WEB SÉMANTIQUE ET ONTOLOGIES

WEB DES DONNÉES

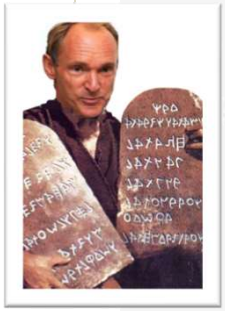
DONNÉES LIÉES (LINKED DATA)

4 – INTERROGER LES DONNÉES AVEC SPARQL

Philippe GENOUD – Danielle ZIEBELIN - LIG-Steamer

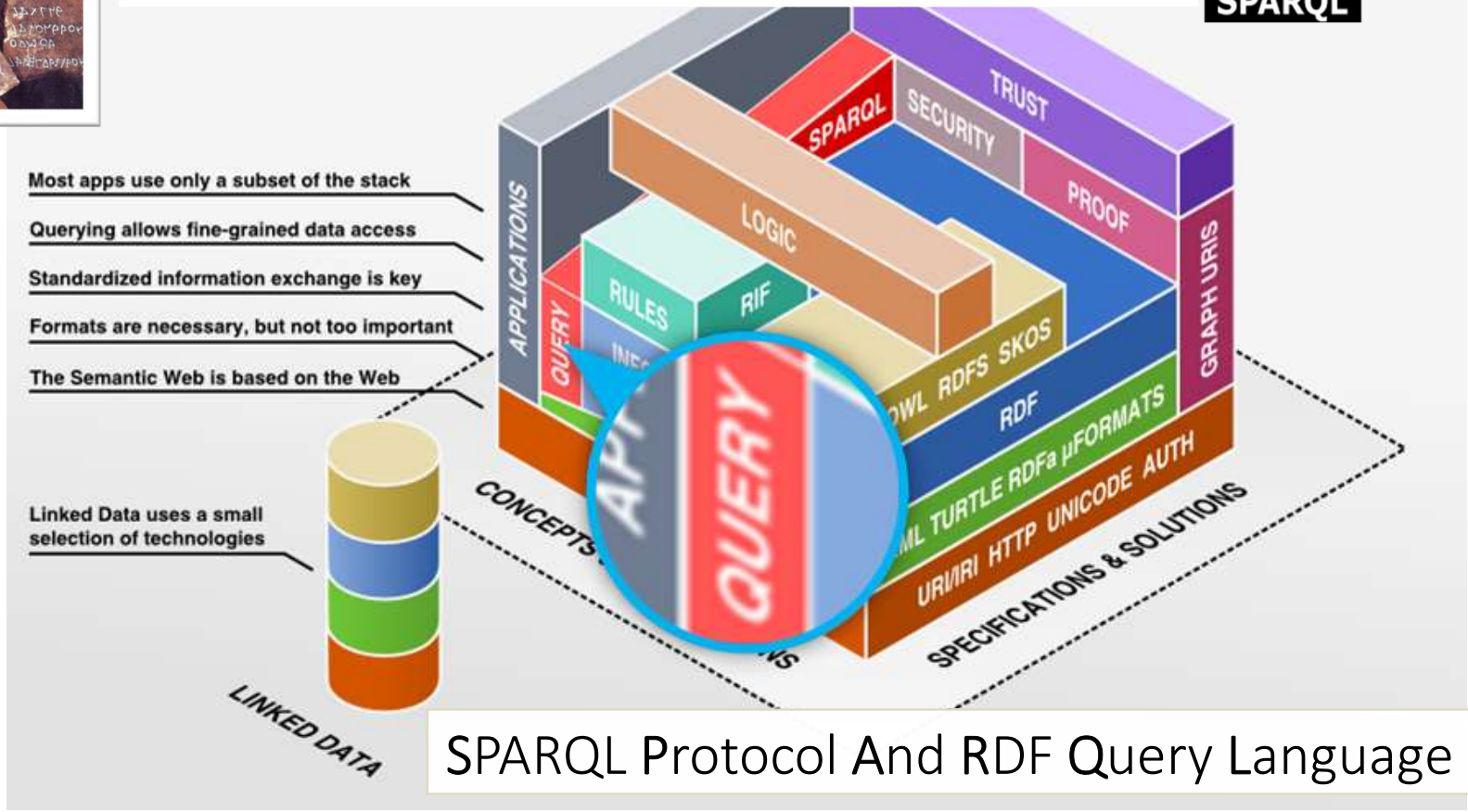
Prenom.Nom@imag.fr

Querying Linked Data with SPARQL



Linked Data: 3rd Principle

*When someone looks up a URI, provide useful information, using the standards (RDF, **SPARQL**).*

