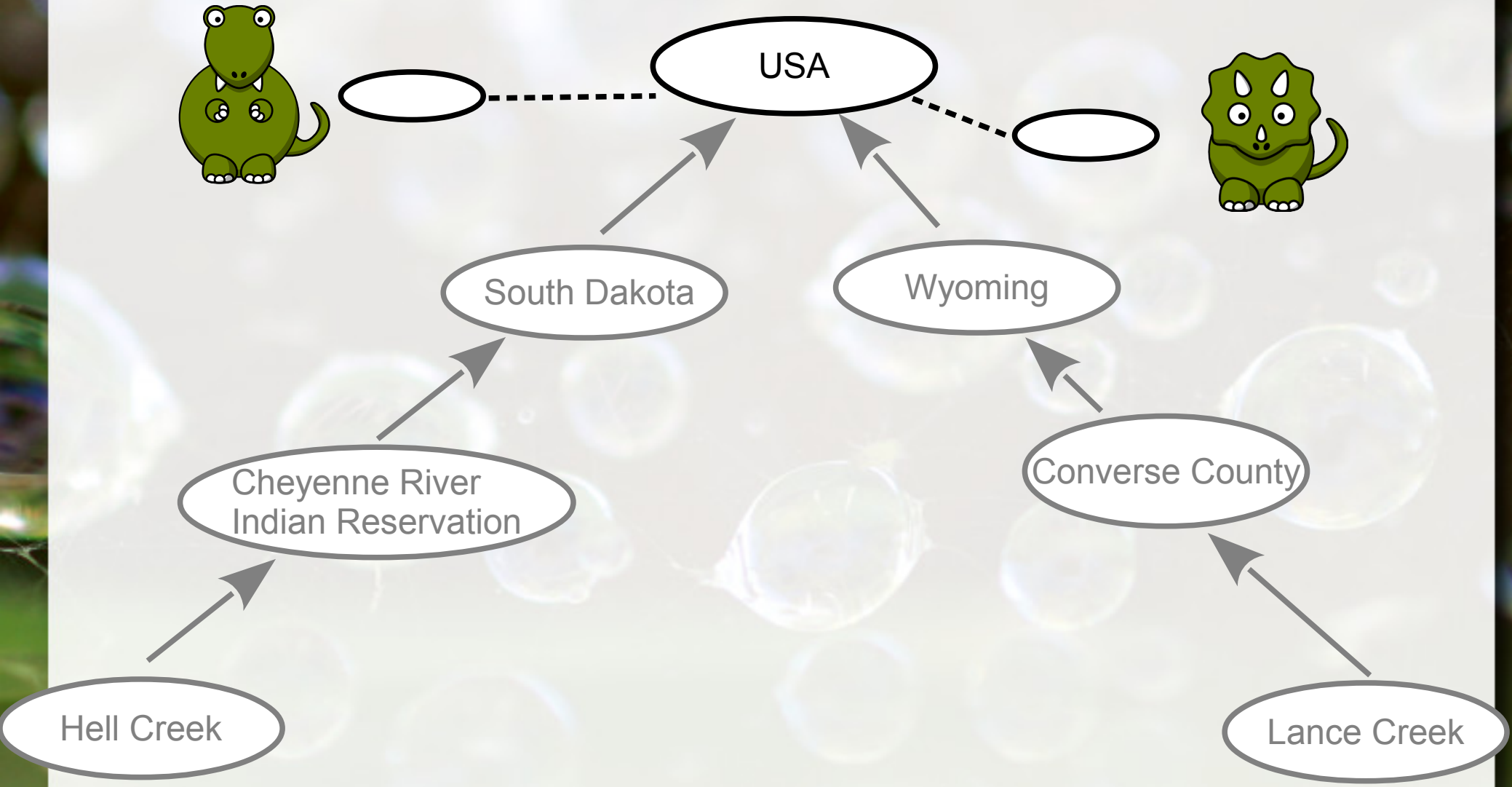
Tyrannosaurus in a Graph



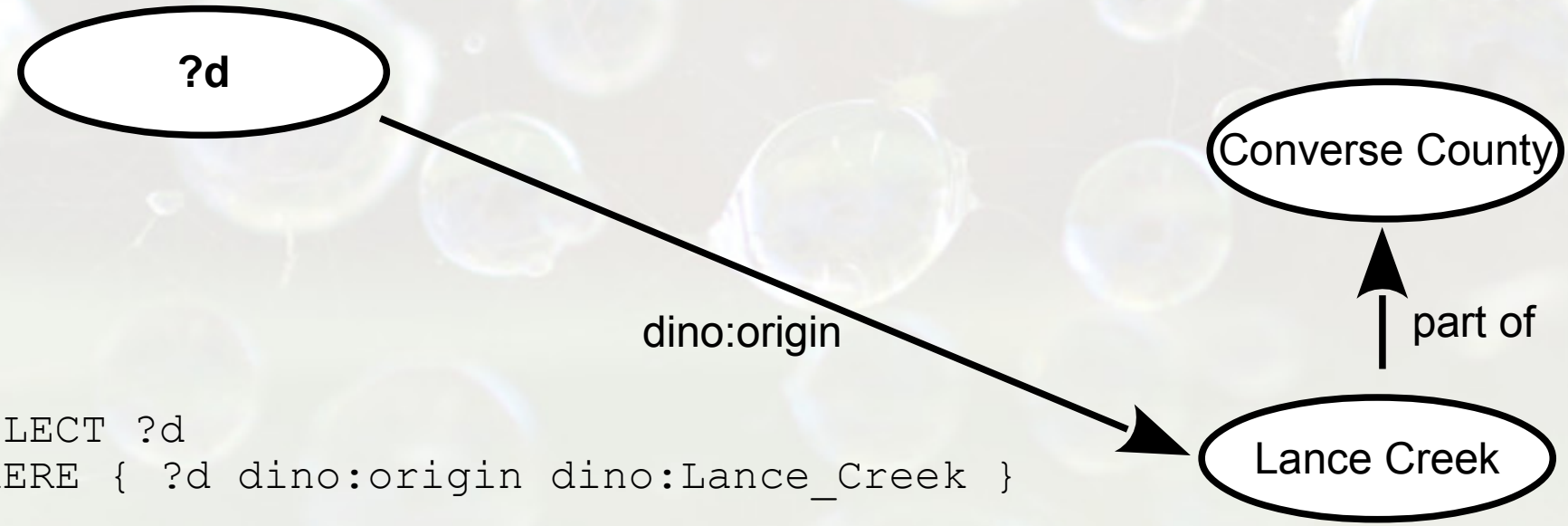
Analyzing in the Space



Analyzing in the Space

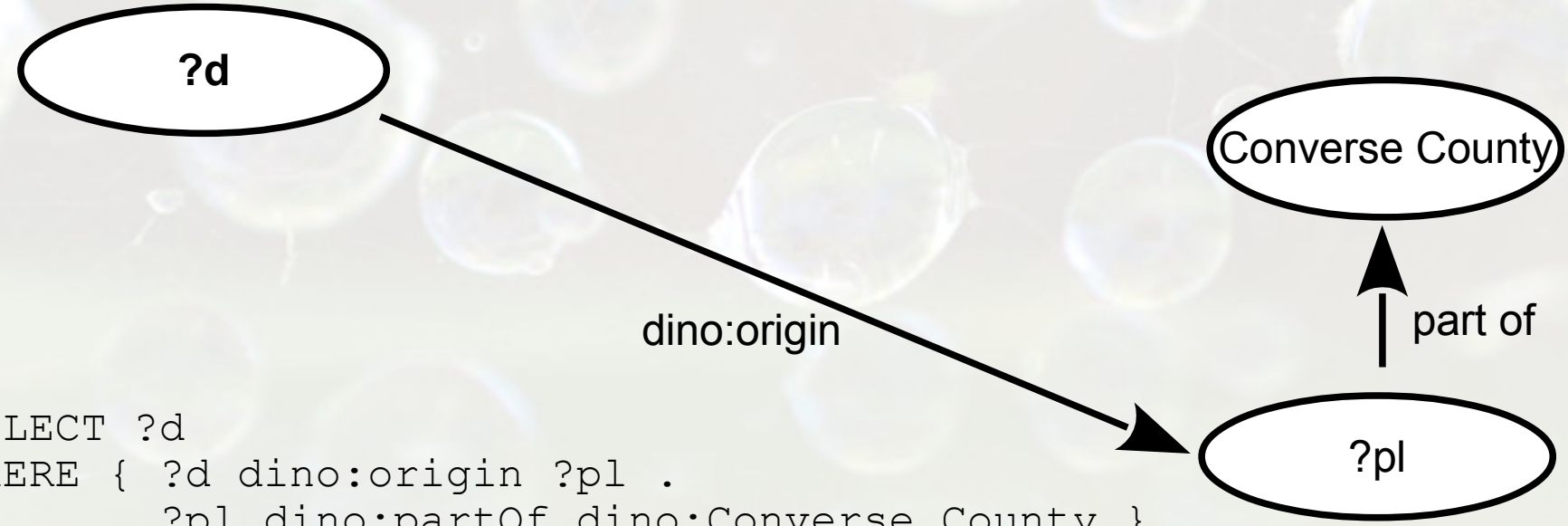


Origin: Lance Creek



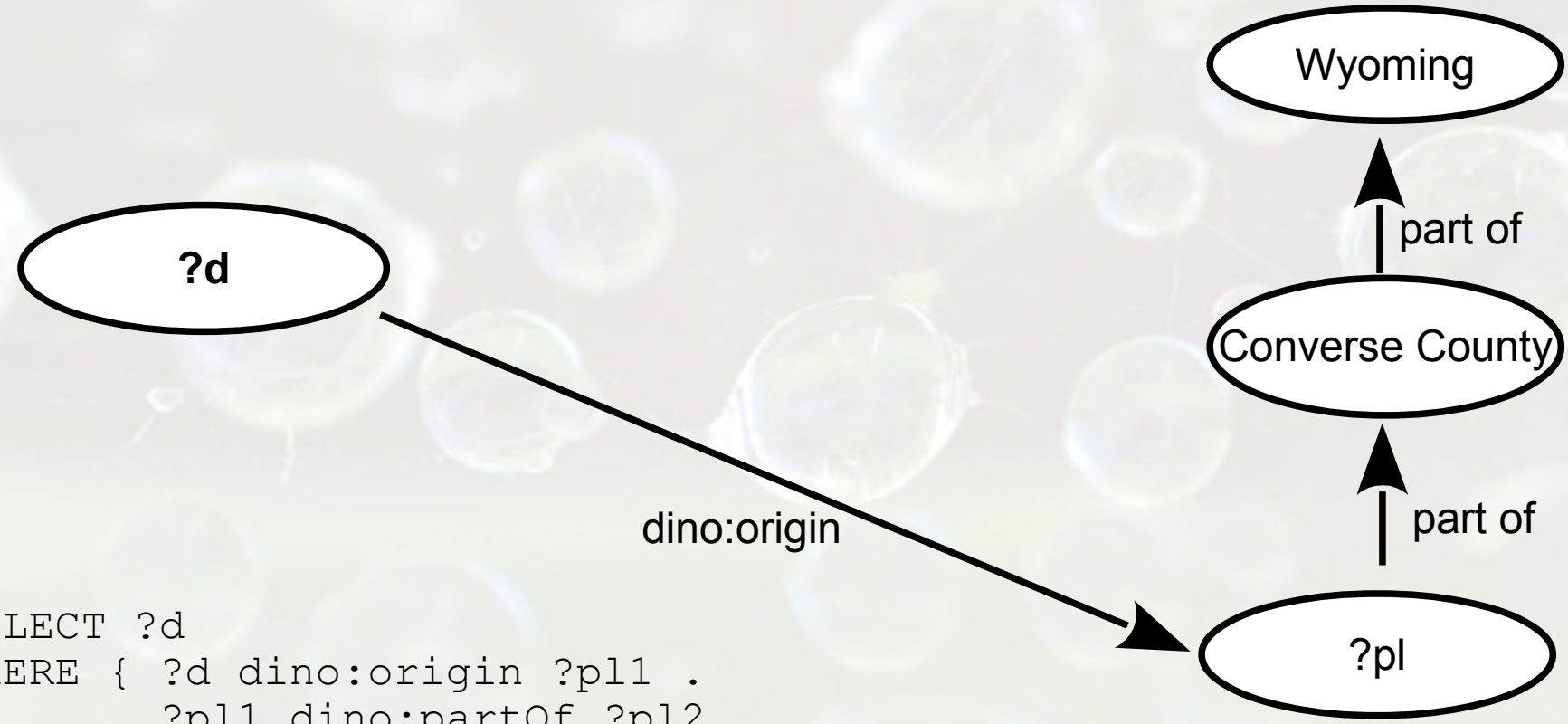

```
SELECT ?d
WHERE { ?d dino:origin dino:Lance_Creek }
```

Origin: part of Converse County



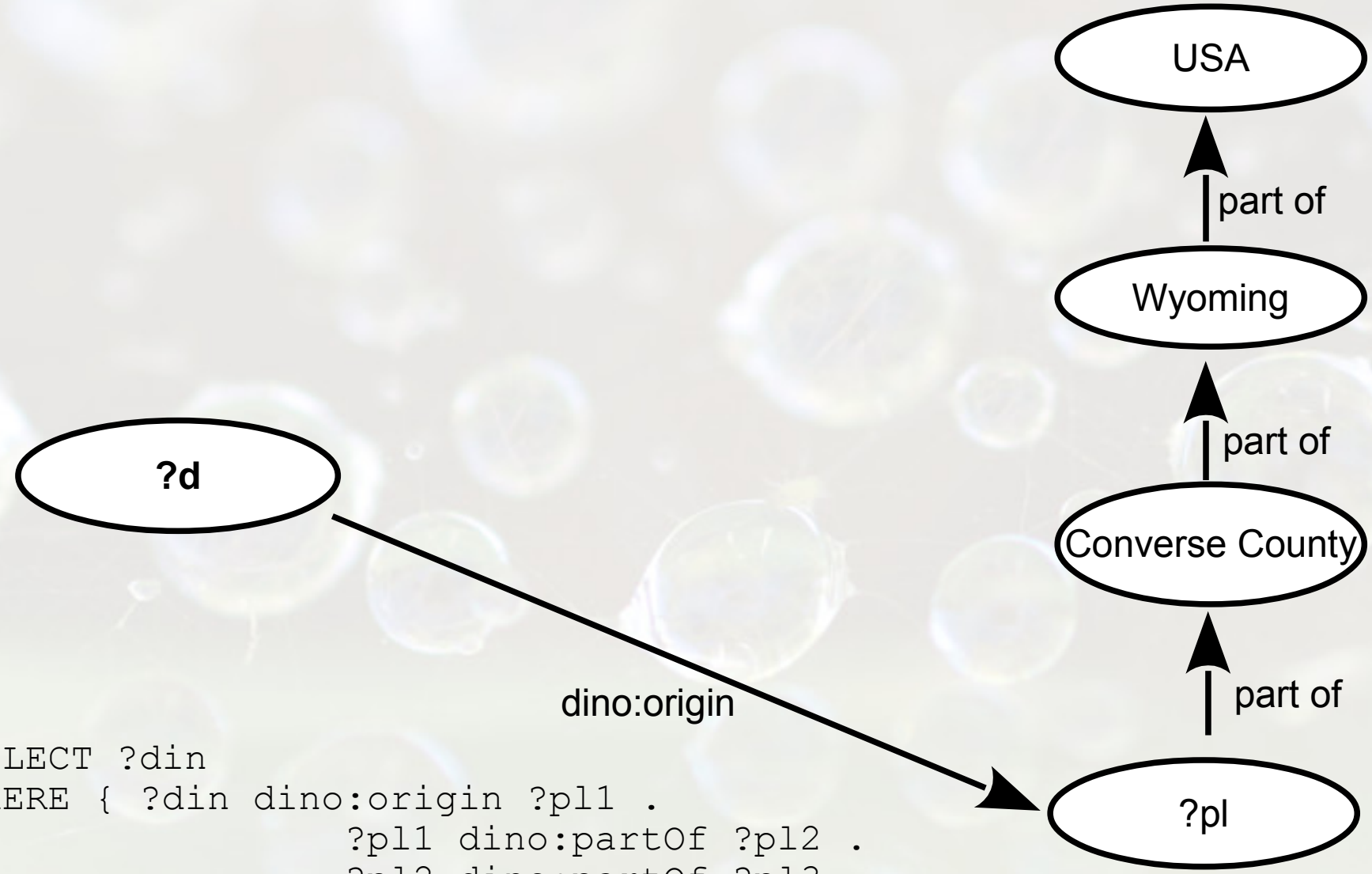

```
SELECT ?d
WHERE { ?d dino:origin ?pl .
        ?pl dino:partOf dino:Converse_County }
```

Origin: part of Wyoming



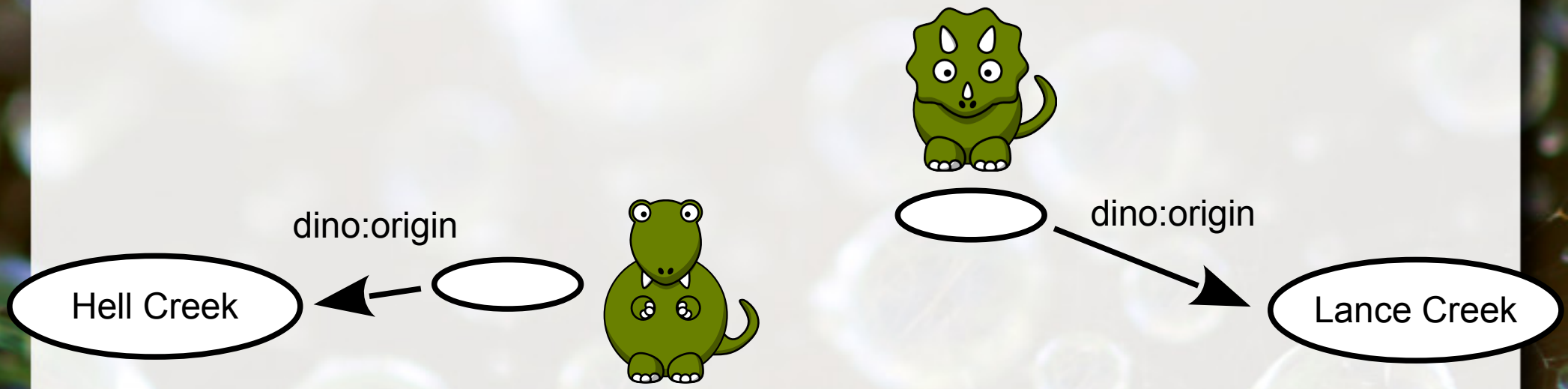

```
SELECT ?d
WHERE { ?d dino:origin ?pl1 .
        ?pl1 dino:partOf ?pl2 .
        ?pl2 dino:partOf dino:Wyoming }
```

Origin: part of USA



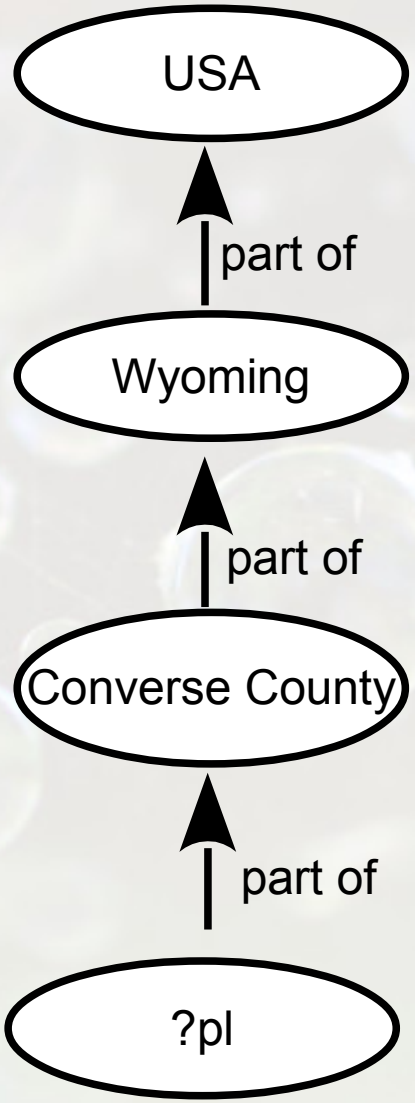

```
SELECT ?din
WHERE { ?din dino:origin ?pl1 .
                ?pl1 dino:partOf ?pl2 .
                ?pl2 dino:partOf ?pl3 .
                ?pl3 dino:partOf dino:USA }
```

Origin: Lance Creek or Hell Creek




```
SELECT ?d
WHERE {
        { ?d dino:origin dino:Lance_Creek }
                UNION
        { ?d dino:origin dino:Hell_Creek }
              }
```

Resources with the part of relationship



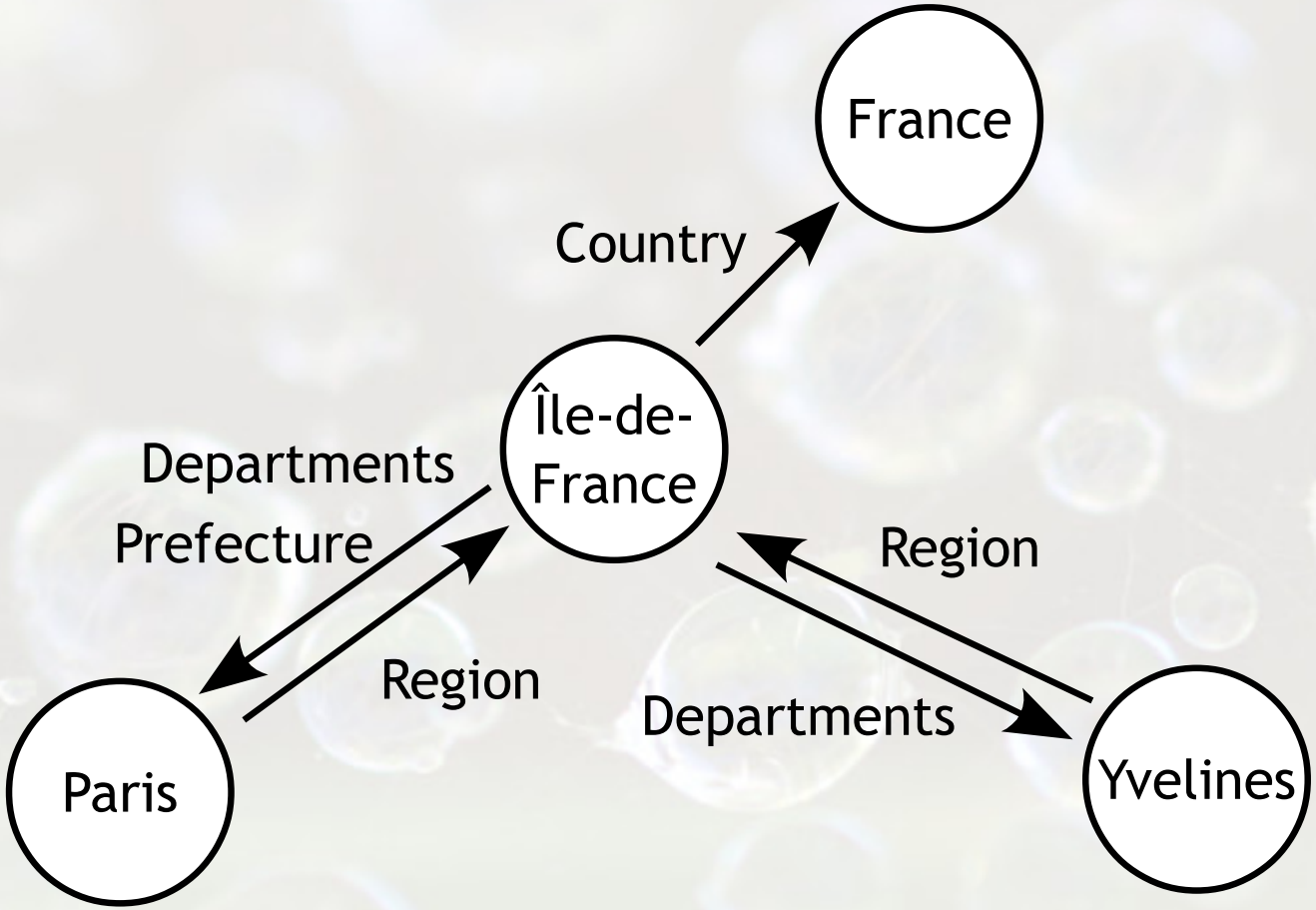

```
SELECT ?l1 ?l2
   WHERE { ?l1 dino:partOf ?l2 }
```

DBPedia

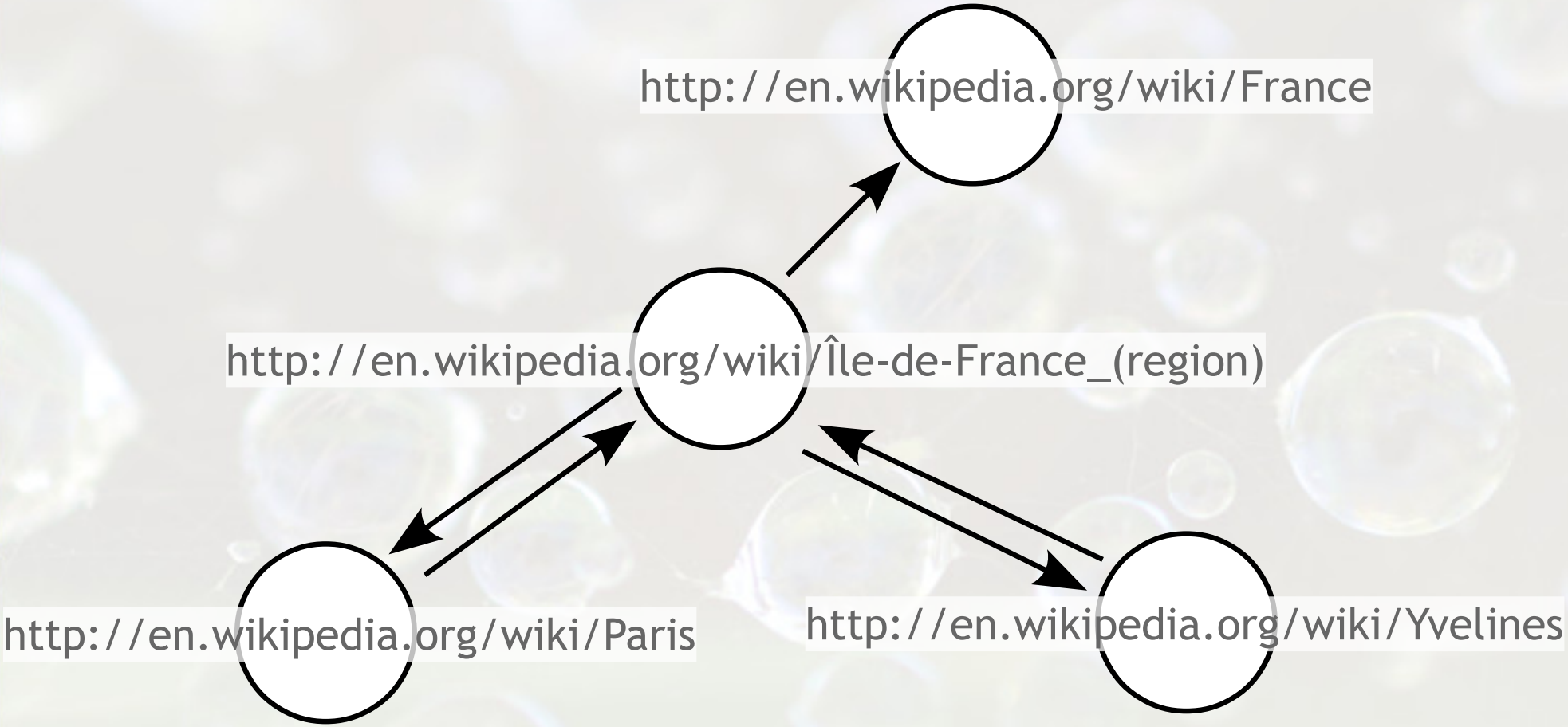
Wikipedia

Infobox

DBPedia


France
Country
Île-de-France
Departments
Prefecture
Region
Region
Departments
Paris
Yvelines

DBPedia (URIs)



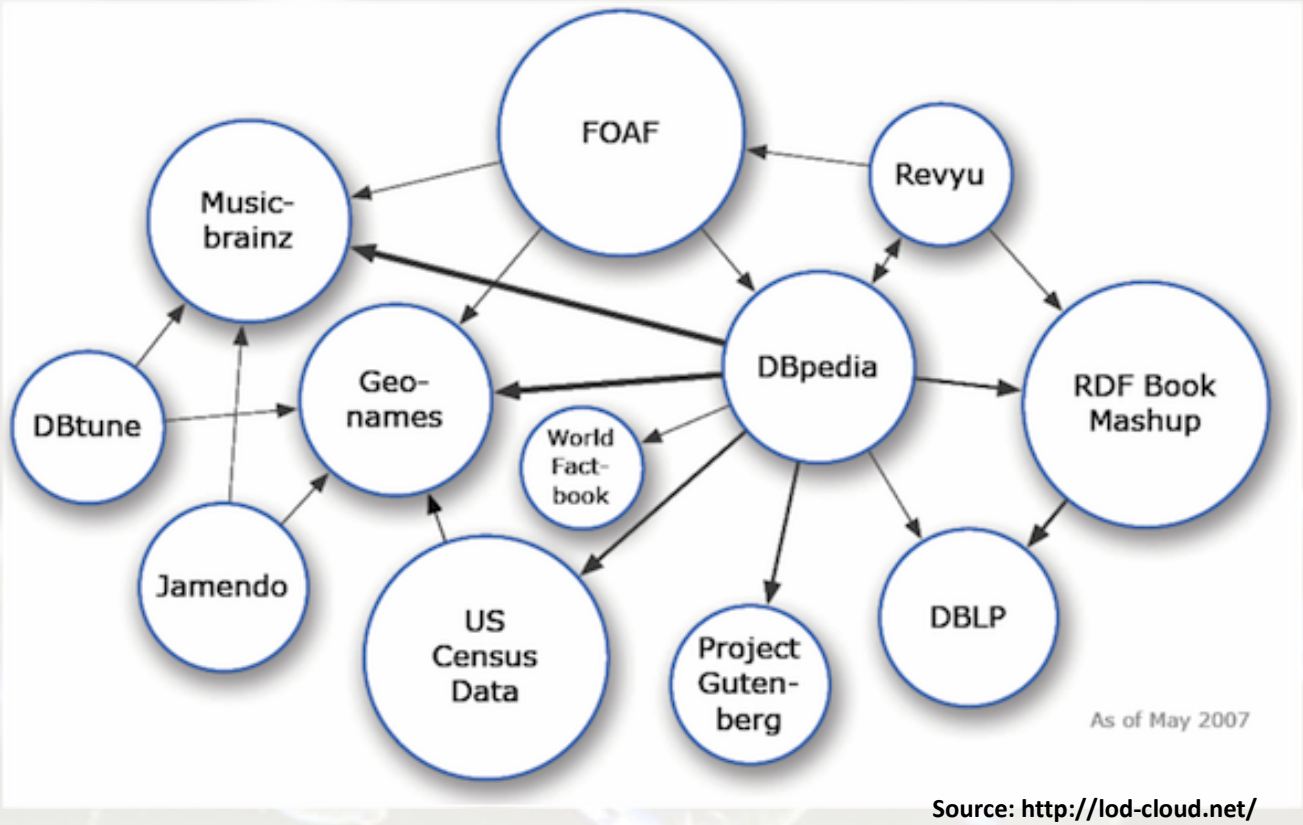
DBPedia – English

- **4 million things**
- 3.22 million classified in a consistent ontology
  - 832,000 persons
  - 639,000 places (427,000 populated)
  - 372,000 creative works
    - 116,000 music albums; 78,000 films; 18,500 video games
  - 209,000 organizations
  - 226,000 species
  - 5,600 diseases.

DBPedia – International

- 119 languages
- 24.9 million things
- 16.8 million interlinked with English
- 12.6 million unique things

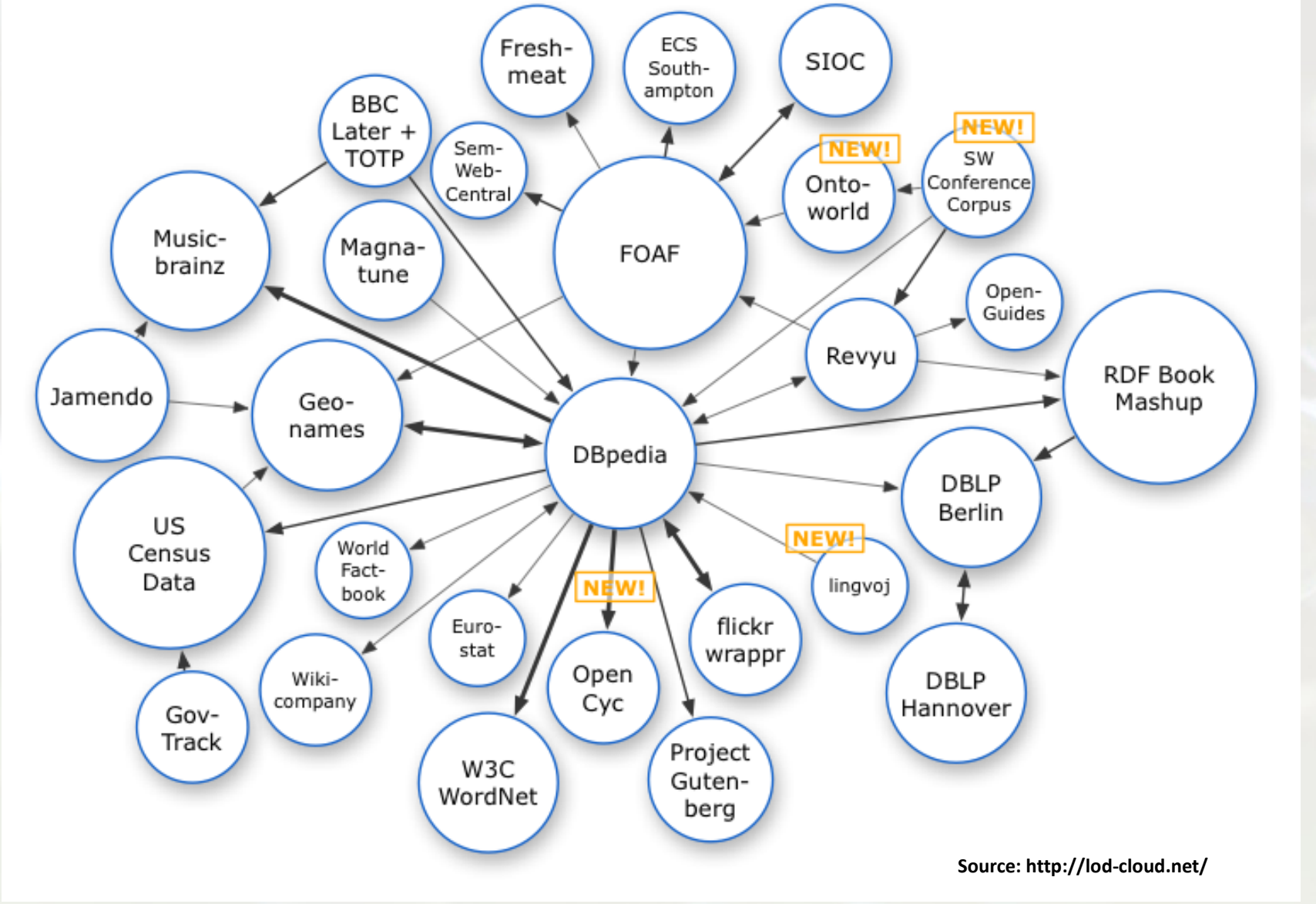
Linked Data



Source: http://lod-cloud.net/

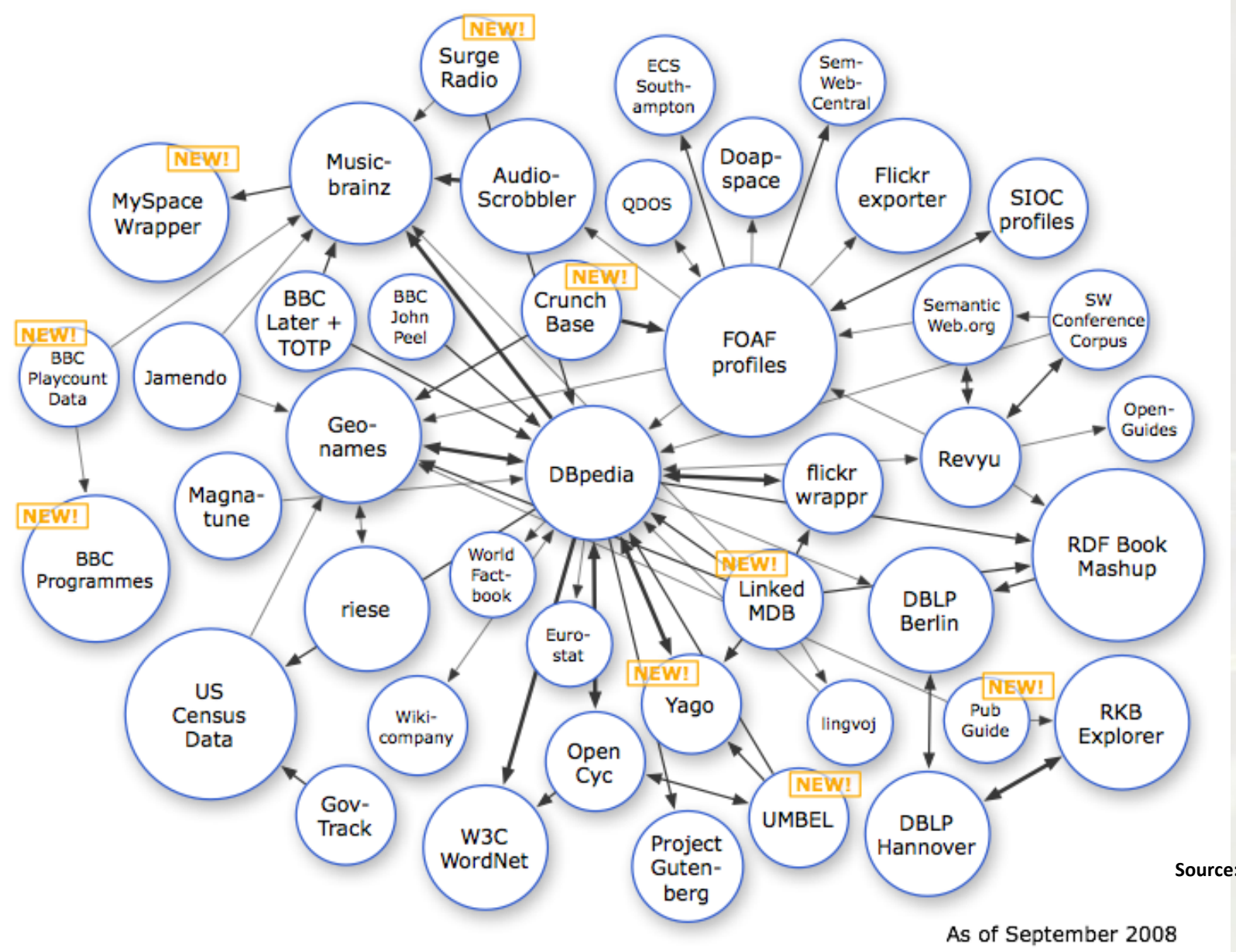
Datasets published following Linked Data 'format': **05/2007**