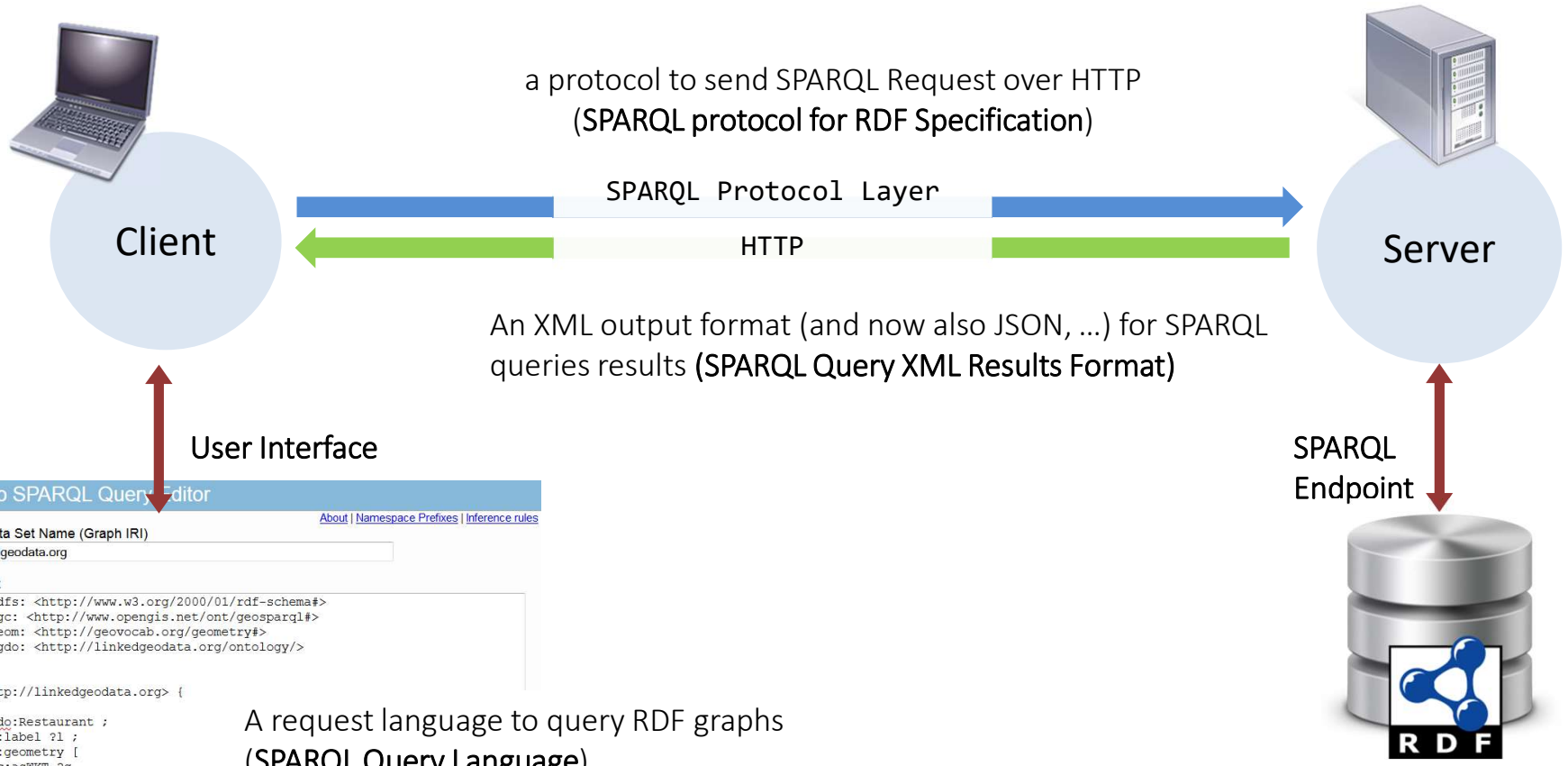


SPARQL Protocol And RDF Query Language

SPARQL : introduction

- RDF (Resource Description Framework)
  - Flexible and extensible way to represent information about resources of the web
- SPARQL (SPARQL Protocol And RDF Query Language)
  - A W3C standard
    - SPARQL 1.0 recommendation - January 2008,
    - SPARQL 1.1 recommendation – March 2013 http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/
  - a request language to access a RDF graph (SPARQL Query Language Specification) inspired from SQL
  - a protocol to submit request through HTTP GET, HTTP POST or SOAP (SPARQL protocol for RDF Specification)
  - an XML format for the results (SPARQL Query XML Results Format), and now JSON

SPARQL Protocol And RDF Query Language
overview



SPARQL endpoint

- Pour poser des requêtes: le langageSPARQL accessible par des points d'accès via un service web
  - INSEE http://rdf.insee.fr/sparql
  - Dbpedia http://fr.dbpedia.org/sparql
  - SPARQLer http://sparql.org/sparql.html General-purpose query endpoint for Web-accessible data
  - bio2rdf http://bio2rdf.org/sparql Bioinformatics data from around 40 public database

Forme la plus courante d'une requête SPARQL (SELECT) :

```
SELECT ?v1 ?v2 … ?vn
FROM <description.rdf>
WHERE {
        (sujet1 | vi) (predicat1 | vj) (objet1 | vk) .
        …
        (sujetx | va) (predicaty | vb) (objetz | vc) .
        }
```

Il existe également d'autres éléments dans le langage SPARQL qui permettent de spécifier des préfixes (PREFIX), des conditions (FILTER), des disjonctions (UNION), des filtres sur la production des résultats (LIMIT et OFFSET).

Types de requêtage dans SPARQL

4 Types de requêtage mais uniquement du requêtage à la différence de SQL

(LDD + LMD) :

SELECT : rechercher de ressources du modèle, résultats restitués sous un format tabulaire

ASK : indiquer si la requête retourne un résultat non vide (test de vacuité)

DESCRIBE : obtenir des informations à propos de ressources présentes dans le modèle (le moins exploité des quatre)

CONSTRUCT : construire de nouveaux graphes RDF à partir des résultats de la requête servant de ”template”

Forme générale d'une requête SPARQL

ici requête SELECT :

Prologue

BASE prefix :<namespace-uri >

PREFIX prefix :<namespace-uri >

Head

SELECT [DISTINCT] variable-list

Body

FROM source-list

WHERE pattern

ORDER BY expression

LIMIT integer > 0

OFFSET integer > 0

Prologue d'une requête SPARQL

BASE suivi d'un raccourci pour une URI de base auxquelles les URIs seront concaténées.

Une seule base par requête.

Exemple : BASE <http ://www.lis.univ-amu/data/>

PREFIX suivi d'une déclaration d'un raccourci pour des espaces de nom.

Il est courant d'avoir plusieurs préfixes.

Exemple : PREFIX fb :<http ://rdf.freebase.com/ns/>

Head d'une requête SPARQL

SELECT suivi d'une liste de variables dont on souhaite connaître les valeurs répondant à la requête.

Note : la requête peut utiliser d'autres variables.

CONSTRUCT suivi d'un template de triplets pour construire un entrepôt des triplets répondant à la requête.

ASK pour tester s'il existe des solutions vérifiant les conditions de la requête.

Body d'une requête SPARQL

Peut contenir les clauses :

FROM (optionnel) permet de spécifier différents graphes de triplets dans lesquels il faut chercher. Si pas mentionné, on suppose qu'un graphe a été spécifié au processeur SPARQL

WHERE suivi d'une déclaration de pattern (motif) pouvant contenir : plusieurs patterns, plusieurs patterns de base, des filtres, des unions, des présences optionnelles

ORDER BY trier par …

LIMIT pour donner un nombre maximal de réponses

OFFSET pour commencer à partir d'un numéro de réponse

Les contraintes OPTIONAL et FILTER

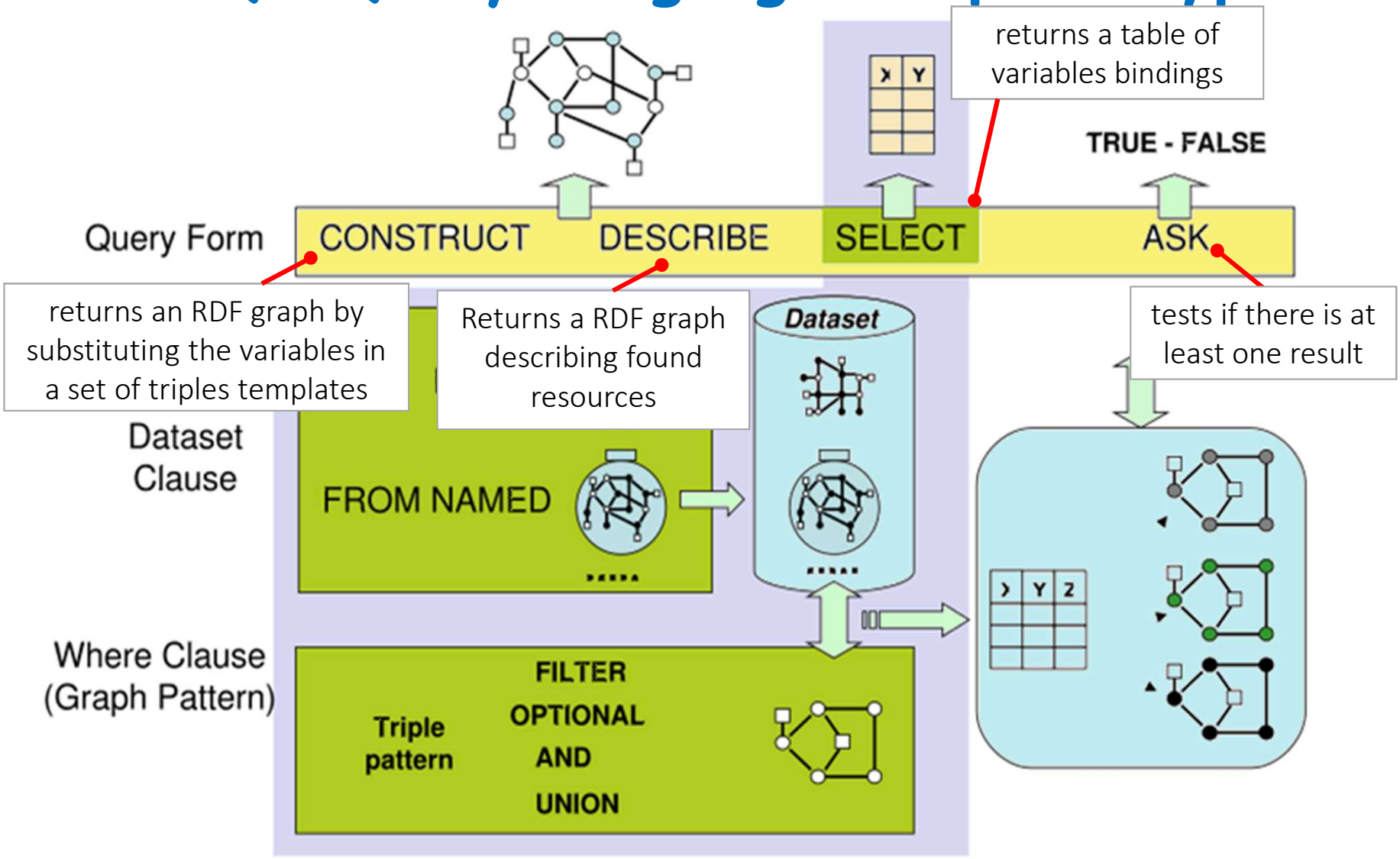
Remarques :