Linked Data



Datasets published following Linked Data 'format': **11/2007**

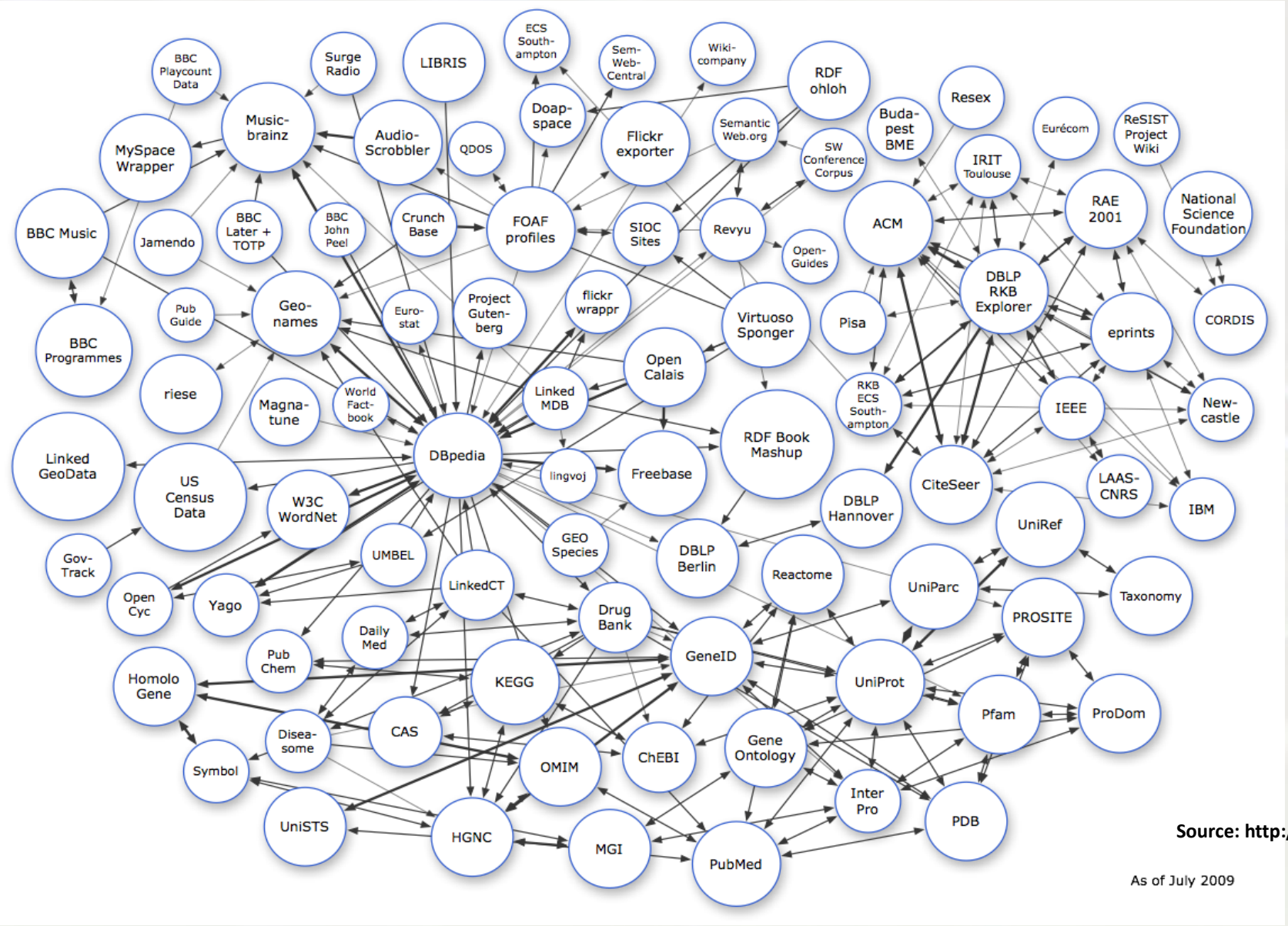
Linked Data



Source: http://lod-cloud.net/

Datasets published following Linked Data 'format': **2008**

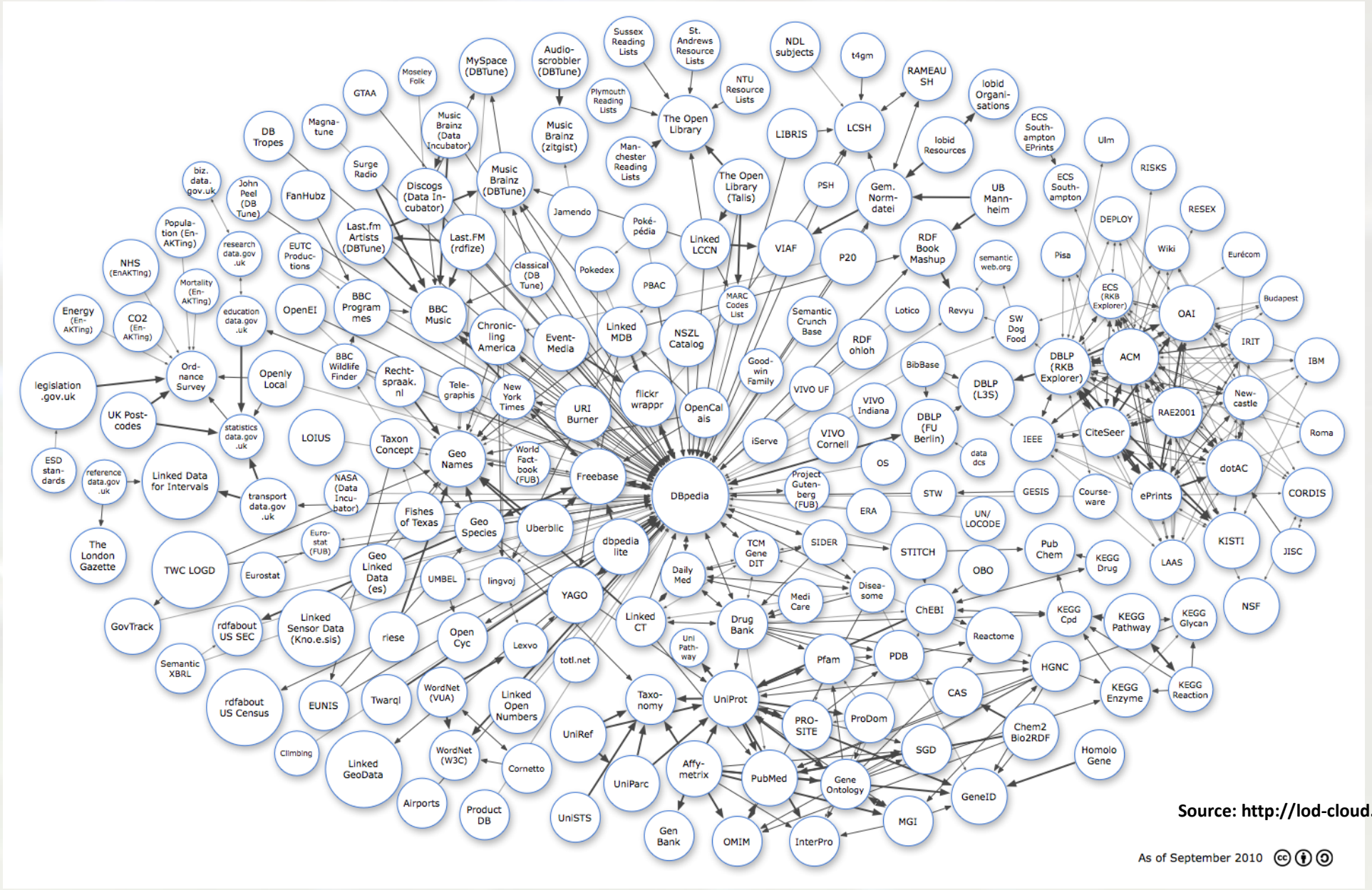
Linked Data



Source: http://lod-cloud.net/

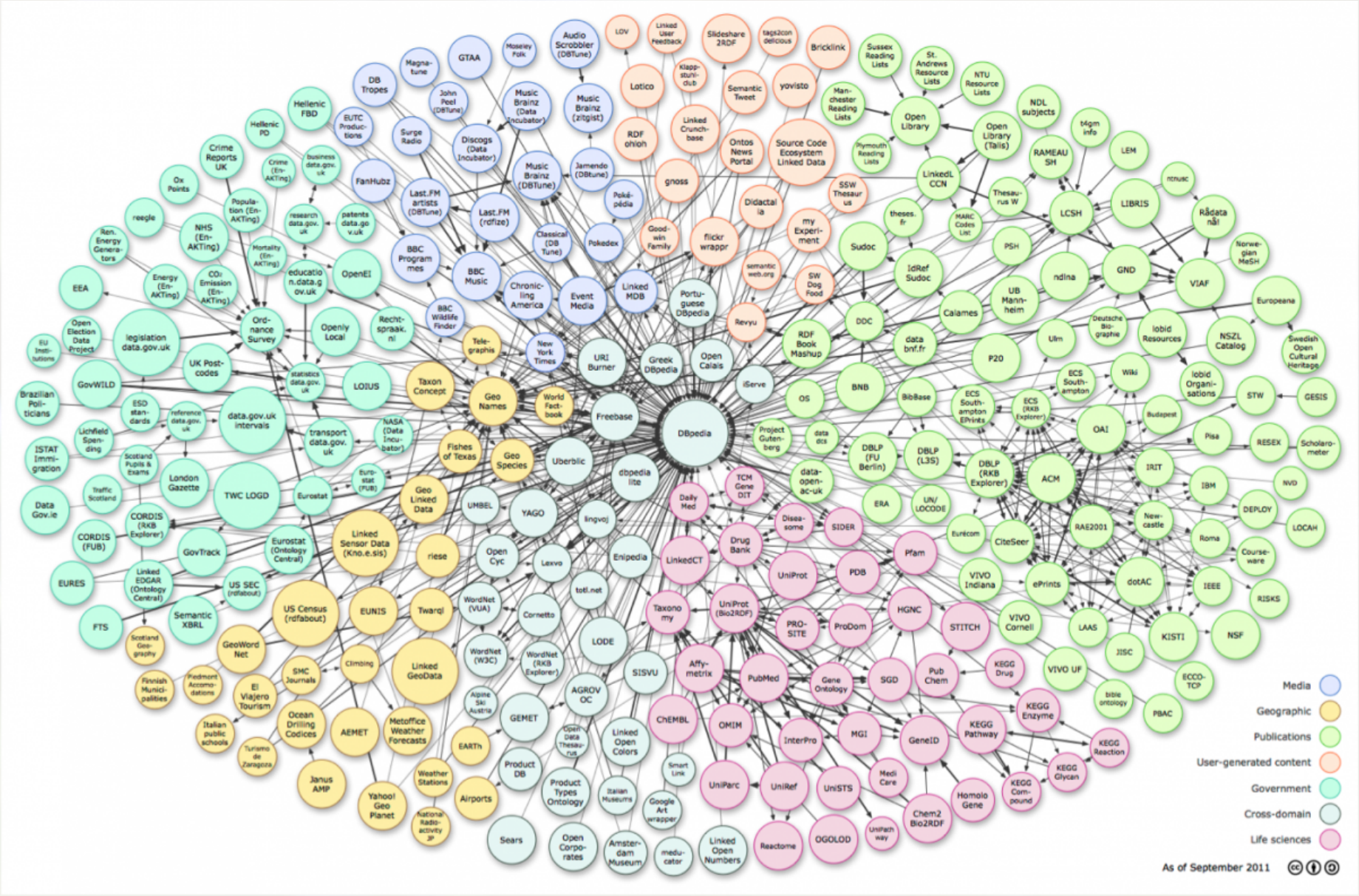
Datasets published following Linked Data ‘format’: **2009**

Linked Data



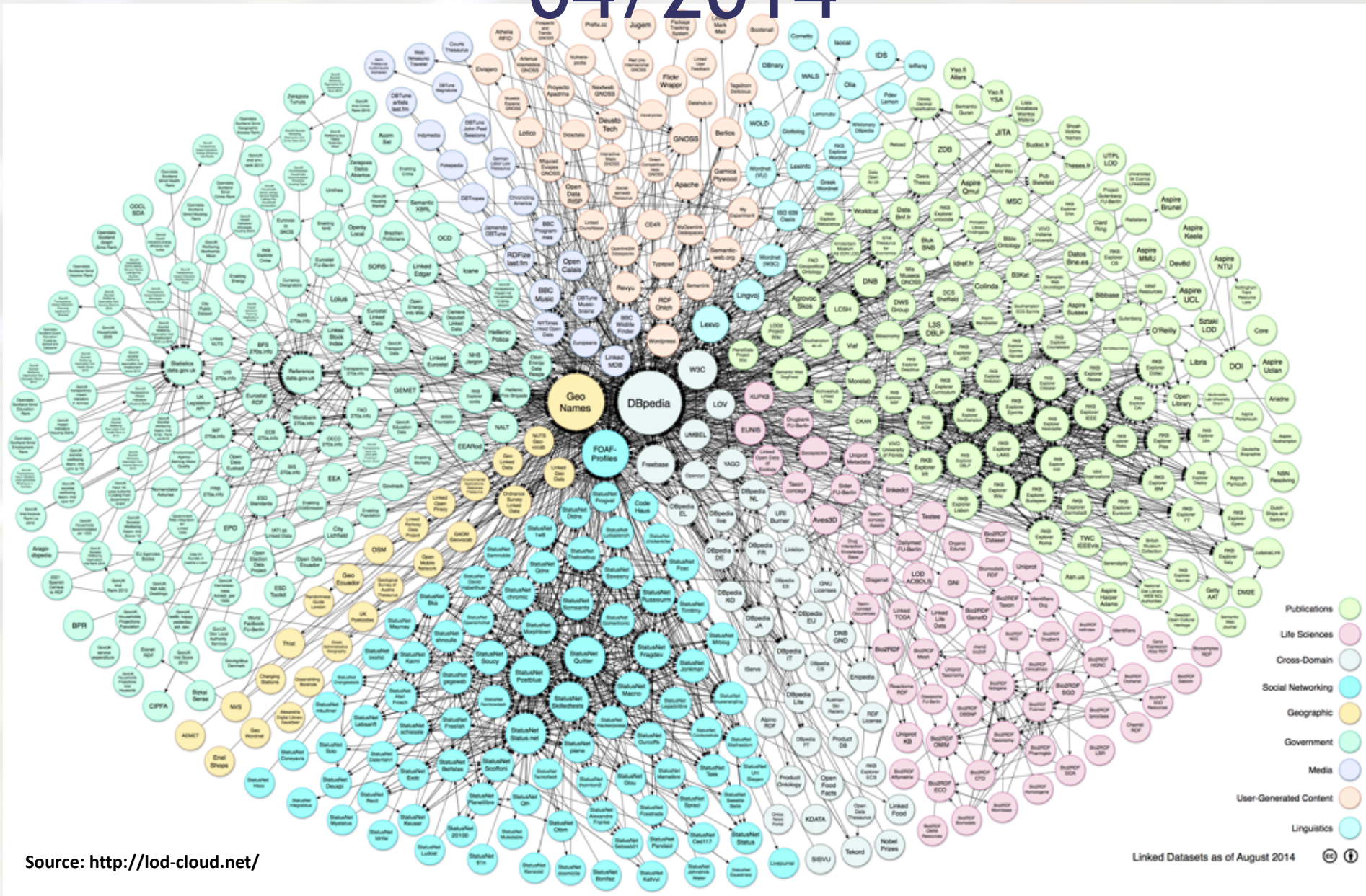
Datasets published following Linked Data ‘format’: **2010**

Linked Data



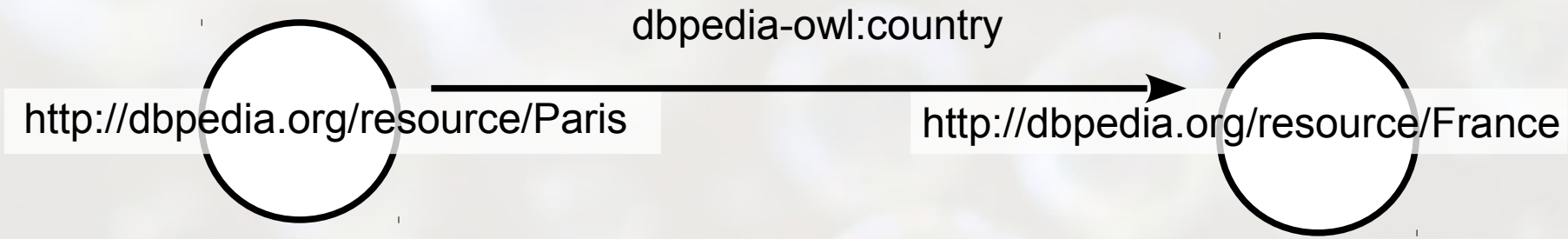
Datasets published following Linked Data ‘format’: **2011**

Linked Data
04/2014



Source: http://lod-cloud.net/

DBPedia (URIs)

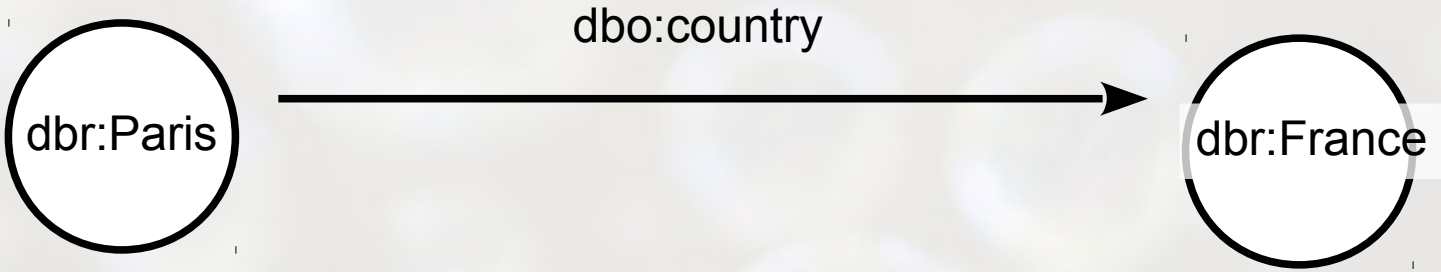


Predefined Namespace Prefixes

http://dbpedia.org/sparql?nsdecl

- dbpedia → http://dbpedia.org/resource/
- dbpedia-owl → http://dbpedia.org/ontology/
- dbprop → http://dbpedia.org/property/

Paris – France
URIs



Ontology Classes

http://mappings.dbpedia.org/server/ontology/classes/

Class	Examples
City	Cambridge, Berlin, Manchester
Country	Spain, Iceland, South Korea
Politician	George W. Bush, Nicolas Sarkozy, Angela Merkel
Musician	AC/DC, Diana Ross, Röyksopp
Music album	Led Zeppelin III, Like a Virgin, Thriller
Director	Woody Allen, Oliver Stone, Takashi Miike
Film	The Great Beauty, Hysterical Blindness, Breakfast at Tiffany's
Book	The Lord of the Rings, The Adventures of Tom Sawyer, the Bible
Computer Game	Tetris, World of Warcraft, Sam & Max hit the Road
Technical Standard	HTML, RDF, URI

(DBPedia, http://wiki.dbpedia.org/Datasets2014, 09/05/2015)

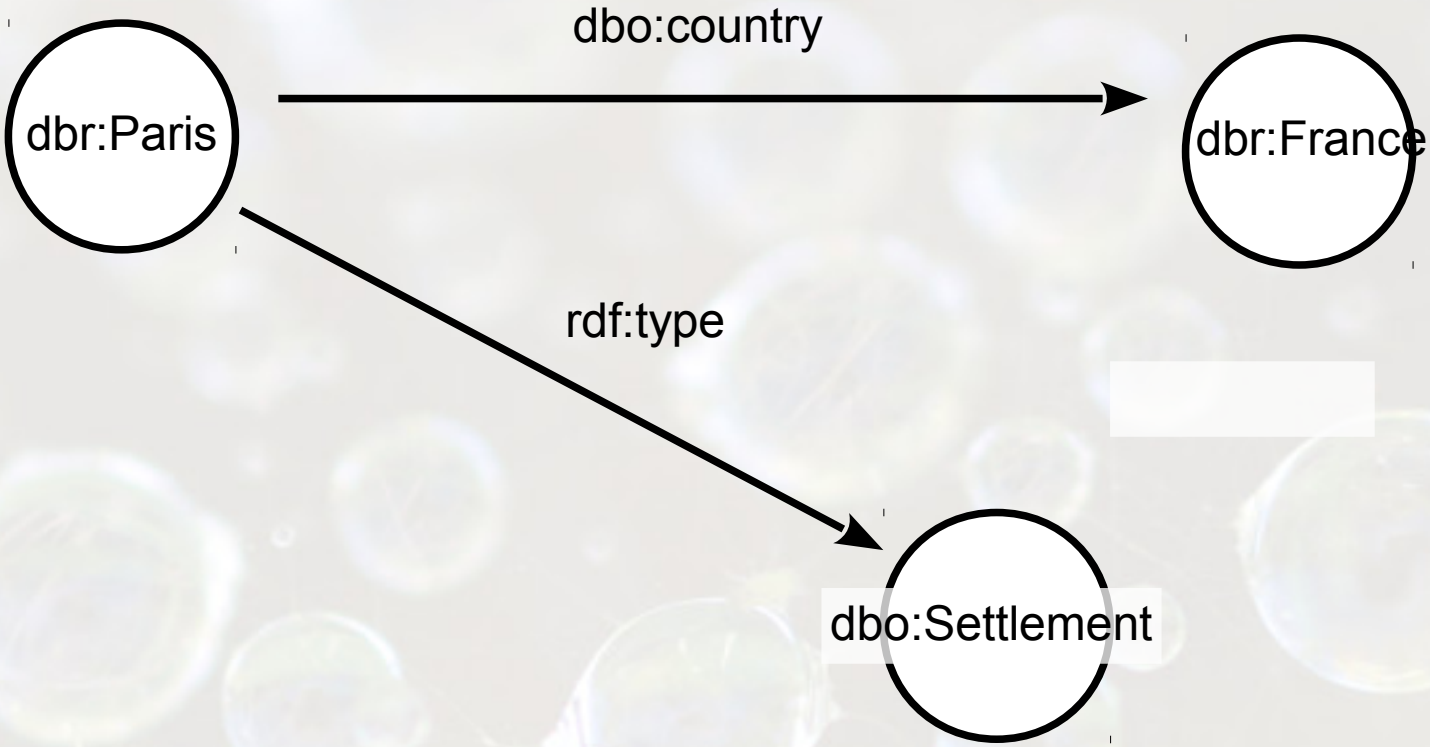
Ontology Classes

http://mappings.dbpedia.org/server/ontology/classes/

Class	Instances
Resource (overall)	4.233.000
Place	735
Person	1.450.000
Work	411
Species	251
Organisation	241