pattern = ensemble de triplets contenant des noms de variables

les triplets sont séparés par le symbole « . »

un tuple de variables satisfait le pattern si tous les triplets sont satisfaits par le tuple : on parle de requête conjonctive

SPARQL Query Language: request types

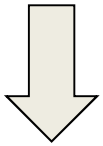


from: Pérez, Arenas and Gutierrez, Chapter 1: On the Semantics of SPARQL, Semantic Web Information Management: A Model Based Perspective, Springer 2010

SPARQL Query Language
ASK query

- ASK – ask a boolean query.
  - to verify that there is at least one response.
  - Is there a student over 30 years?

```
PREFIX ufrimag: <http://www.ufrimag.fr/onto#>
ASK {
 ?etudiant ufrimag:inscrit ?x .
           ufrimag:age ?age .
 FILTER (?age > 30)
}
```



the boolean result

```
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
 <head> … </head>
 <boolean> true </boolean>
</sparql>
```

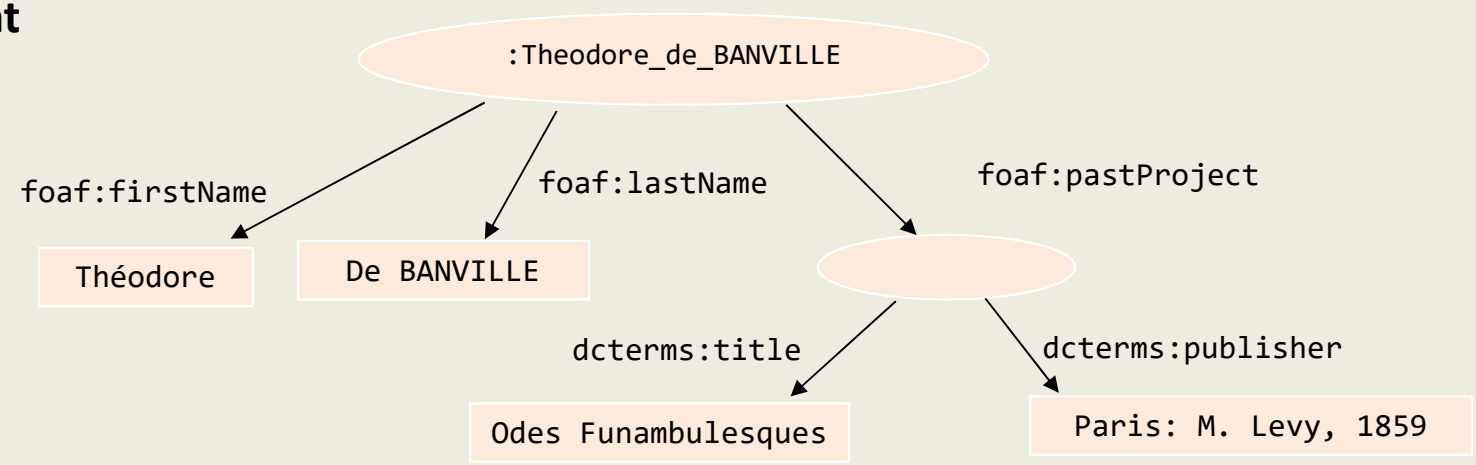
SPARQL Query Language
CONSTRUCT query

- SELECT returns a flat list of variables bindings
  - the application program is in charge of processing these bindings (sometimes by converting solution tuples into triples and adding them to a RDF graph)