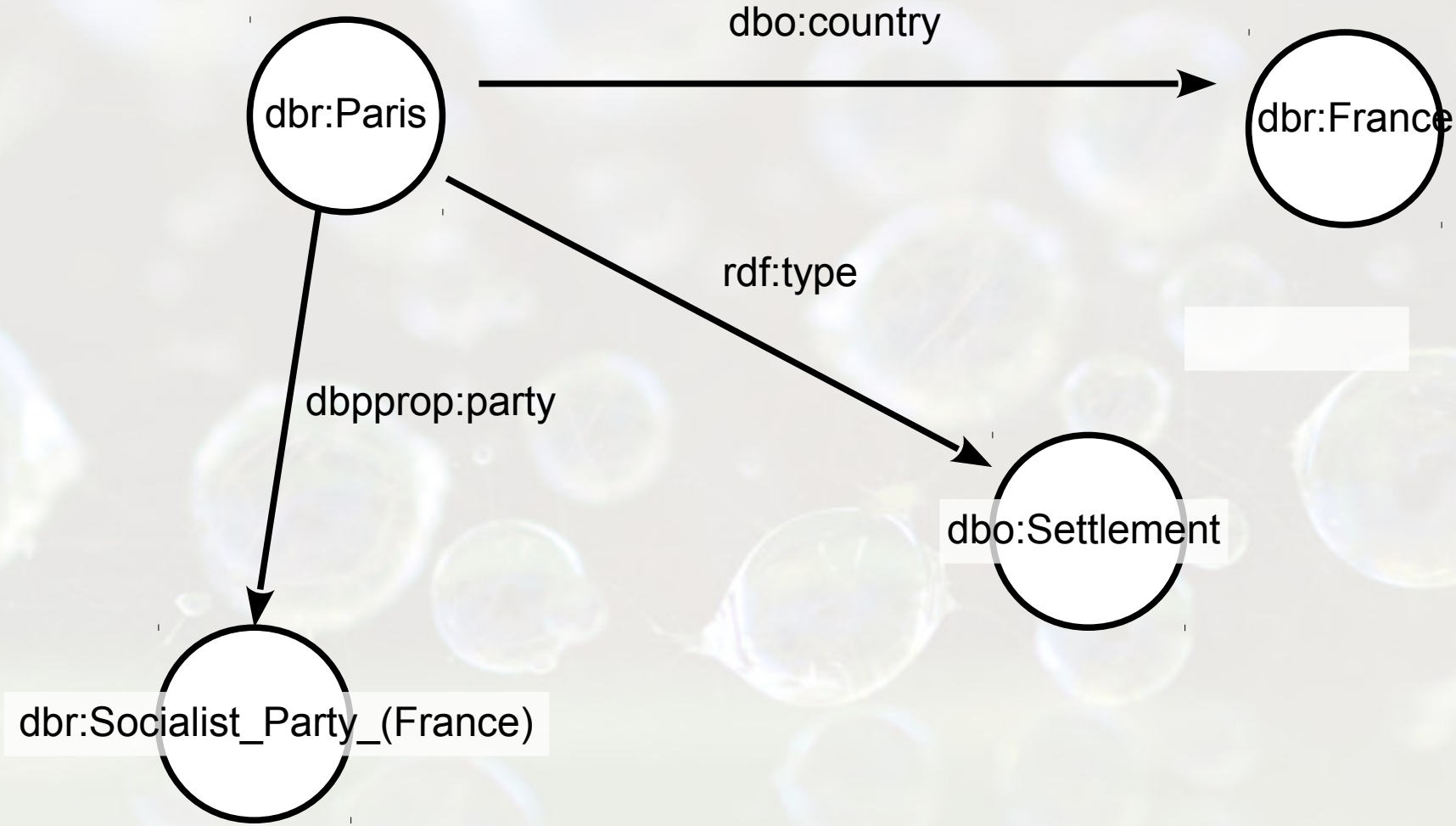
(DBPedia, http://wiki.dbpedia.org/Ontology, 09/05/2015)

Settlement Class



The `rdf:type` edge denotes that a resource (Paris) belongs to a Class (Settlement).

Property party

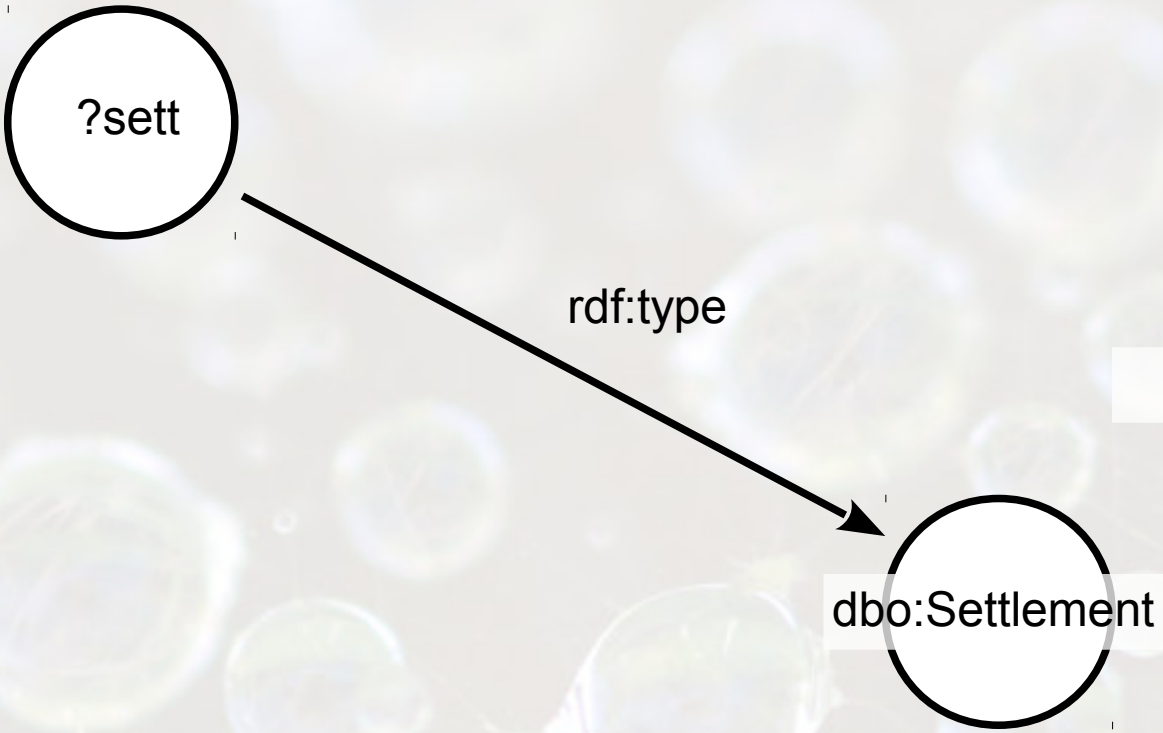


DBPedia SPARQL Query Editor

http://dbpedia.org/sparql

Paris Example

Instances of the Settlement class




```
select distinct ?sett
  where {?sett a dbo:Settlement}
```

Mayor of Paris

```
SELECT ?m
WHERE {dbr:Paris dbo:mayor ?m}
```

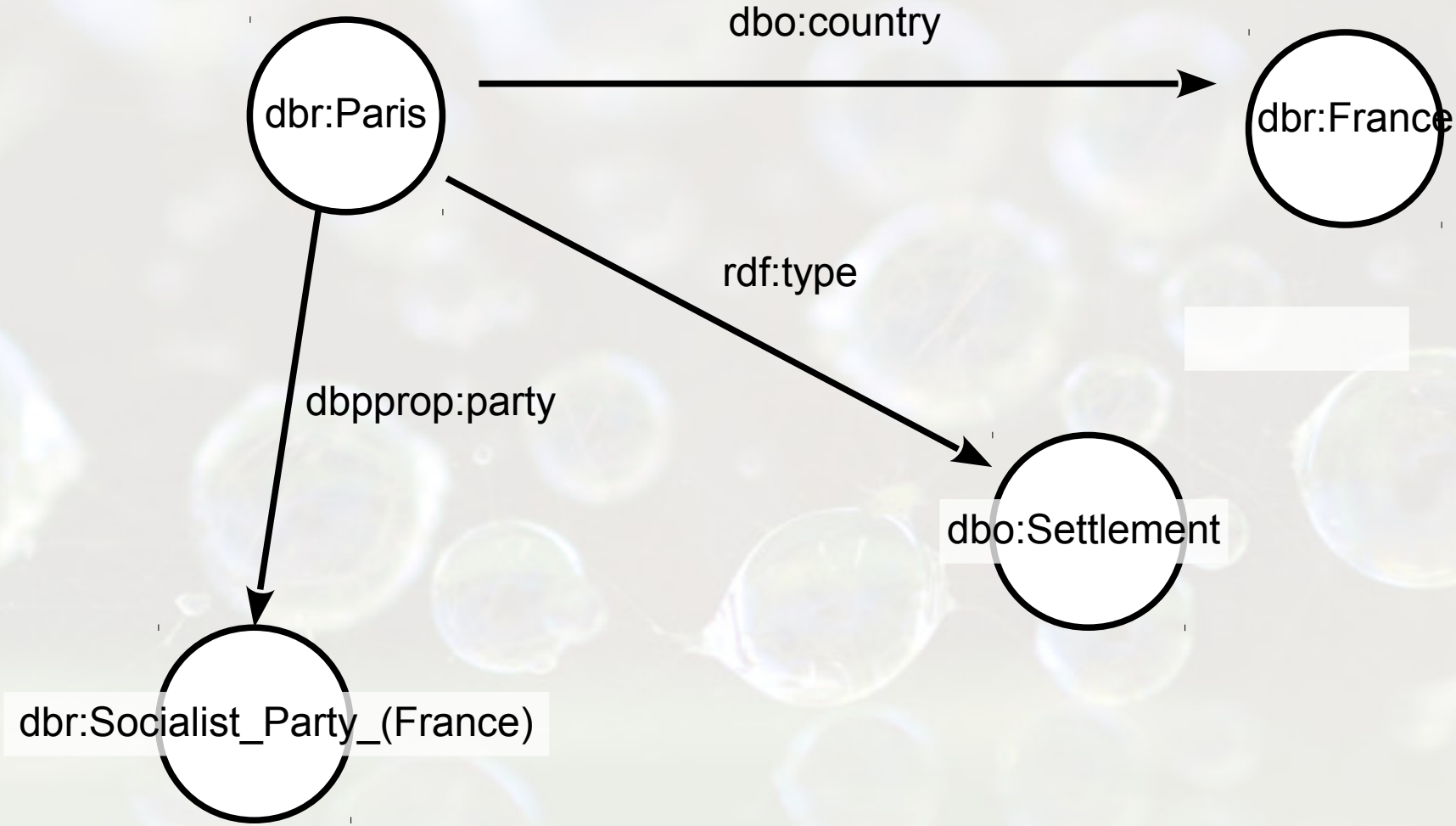
Mayor of Paris

```
SELECT ?m
WHERE {dbr:Paris dbo:mayor ?m}
```

Cities whose mayor was born in Paris

```
SELECT ?c
   where {?c dbo:mayor ?m .
          ?m dbo:birthPlace dbr:Paris}
```

Property party



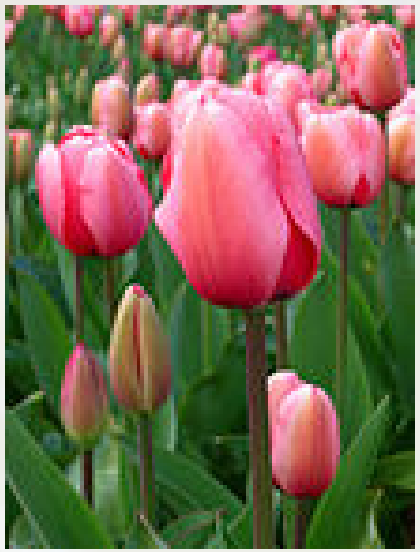
Cities managed by the French Socialist Party

```
select distinct ?organ
  where {?organ a dbo:Settlement .
         ?organ dbp:party <http://dbpedia.org/resource/Socialist_Party_(France)>}
```

Tulip Example

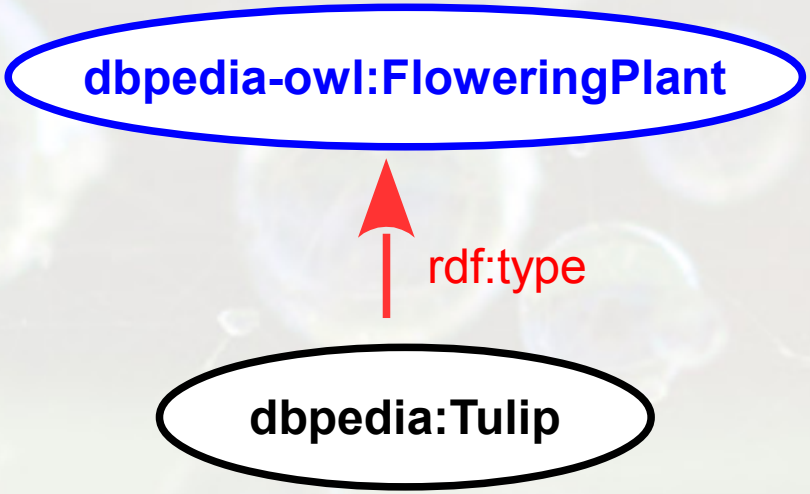
Tulip

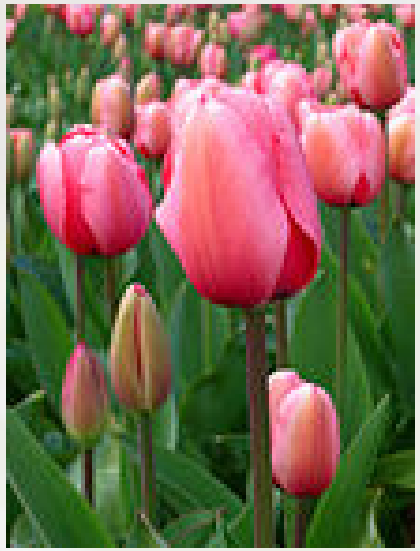
- http://en.wikipedia.org/wiki/Tulip
- dbpedia:Tulip
  - http://dbpedia.org/resource/Tulip
- http://dbpedia.org/page/Tulip



Tuplip – Class FloweringPlant

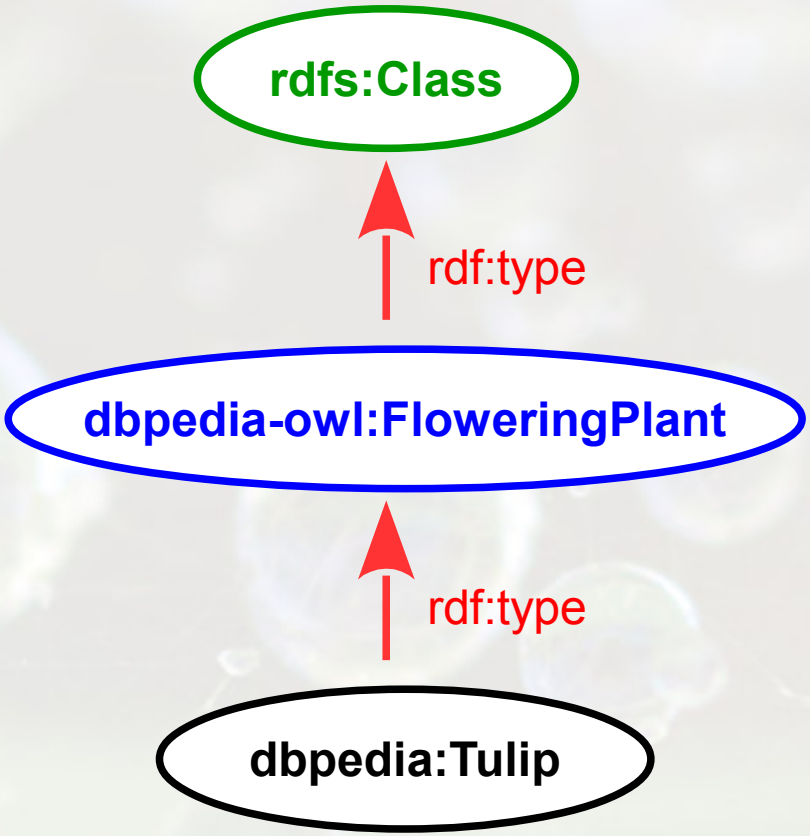
the `rdf:type` edge denotes that a resource (node) belongs to a Class





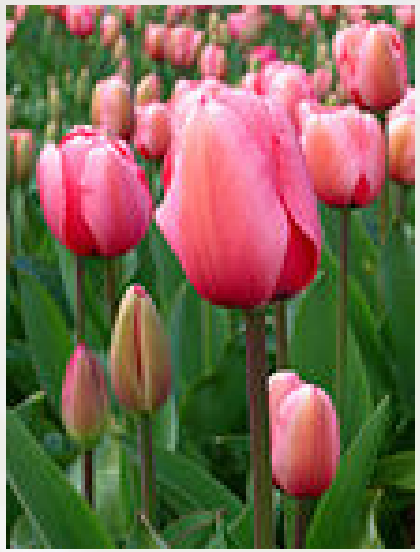
dbpedia:Tulip rdf:type dbpedia-owl:FloweringPlant .

Tuplip – Class FloweringPlant
Triple Representation

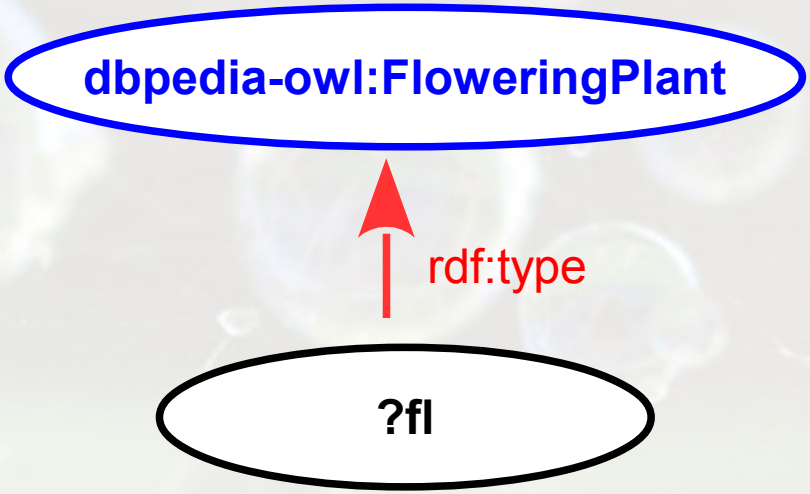


dbpedia:Tulip rdf:type dbpedia-owl:FloweringPlant .

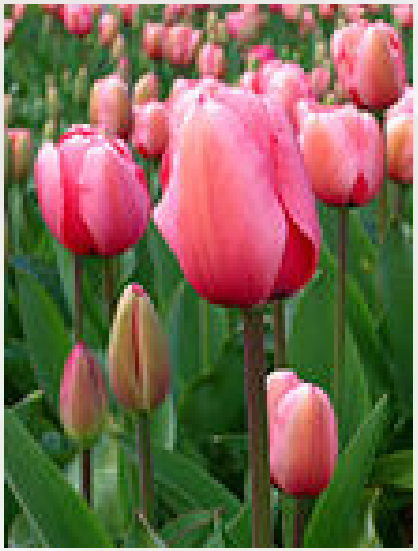
dbpedia-owl:FloweringPlant rdf:type rdfs:Class .