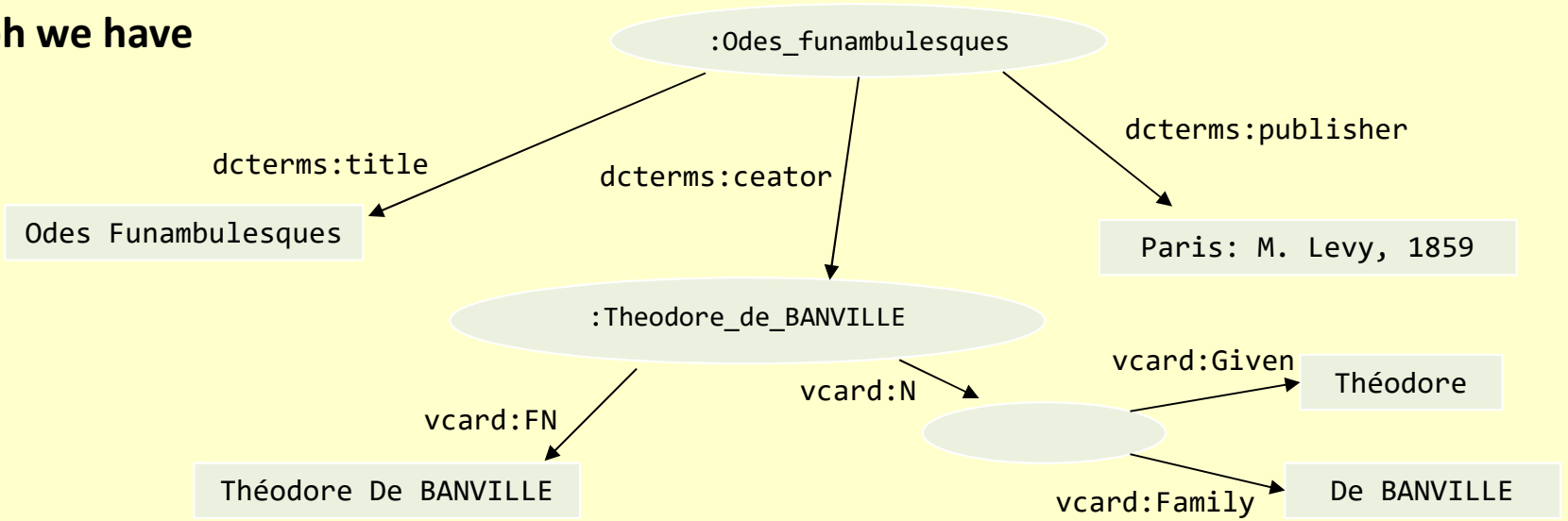
- CONSTRUCT allows you to directly product a RDF graph containing the variables values
  - the WHERE and FILTER clause works the same way as the SELECT form
  - bindings of the variables are inserted into a new graph constructed from template triples specified in the CONSTRUCT clause (which replace the SELECT clause).