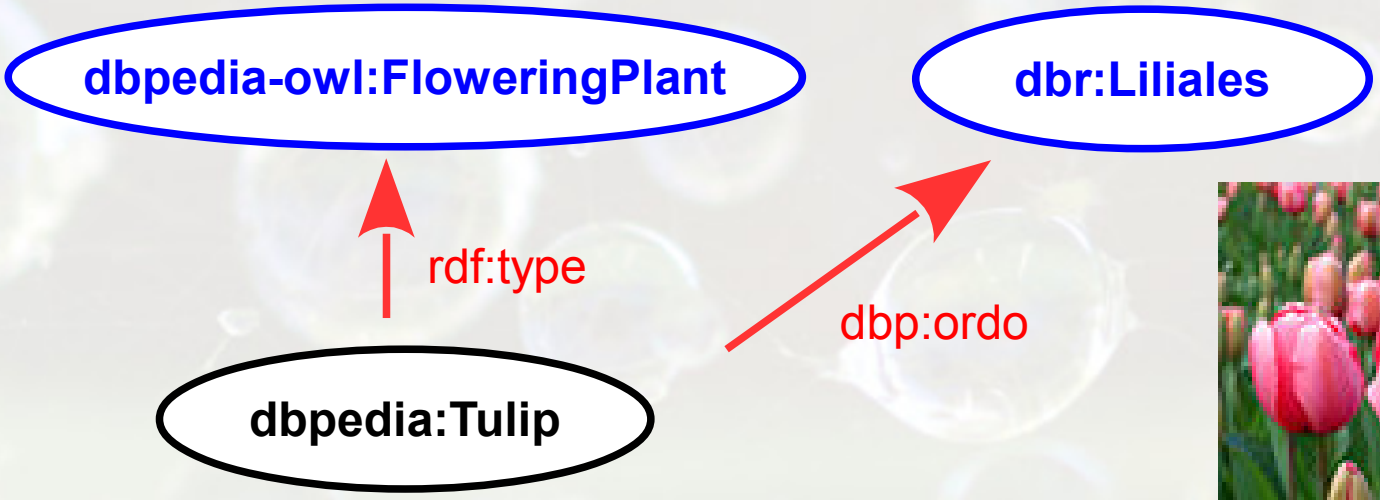
All FloweringPlants
instances of the FloweringPlant class

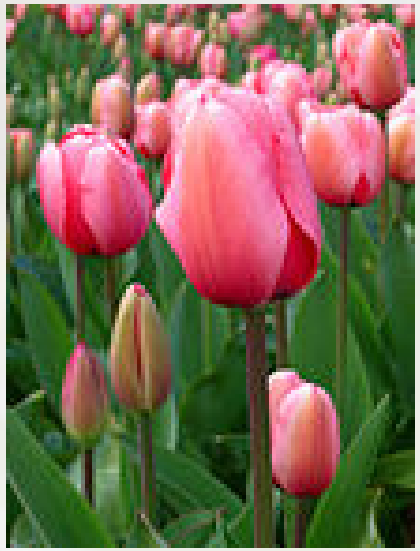



```
select ?fl
  where {?fl a dbo:FloweringPlant}
```

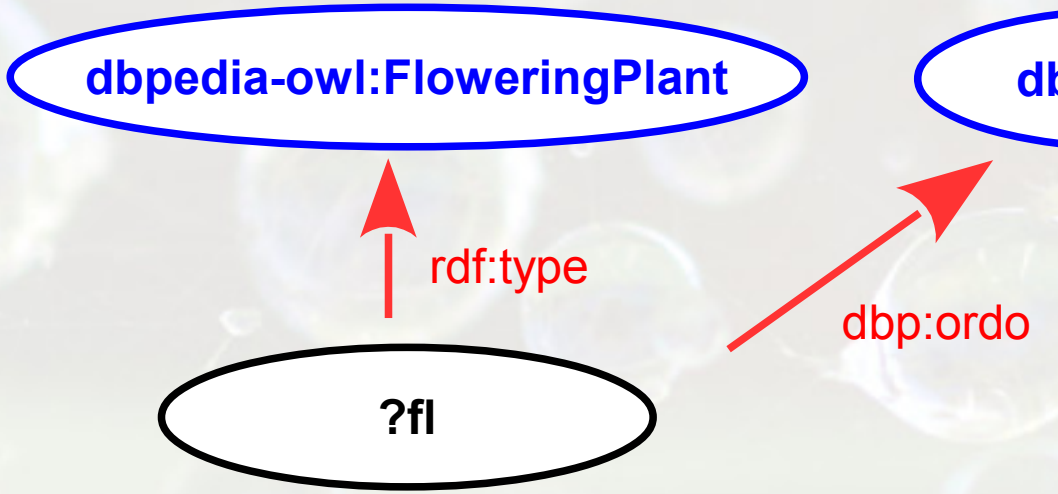


Property ordo

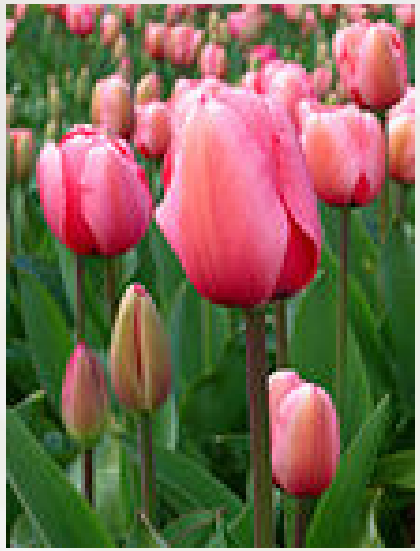




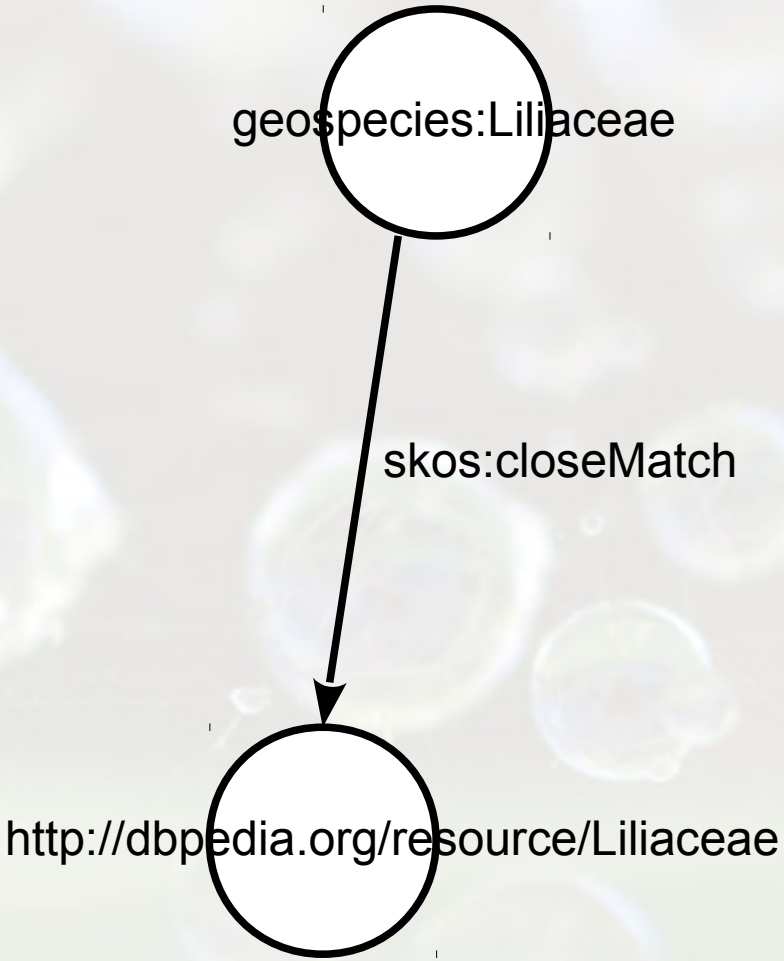
All FloweringPlants ordo Liliales




```
SELECT ?fl
WHERE {
  ?fl rdf:type dbo:FloweringPlant ;
  dbp:ordo dbr:Liliales .
}
```

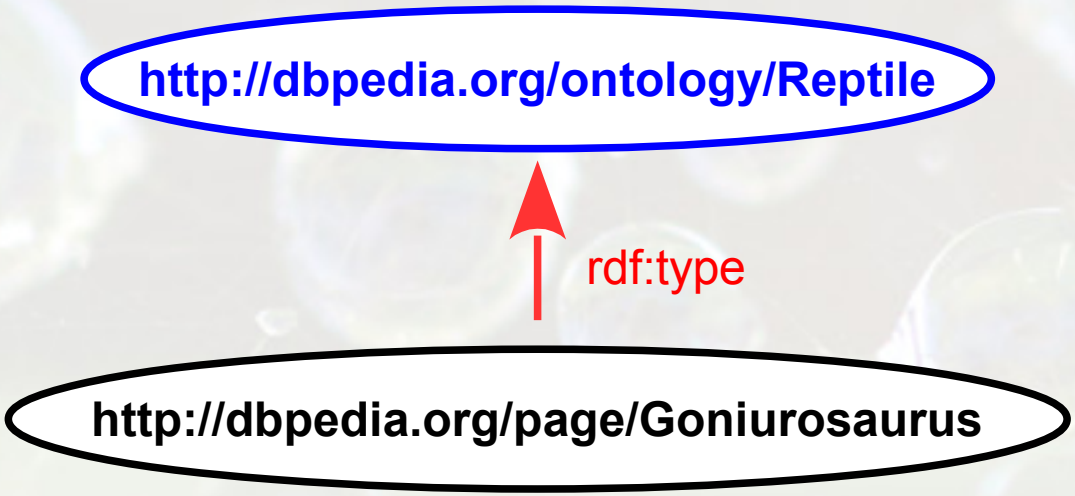


Challenge: how to exploit?



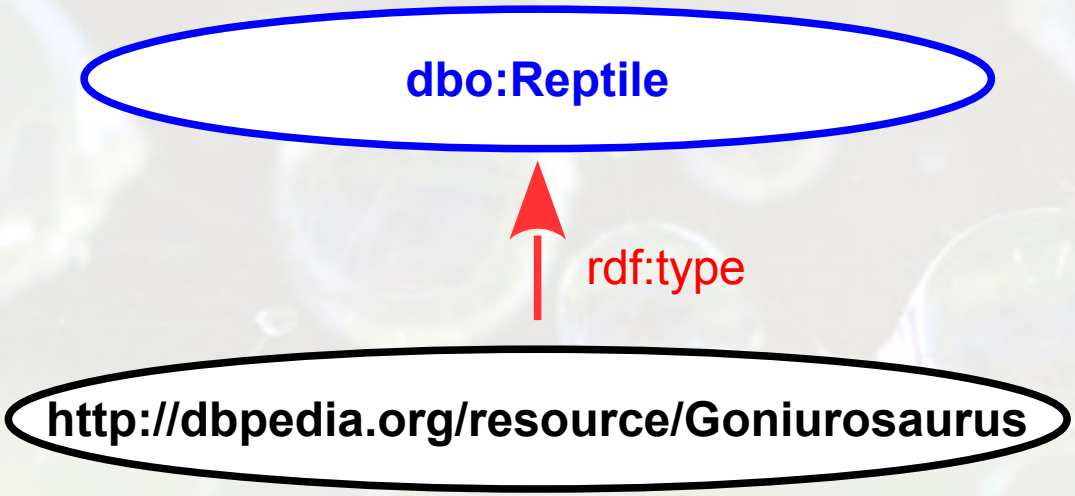
Reptile Example

Reptile Class
without namespaces



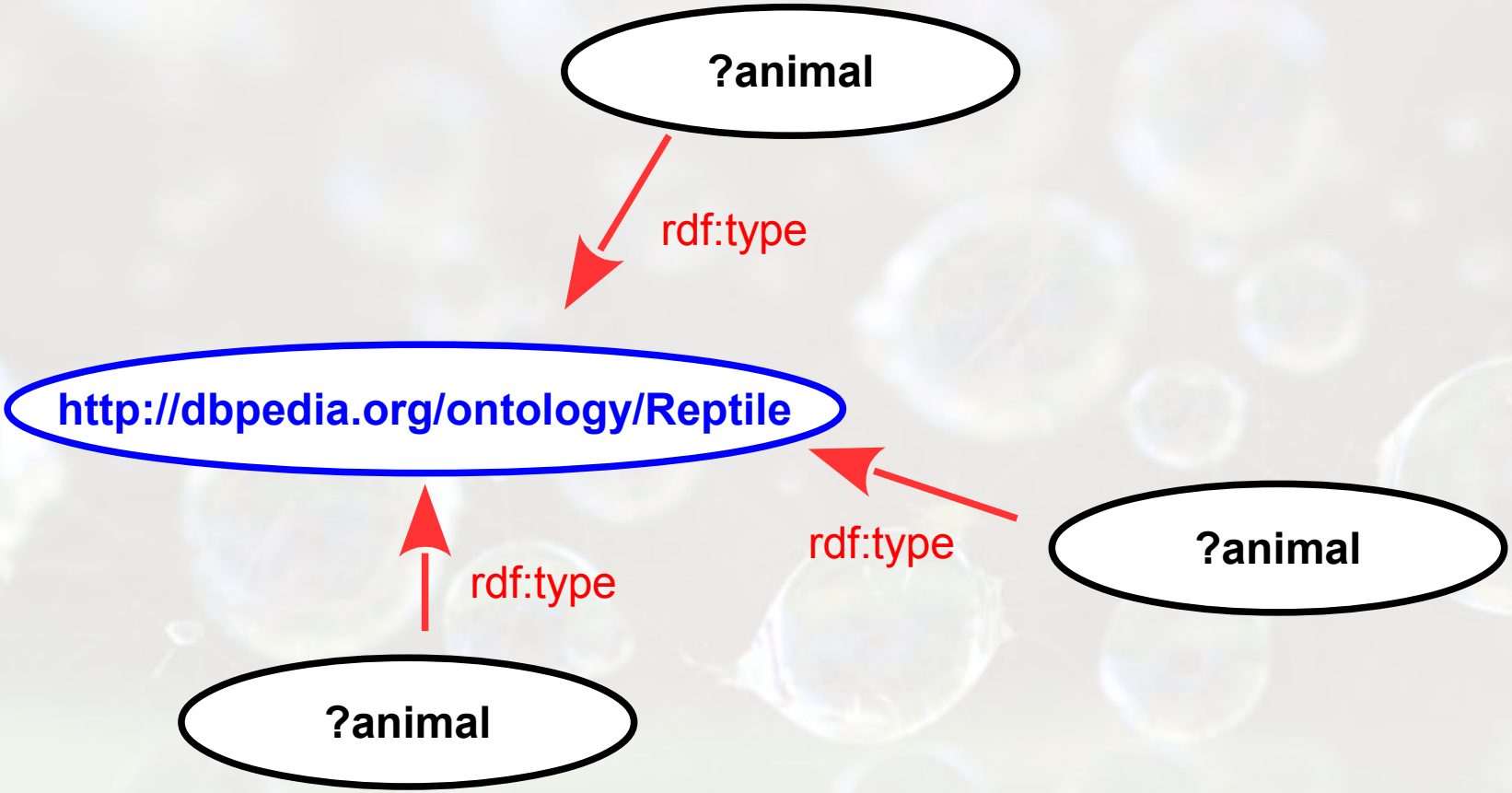
<http://dbpedia.org/page/Goniurosaurus> rdf:type **<http://dbpedia.org/ontology/Reptile>**

Reptile Class
with namespaces



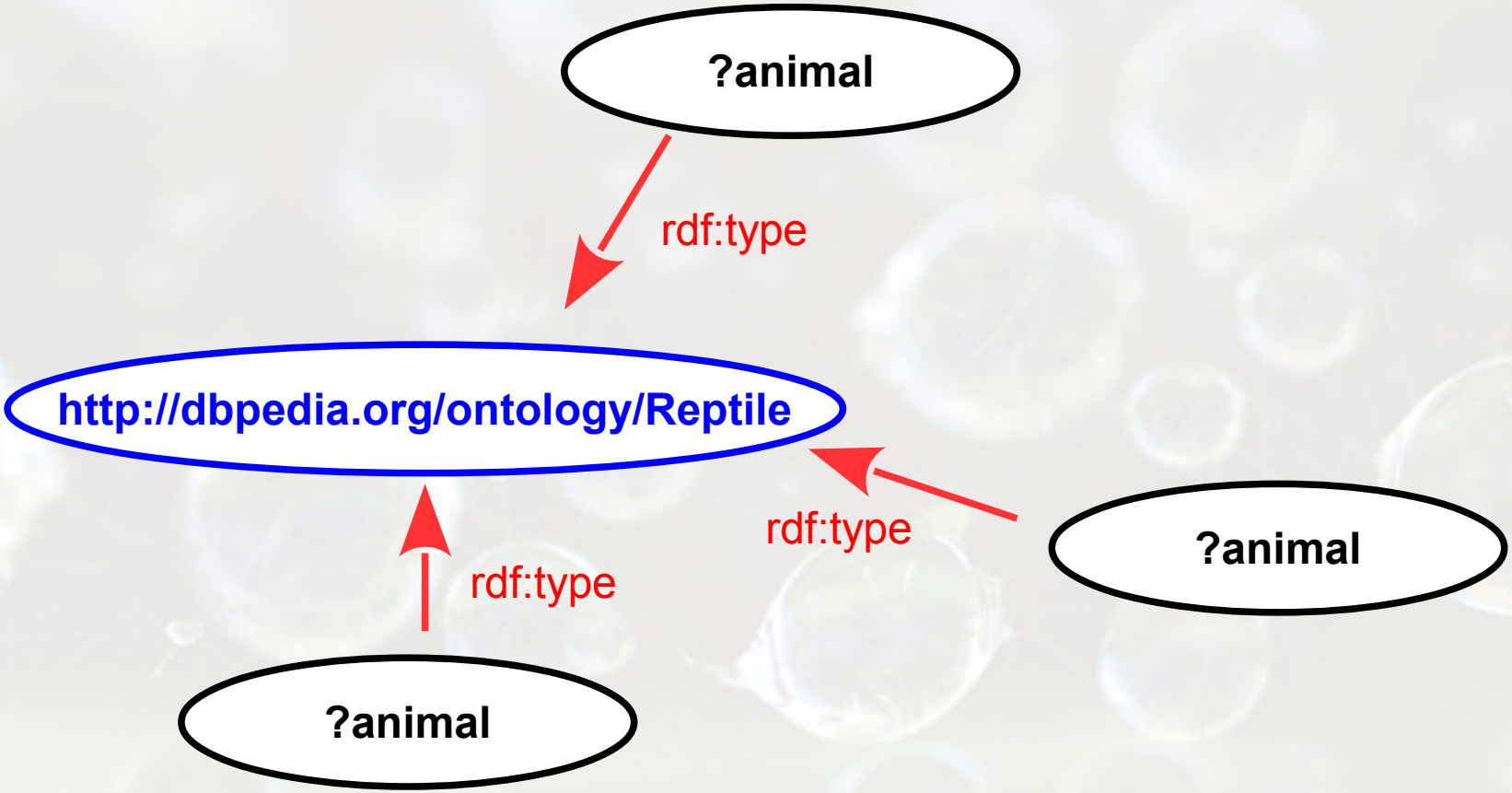
dbr:Goniurosaurus rdf:type **dbo:Reptile**