SPARQL Query Language
CONSTRUCT query

The graph we want



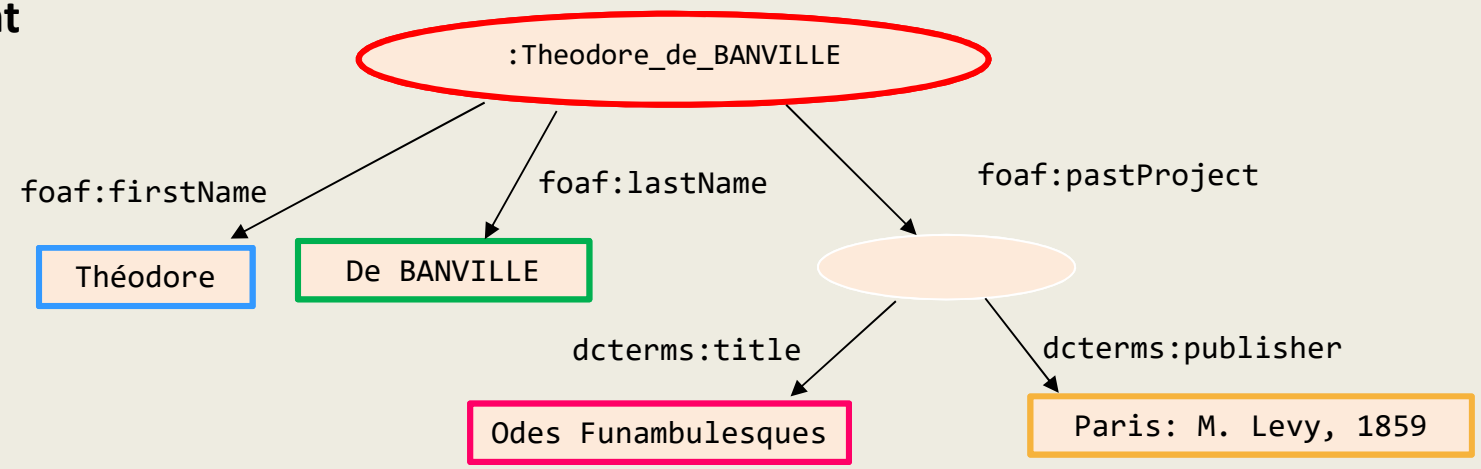
The graph we have



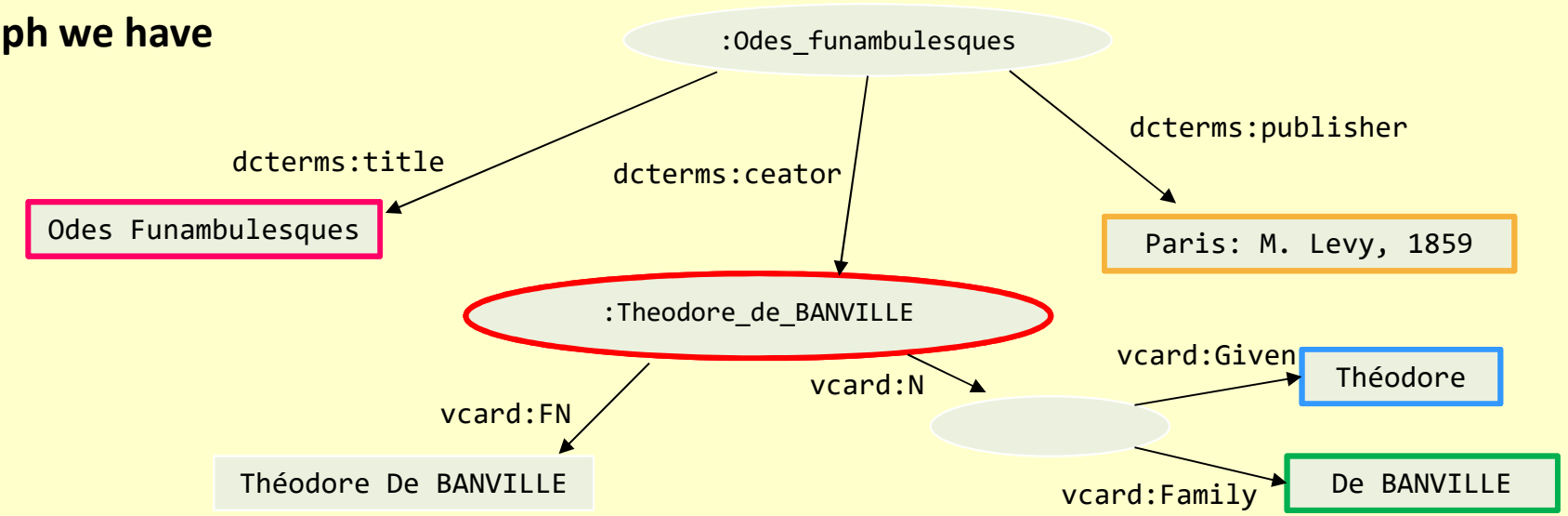
D'après : http://pagesperso-systeme.lip6.fr/Jean-Francois.Perrot/inalco/XML/RDF/SPARQL/IntroSPARQL.html

SPARQL Query Language
CONSTRUCT query

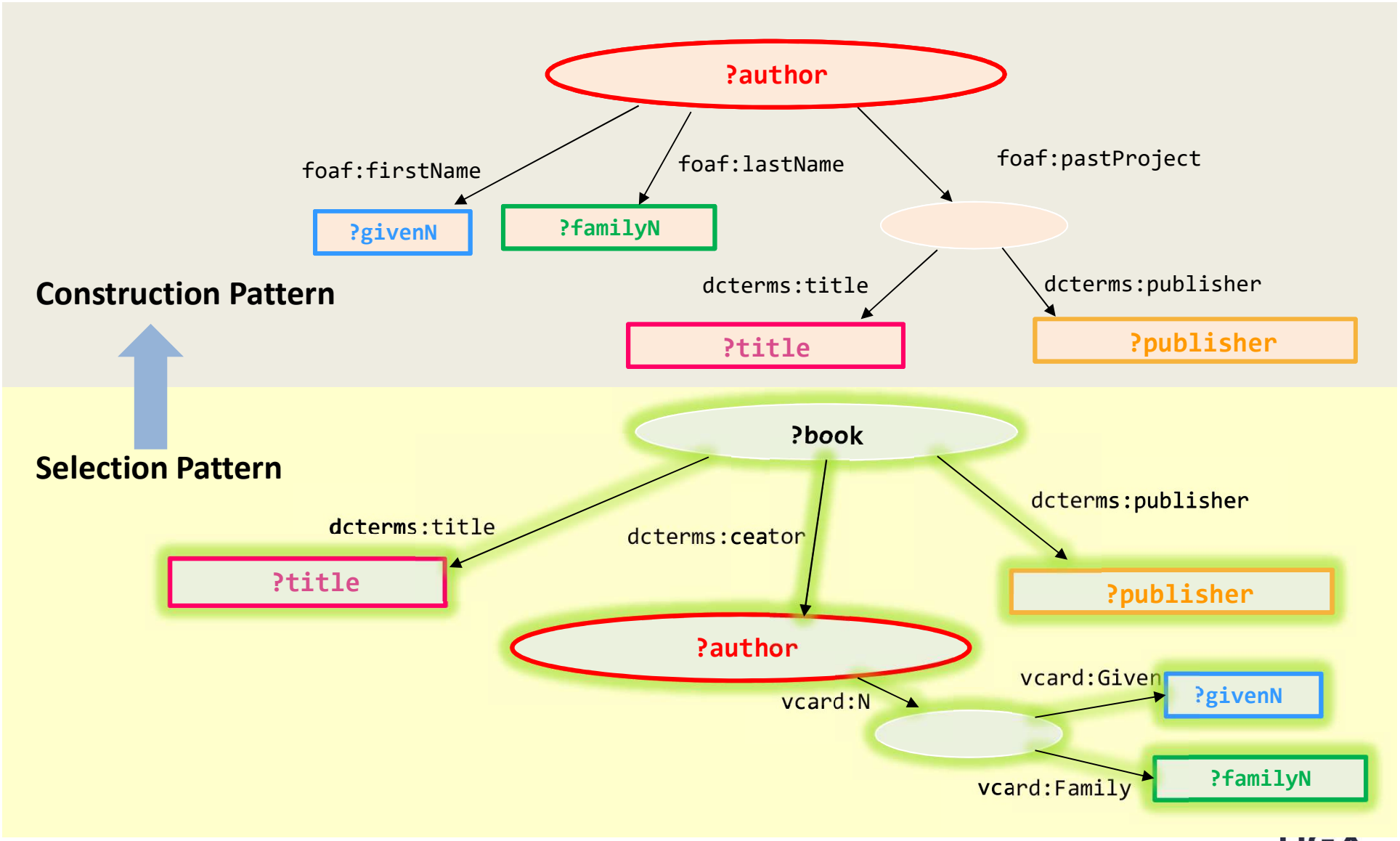
The graph we want



The graph we have



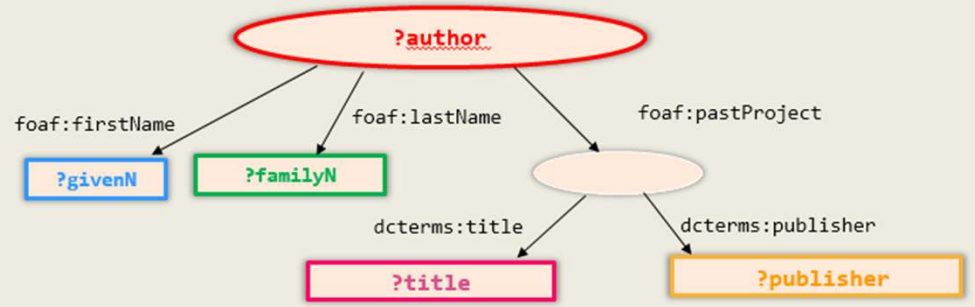
SPARQL Query Language
CONSTRUCT query



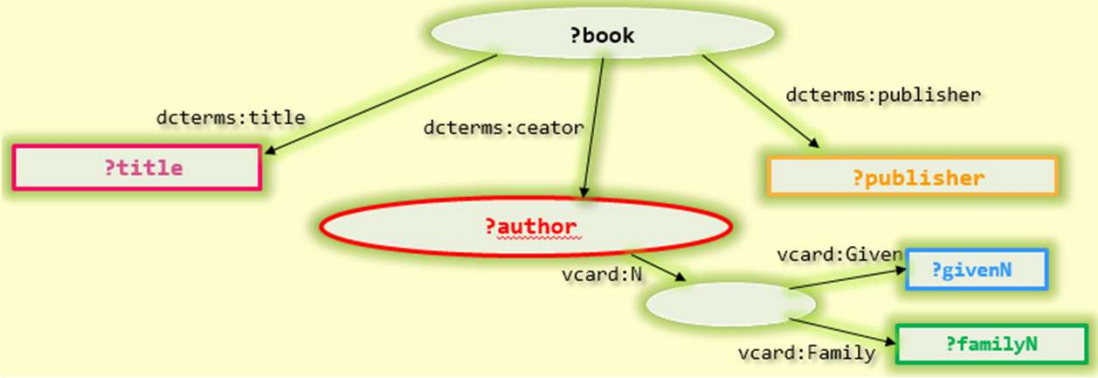
SPARQL Query Language
CONSTRUCT query

```
PREFIX foaf:    <http://xmlns.com/foaf/0.1/>
PREFIX dc:      <http://purl.org/dc/elements/1.1/>
PREFIX vcard:   <http://www.w3.org/2001/vcard-rdf/3.0#>
```

```
CONSTRUCT {
  ?author foaf:firstName ?givenN;
          foaf:lastName  ?familyN;
          foaf:pastProject [
              dc:title ?title;
              dc:publisher ?publisher
          ]
}
```