Nodes of the Reptile Class - SPARQL
without namespaces



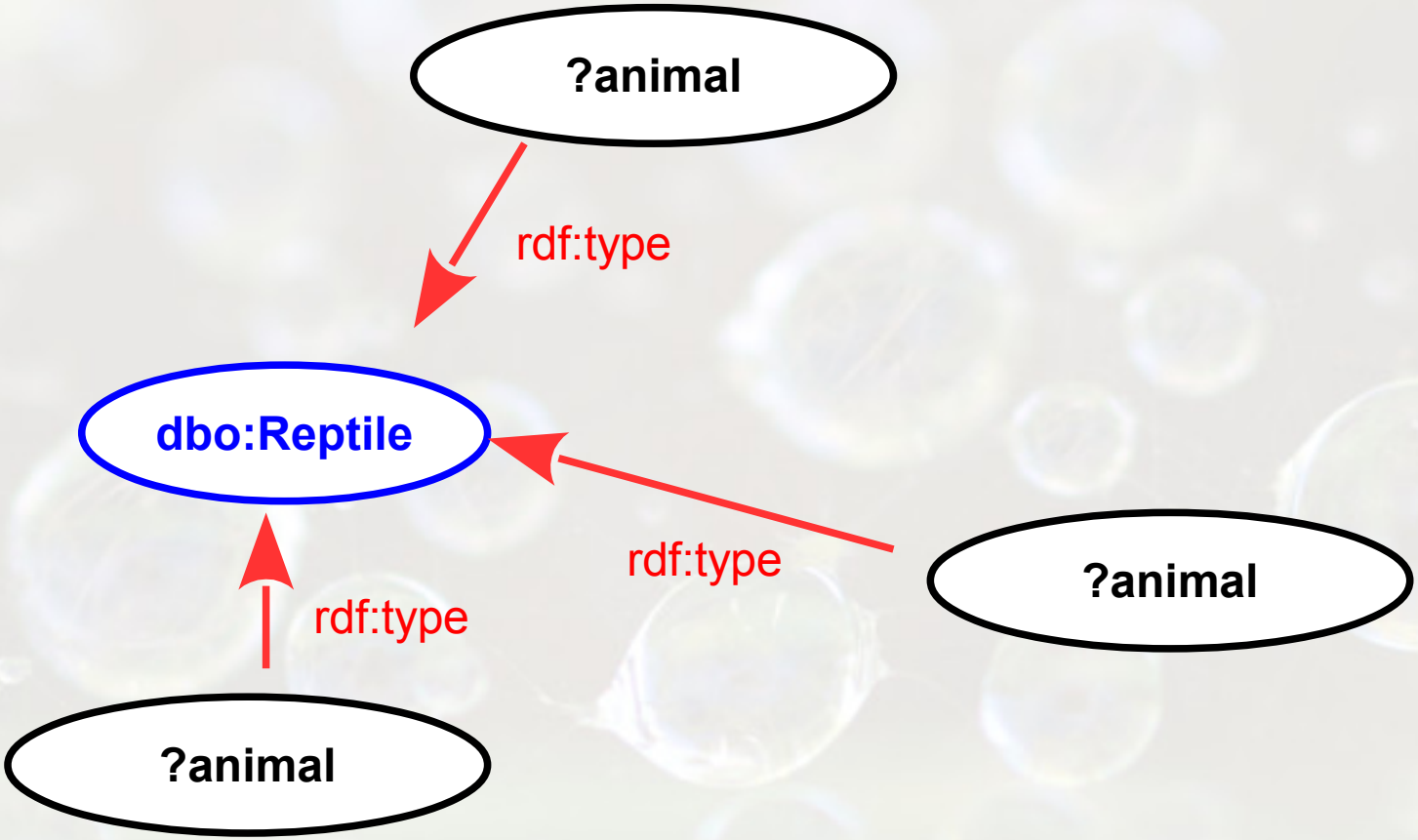

```
SELECT ?animal
  WHERE {?animal rdf:type <http://dbpedia.org/ontology/Reptile>}
```

SPARQL
“a” replacing “rdf:type”



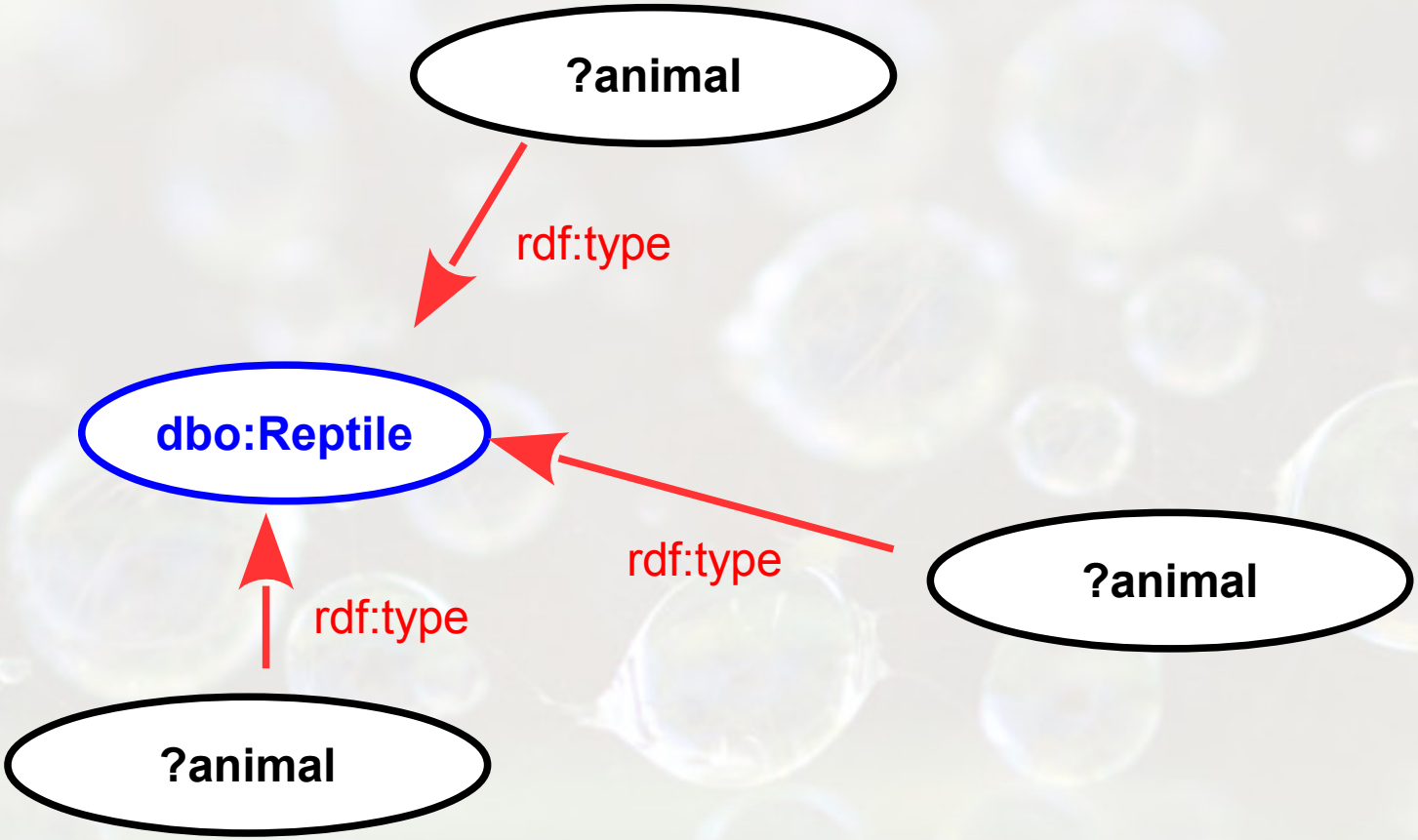

```
SELECT ?animal
  WHERE {?animal a <http://dbpedia.org/ontology/Reptile>}
```

Nodes of the Reptile Class - SPARQL with namespaces



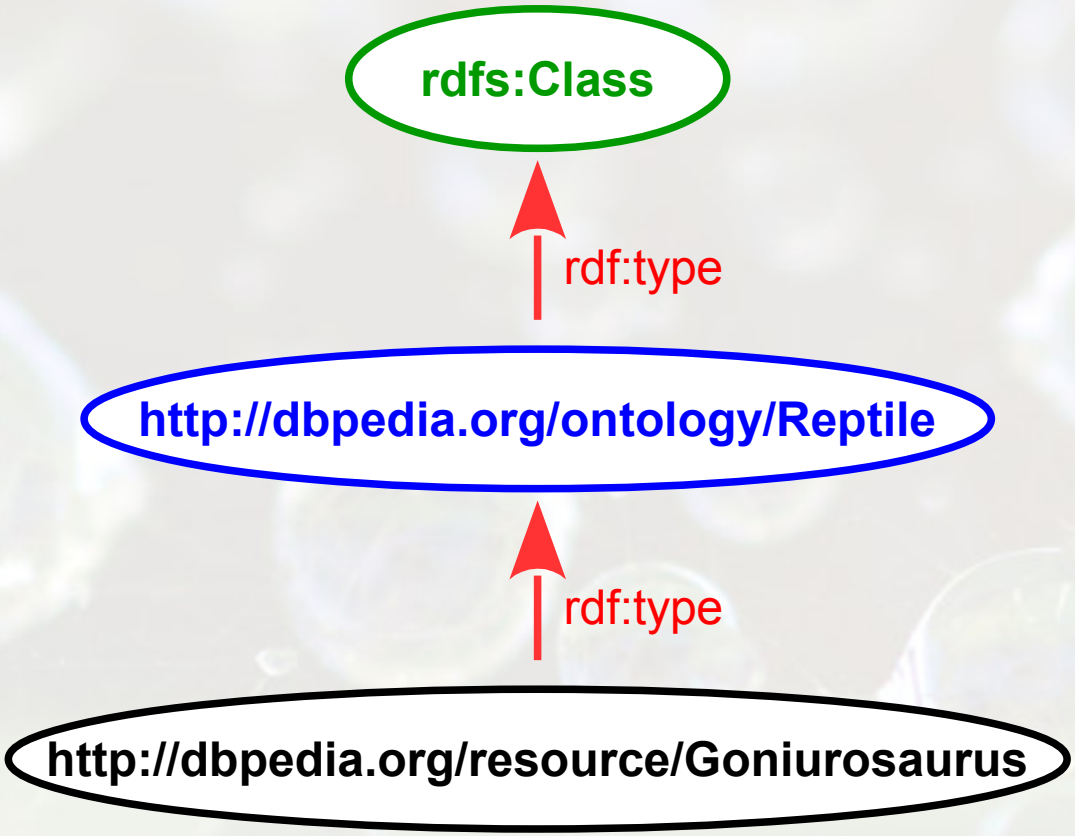

```
SELECT ?animal
  WHERE {?animal rdf:type dbo:Reptile}
```

SPARQL
“a” replacing “rdf:type”




```
SELECT ?animal
  WHERE {?animal a dbo:Reptile}
```

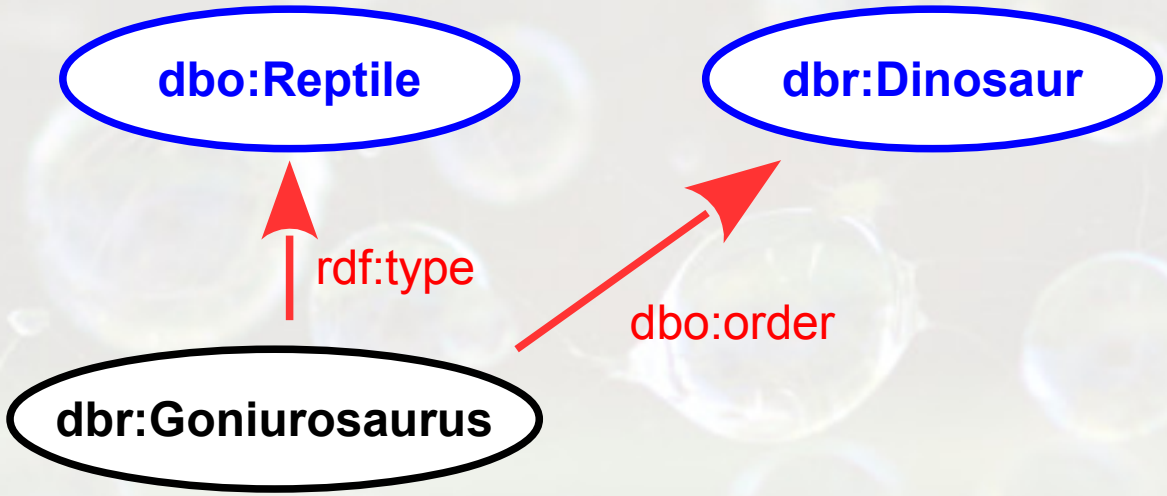
Reptile Class



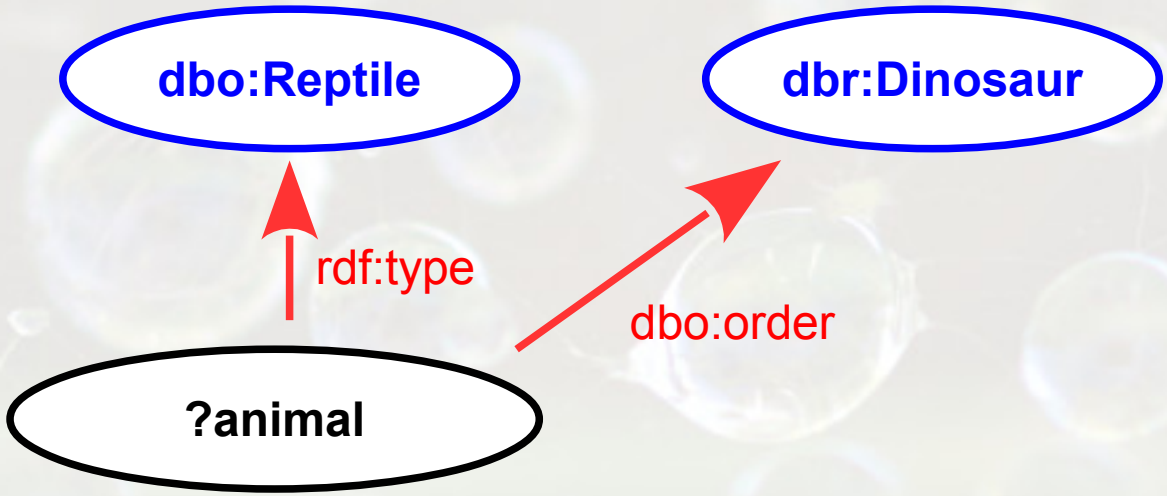
http://dbpedia.org/resource/Goniurosaurus rdf:type **http://dbpedia.org/ontology/Reptile** .

http://dbpedia.org/ontology/Reptile rdf:type rdfs:Class .

Property Order: Dinosaur

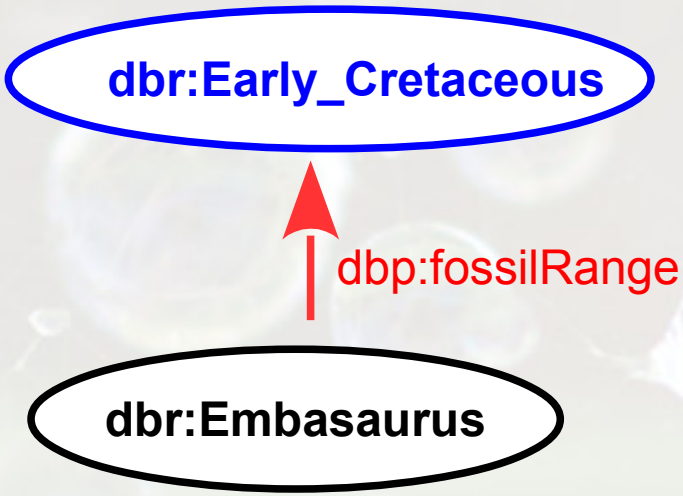


Nodes of the Reptile Class and Order Dinosaur – SPARQL

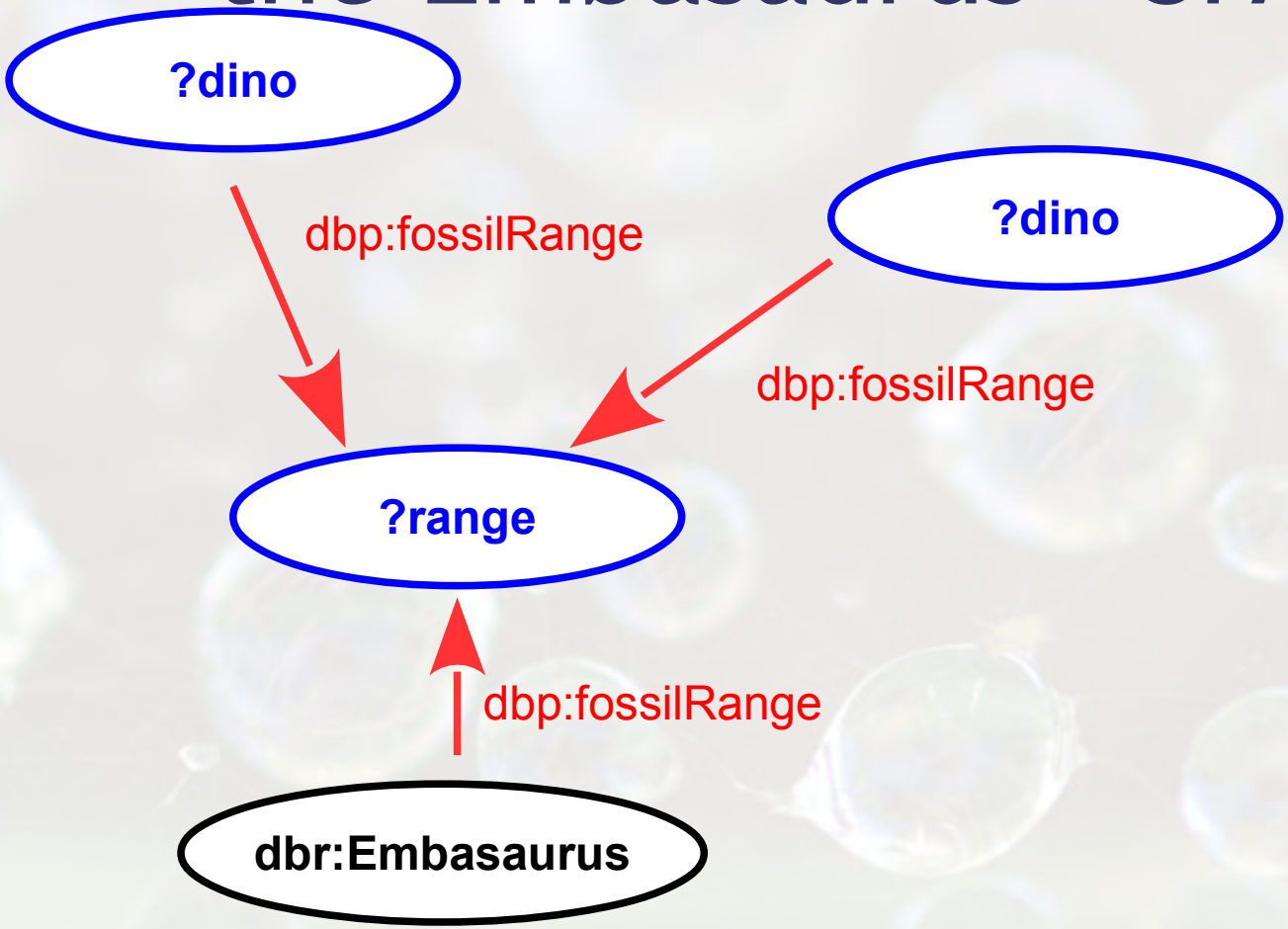



```
SELECT ?animal
  WHERE {?animal rdf:type dbo:Reptile .
         ?animal dbo:order dbr:Dinosaur}
```

Property fossilRange: Early_Cretaceous



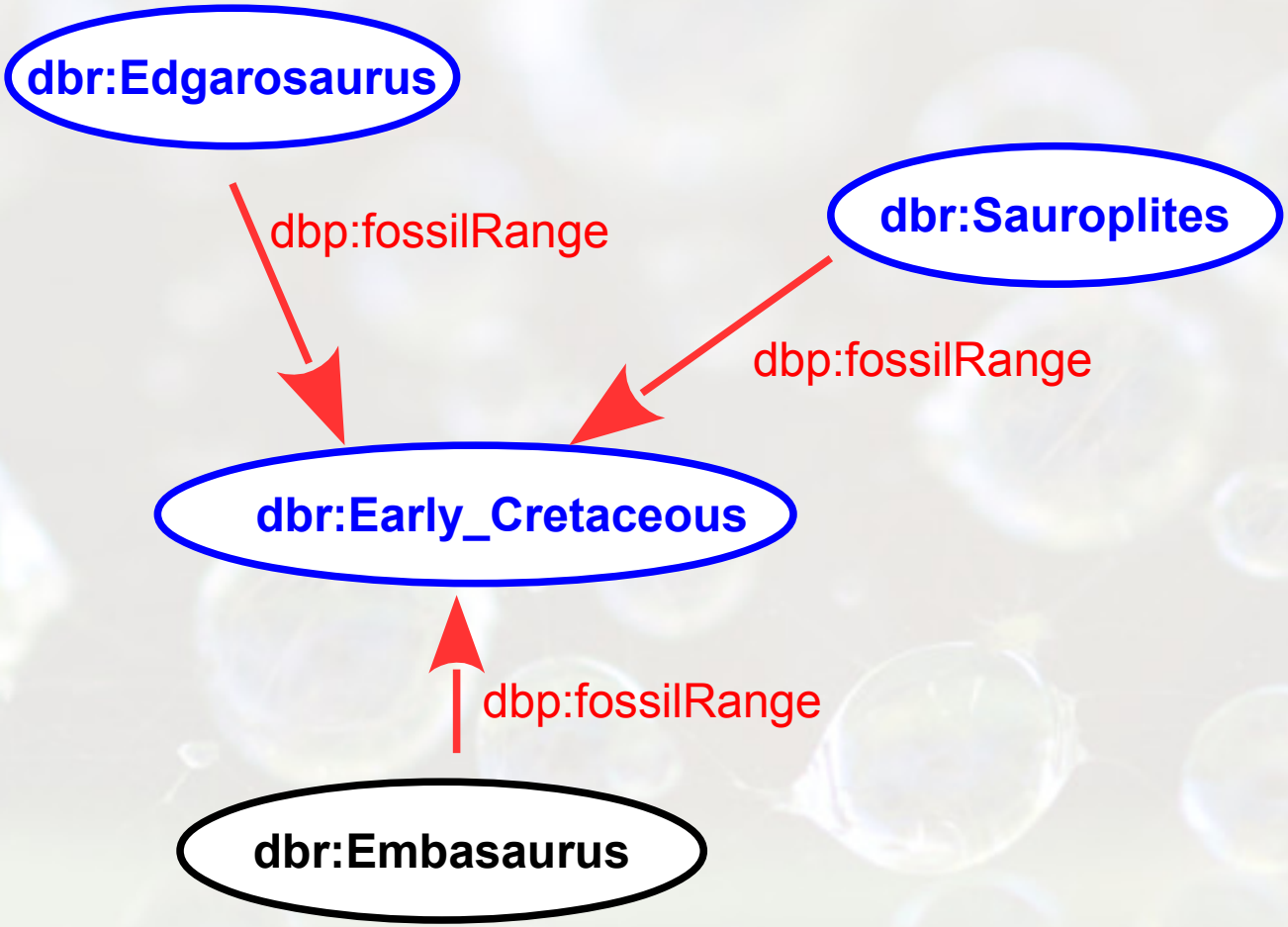
Nodes of the same fossilRange of the Embasaurus - SPARQL




```
SELECT ?dinosaur
  WHERE {dbr:Embasaurus dbp:fossilRange ?range .
         ?dinosaur dbp:fossilRange ?range}
```

Nodes of the same fossilRange of the Embasaurus - SPARQL

Some Results




```
SELECT ?dinosaur
  WHERE {dbr:Embasaurus dbp:fossilRange ?range .
         ?dinosaur dbp:fossilRange ?range}
```

Processing the Query

- Process by pattern
  - Find [species] whose [origin] → (0..*) part of → EUA
- Process by inference
  - Find [species] whose [origin] → EUA
  - Rules:
    - If (A) origin (B) and (B) part of (C) => (A) origin (C)
    - If (A) part of (B) and (B) part of (C) => (A) part of (C)

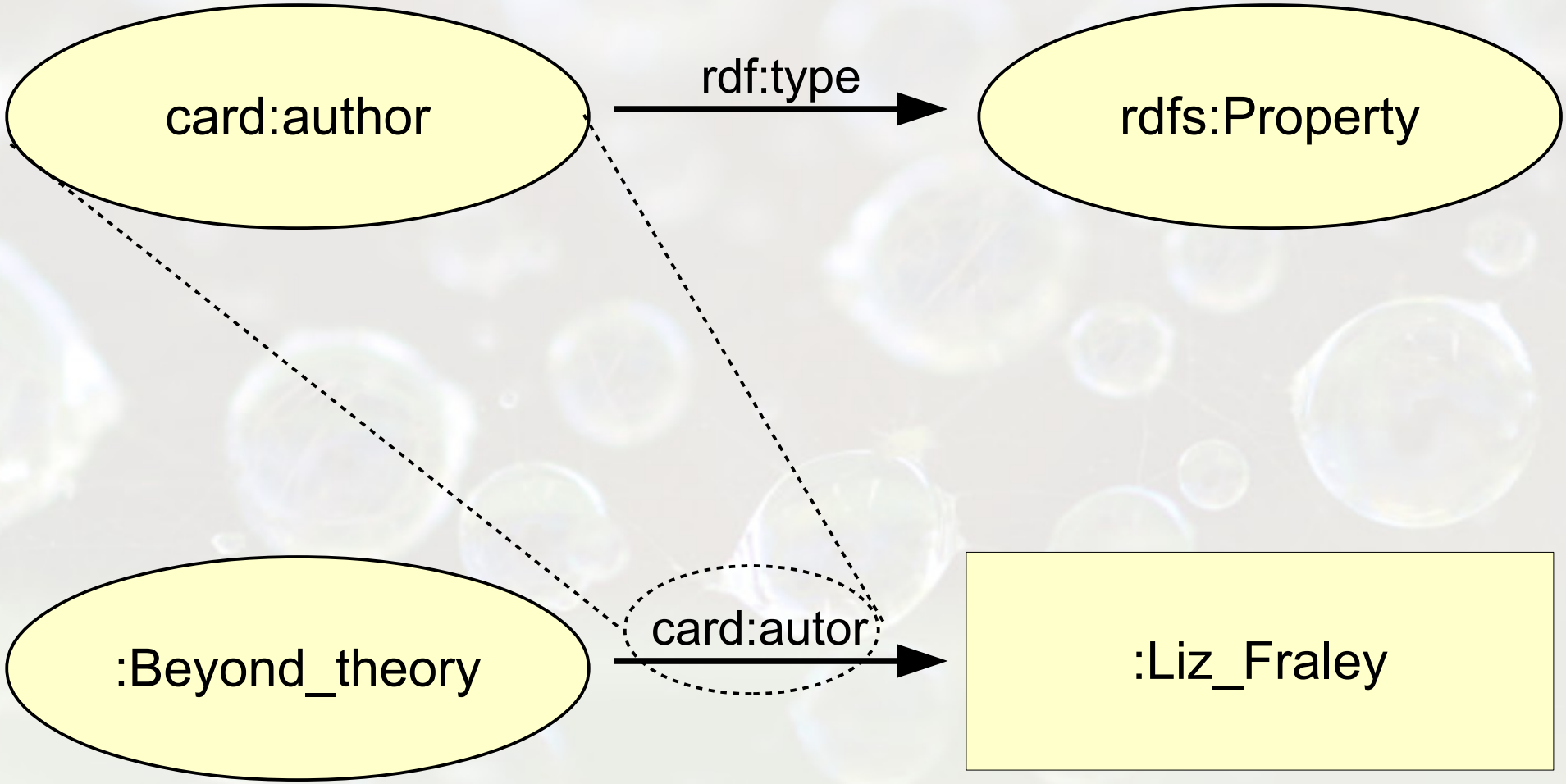
Cypher

```
MATCH (dn:Dinosaur)-[:Found]->(pl:Place)
MATCH (pl)-[:Part_of*0..5]->(pl2:Place) WHERE pl2.id="USA"
RETURN dn.id
```

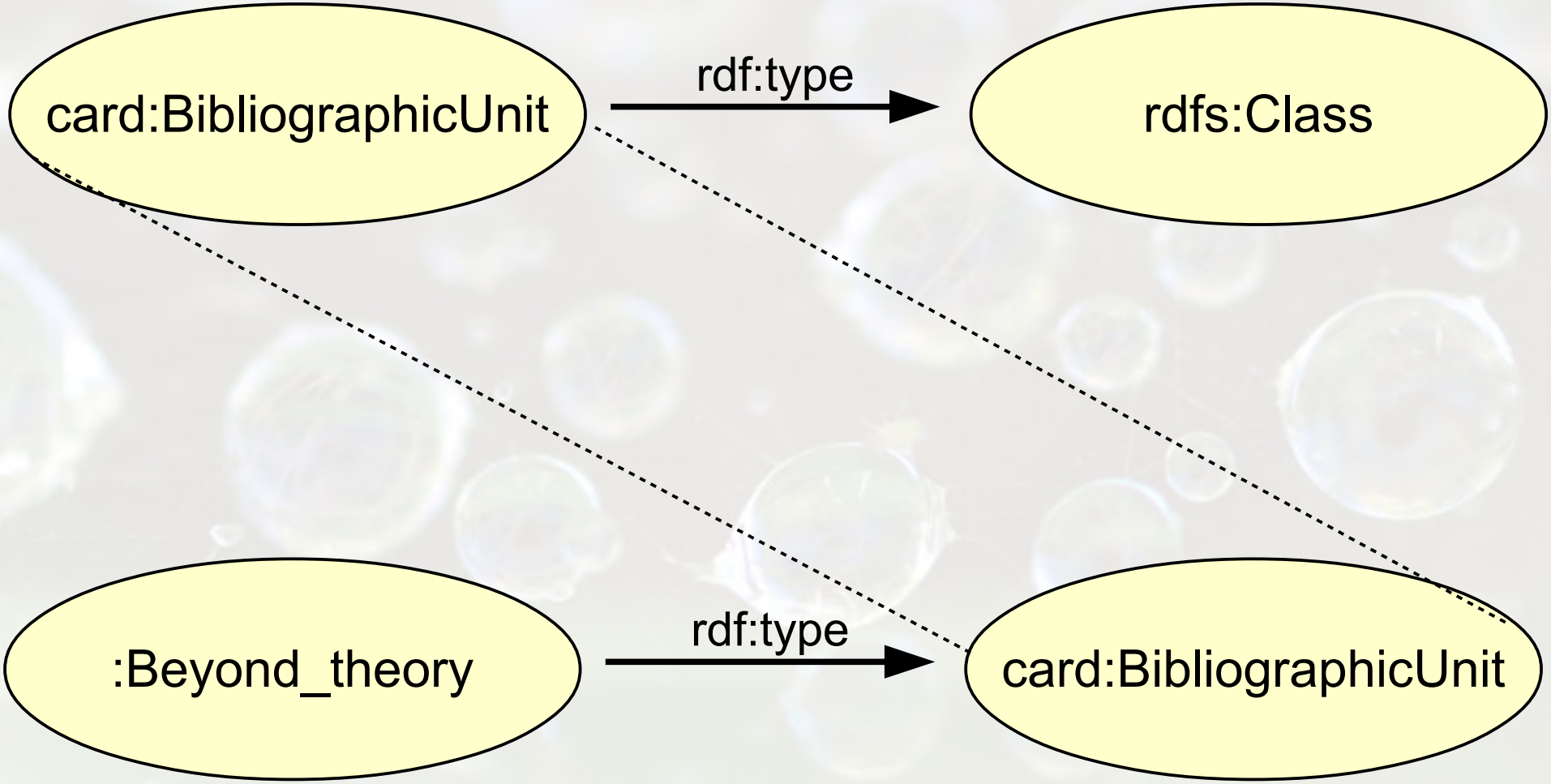
SPARQL

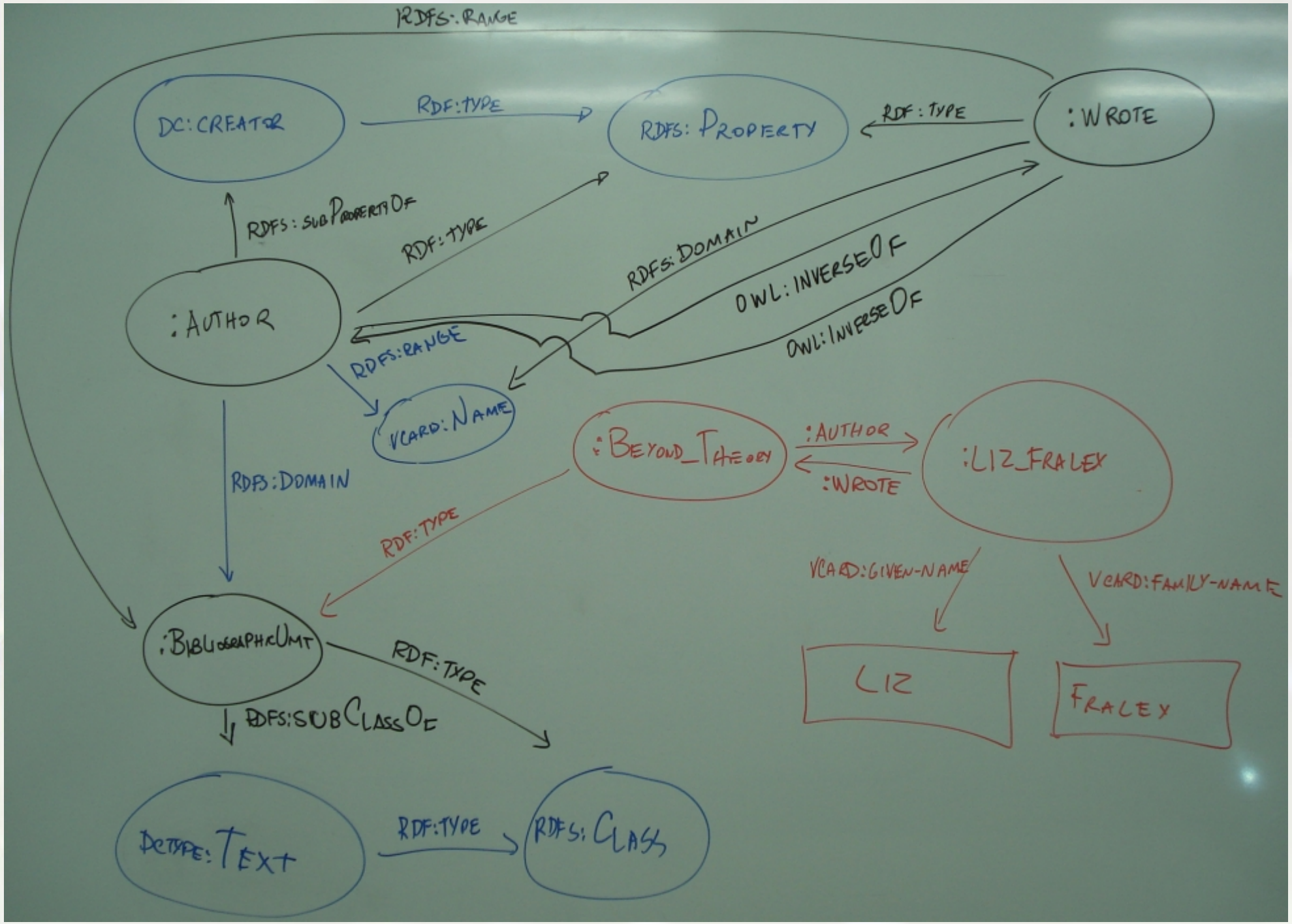
Ontologia ResearchCard

Propriedade

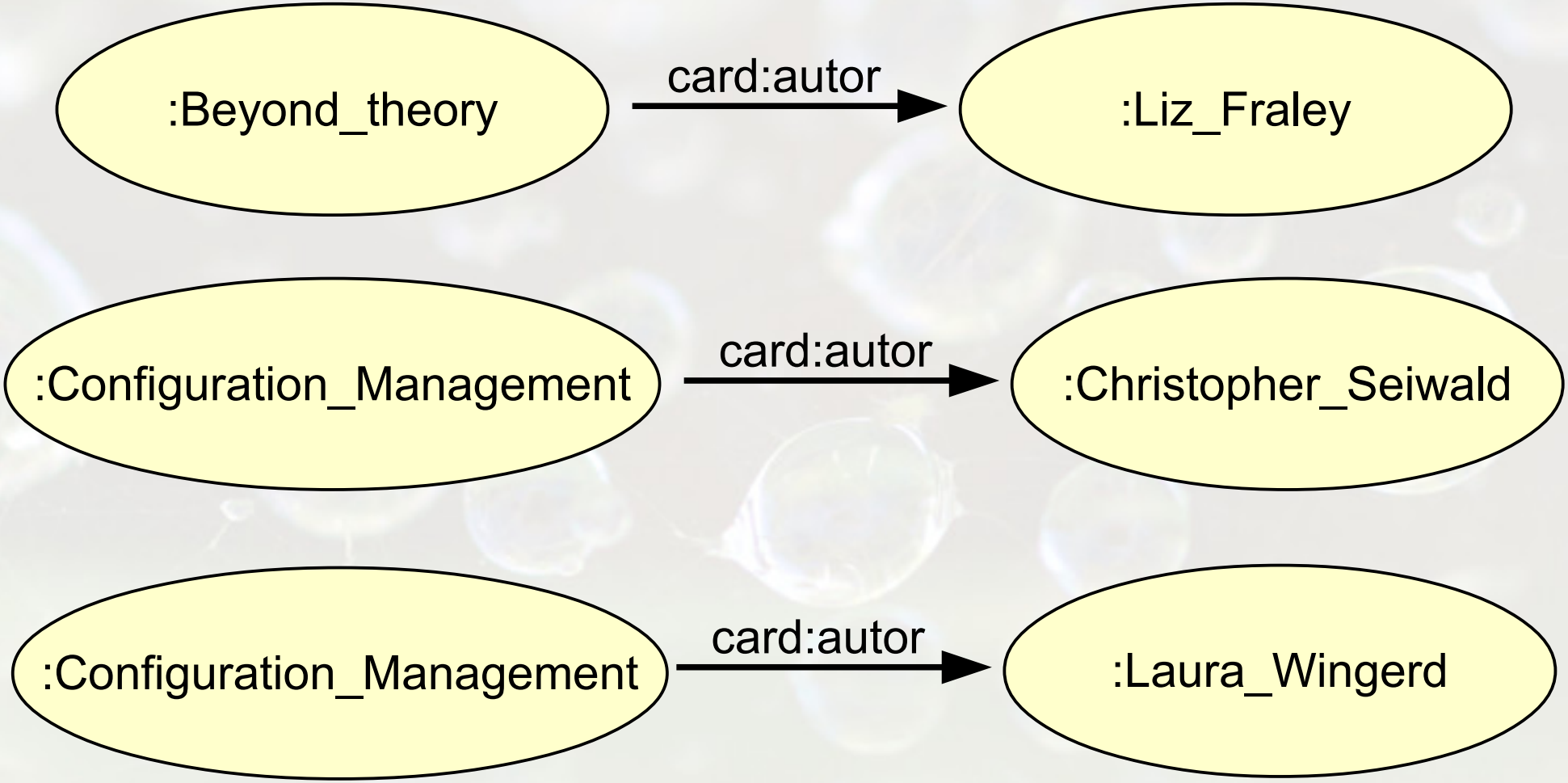


Classe


card:BibliographicUnit
rdf:type
rdfs:Class
:Beyond_theory
rdf:type
card:BibliographicUnit


RDFS:RANGE
DC:CREATOR
RDF:TYPE
RDFS:PROPERTY
RDF:TYPE
:WROTE
RDFS:SUBPROPERTYOF
RDF:TYPE
RDFS:DOMAIN
OWL:INVERSEOF
OWL:INVERSEOF
:AUTHOR
RDFS:RANGE
VCARD:NAME
:BEYOND_THEORY
:AUTHOR
:WROTE
:LIZ_FRALEY
RDFS:DOMAIN
RDF:TYPE
VCARD:GIVEN-NAME
VCARD:FAMILY-NAME
:BIBLIOGRAPHICUMT
RDF:TYPE
LIZ
FRALEY
RDFS:SUBCLASSOF
DCTYPE:TEXT
RDF:TYPE
RDFS:CLASS

Conjunto de Dados



```
SELECT ?nome
WHERE { ?nome rdf:type vcard:Name }
```

```
SELECT ?publicacao
WHERE { ?publicacao card:authors :Liz_Fraley }
```

```
SELECT ?publicacao ?autor
WHERE { ?publicacao card:authors ?autor }
```

```
SELECT ?assunto
WHERE { ?publicacao card:subjects ?assunto }
```

```
SELECT ?assunto
WHERE { ?publicacao card:authors :Liz_Fraley .
        ?publicacao card:subjects ?assunto }
```

```
SELECT ?assunto
WHERE { ?publicacao card:subjects ?assunto .
        ?assunto subj:superset subj:content_management }
```

```
SELECT ?assunto
WHERE { ?assunto rdf:type subj:Subject }
```

```
SELECT ?assunto
WHERE { ?assunto rdf:type subj:Subject .
        ?assunto subj:superset subj:content_management }
```

```
SELECT ?assunto
WHERE { ?assunto rdf:type subj:Subject .
        {
          { ?assunto subj:superset subj:content_management . }
          UNION
          { ?assunto subj:superset ?super .
            ?super subj:superset subj:content_management
          }
        }
}
```

Generic Examples

- All classes

```
SELECT ?classe
WHERE { ?classe rdf:type owl:Class }
```

- All triples (resource, property, value)

```
SELECT ?r ?p ?v
WHERE { ?r ?p ?v }
```

SWRL

- Place(?pl1), Place(?pl2), Place(?pl3), partOf(?pl1, ?pl2), partOf(?pl2, ?pl3) -> partOf(?pl1, ?pl3)

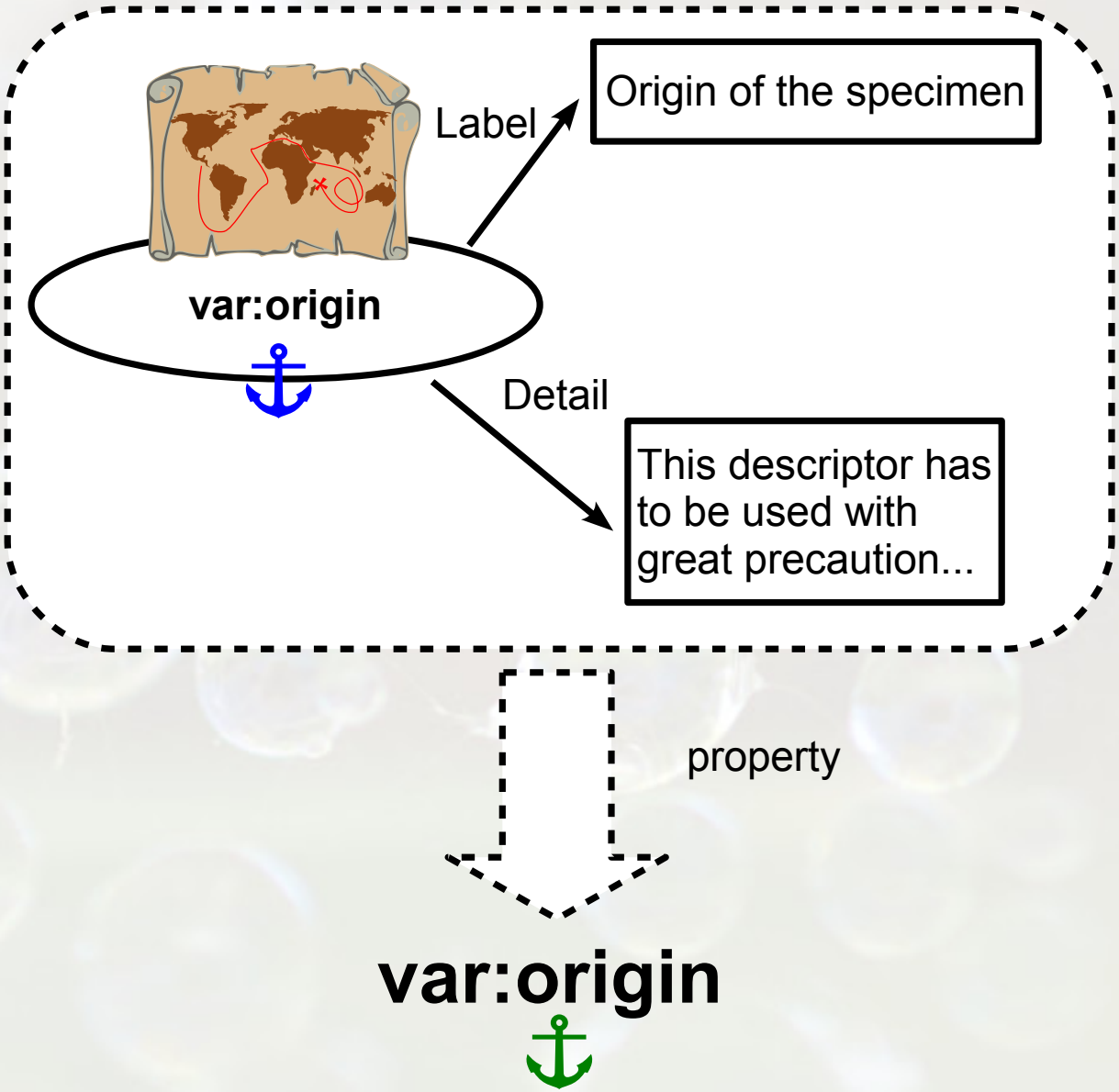
http://dbpedia.org/page/Eye

Agradecimentos

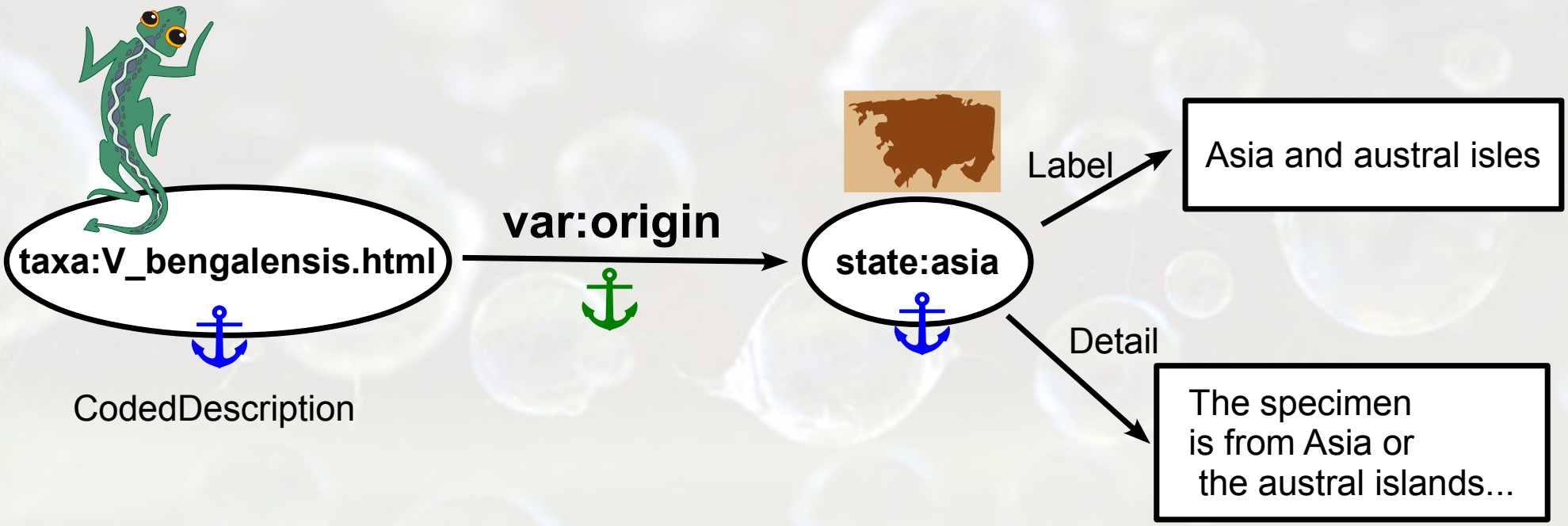
- Luiz Celso Gomes Jr (professor desta disciplina em 2014) pela contribuição na disciplina e nos slides.

André Santanchè
http://www.ic.unicamp.br/~santanche

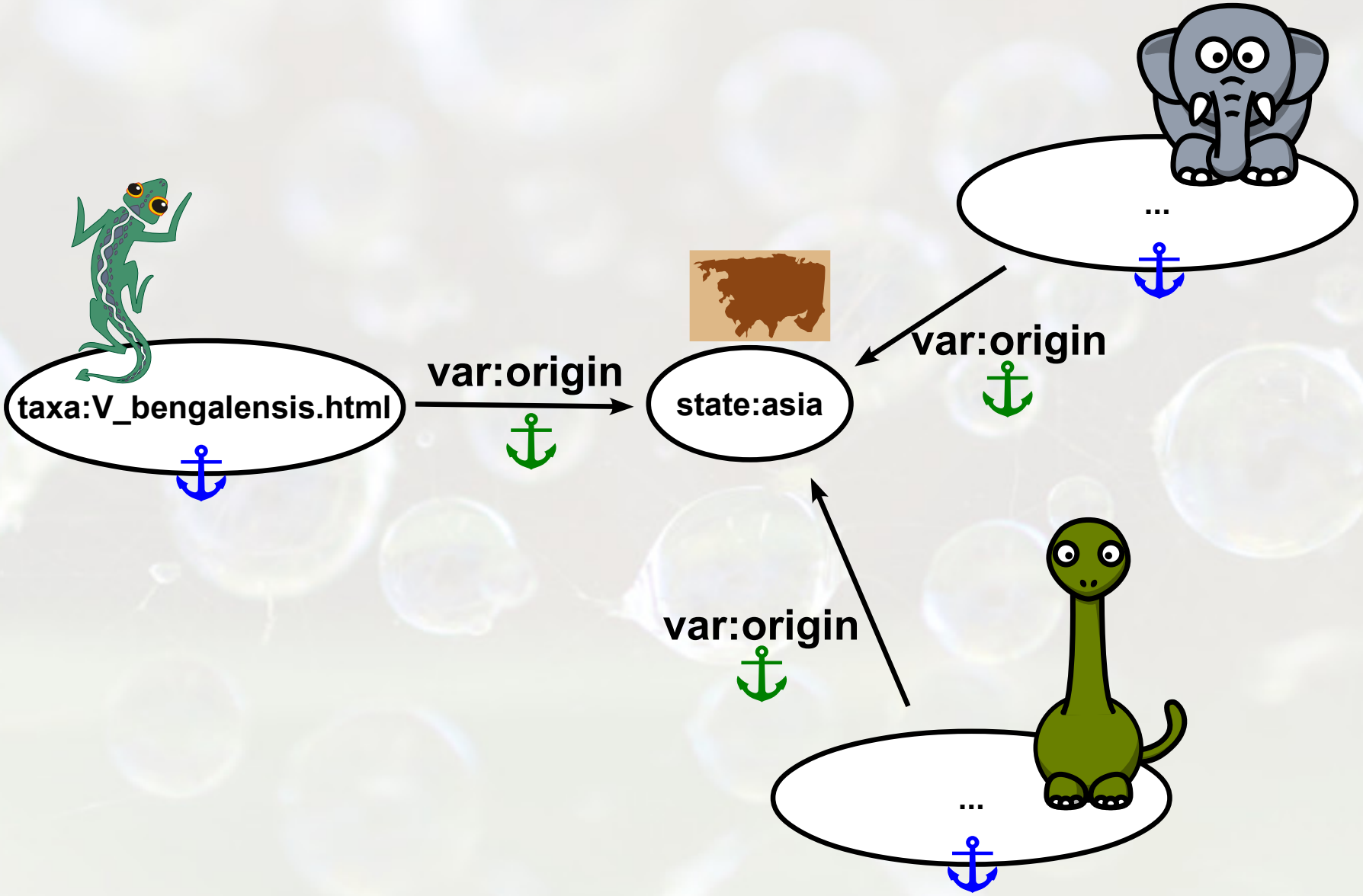
Property


Label
Origin of the specimen
var:origin
Detail
This descriptor has to be used with great precaution...
property
var:origin

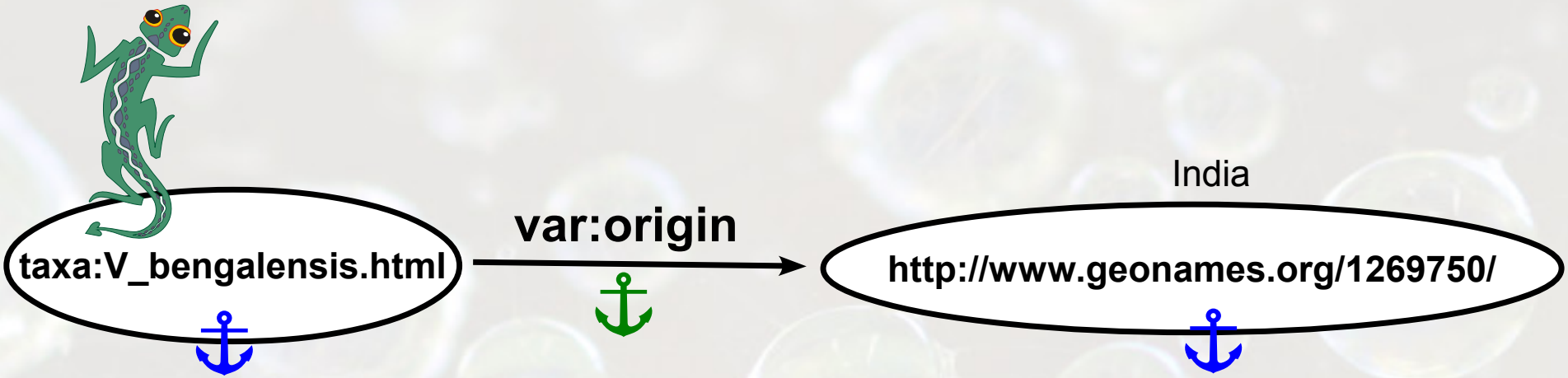
Using the Property



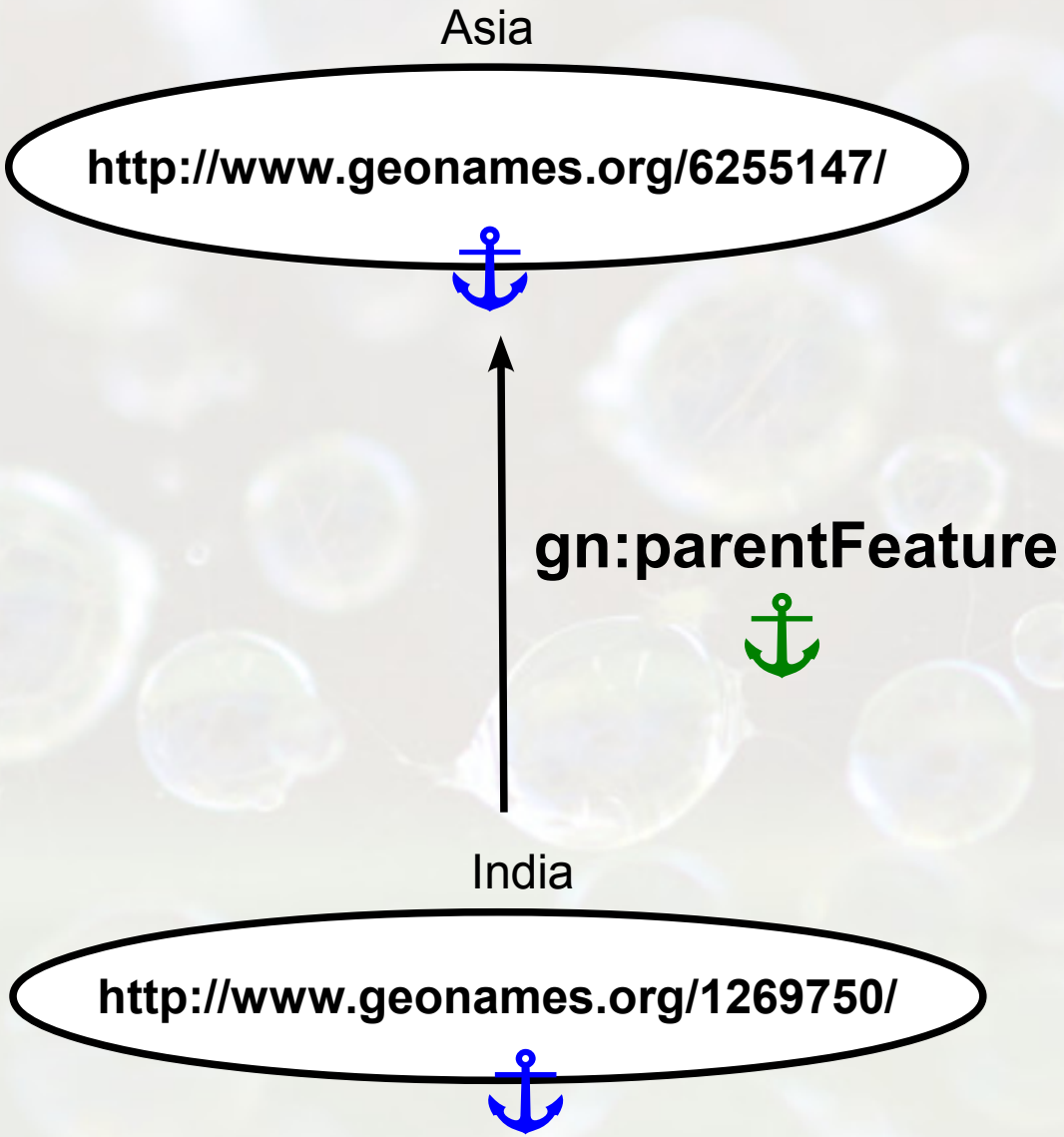
Connected Graph



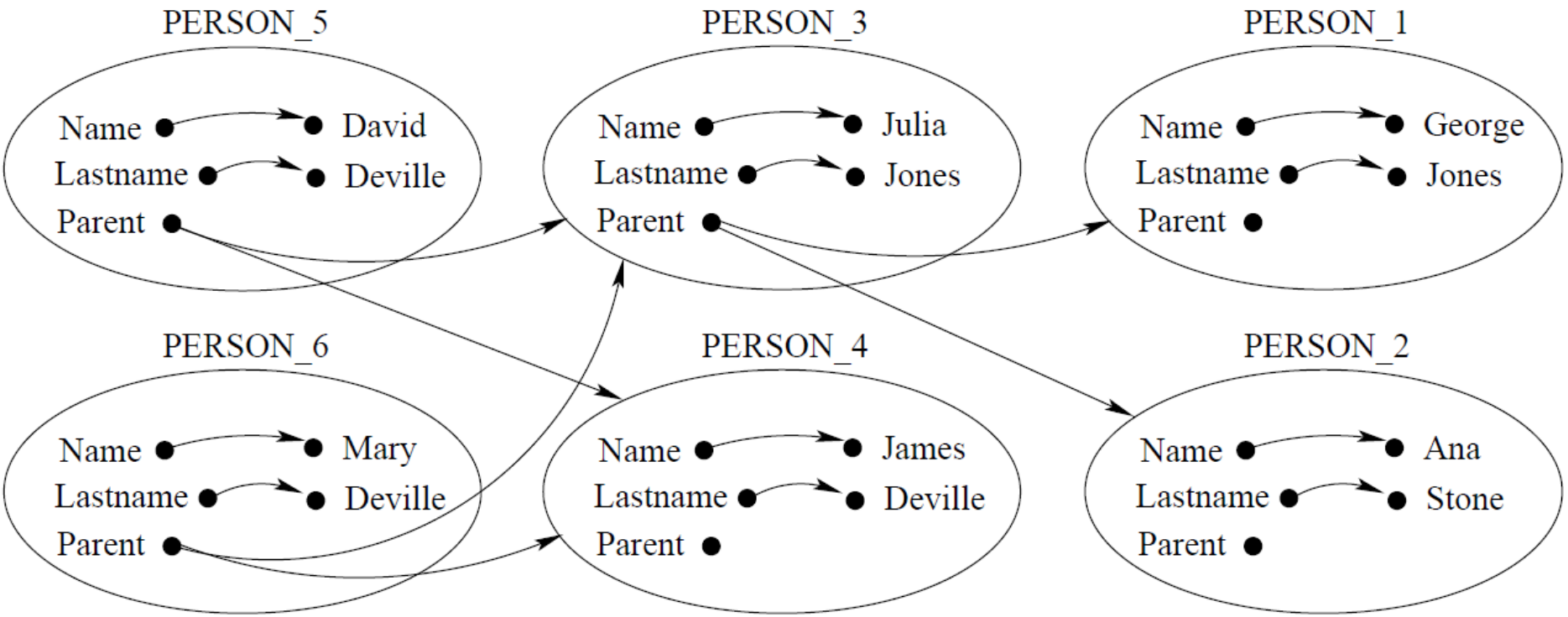
Origin in GeoNames



Geo Tree


Asia
http://www.geonames.org/6255147/
gn:parentFeature
India
http://www.geonames.org/1269750/

Hypergraph



(Jaudete Daltio, 2013)

Licença

- Estes slides são concedidos sob uma Licença Creative Commons. Sob as seguintes condições: Atribuição, Uso Não-Comercial e Compartilhamento pela mesma Licença.
- Mais detalhes sobre a referida licença Creative Commons veja no link: http://creativecommons.org/licenses/by-nc-sa/3.0/

- Fotografia de capa e fundos: web-drops por Jeremy Hiebert [http://www.flickr.com/photos/jeremyhiebert/] dispinível em http://www.flickr.com/photos/jeremyhiebert/6081389428/