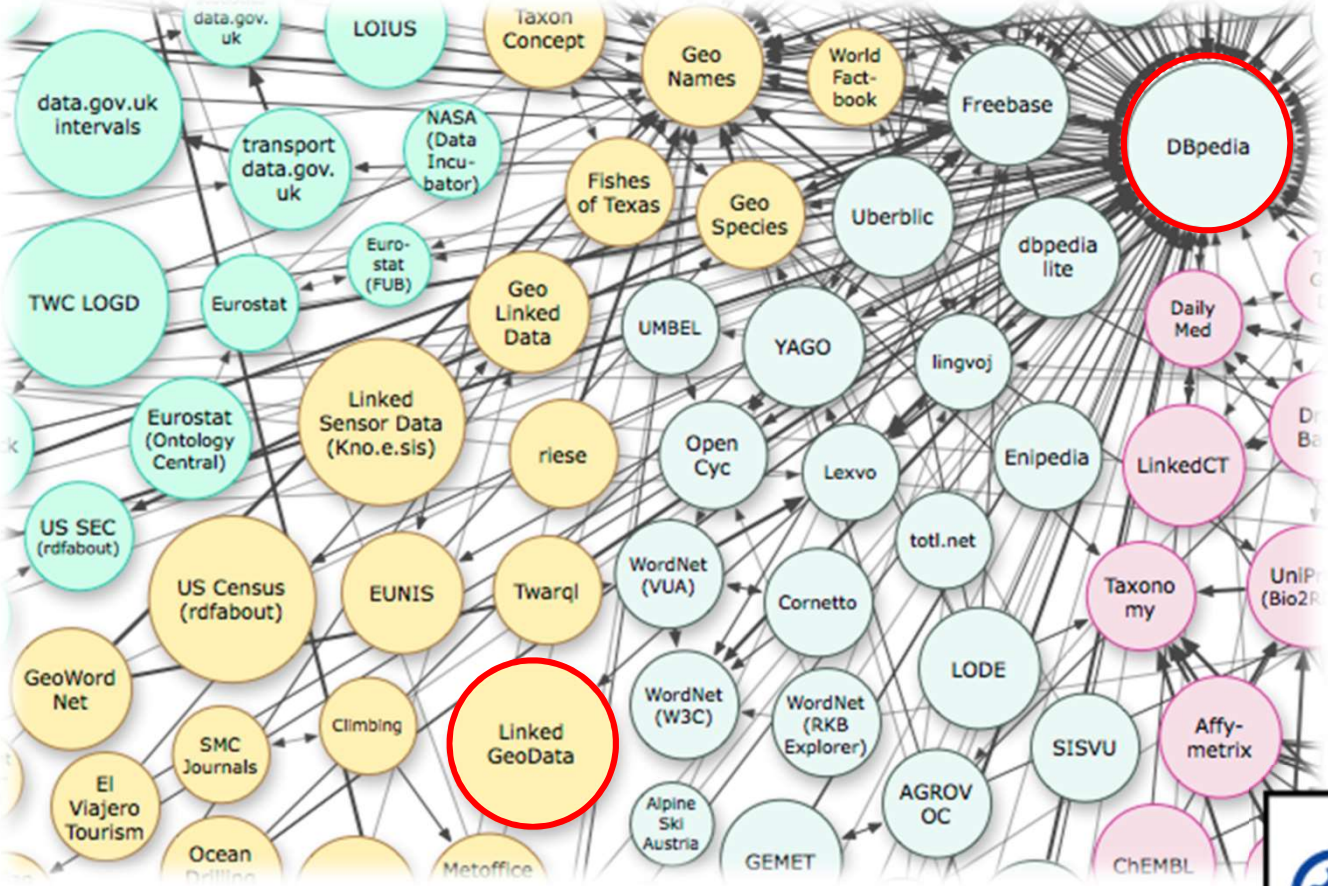

```
WHERE {
  ?book dc:title ?title;
        dc:creator ?author;
        dc:publisher ?publisher .
  ?author vcard:N  [
        vcard:Given ?givenN;
        vcard:Family ?familyN
      ] .
 }
```



SPARQL Query Language
DESCRIBE query

- `DESCRIBE` – Returns an RDF graph, based on what the query processor is configured to return.
  - SPARQL specification says : "the useful information the service has about a resource"
  - in theory this should help you understand the context of the resources returned… but there is no warranty.

SPARQL in action with LinkedGeoData



SPARQL Query Language: SELECT

- find all the restaurants that are less than 1km from Fort Saint-Pierre (Théâtre de la Mer – Sète)

```
Prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix ogc:  <http://www.opengis.net/ont/geosparql#>
Prefix geom: <http://geovocab.org/geometry#>
Prefix lgdo: <http://linkedgeodata.org/ontology/>

Select ?s, ?l From <http://linkedgeodata.org>
Where
{
     ?s    a lgdo:Restaurant ;
           rdfs:label ?l ;
           geom:geometry [
              ogc:asWKT ?g
           ] .

     Filter(bif:st_intersects (?g, bif:st_point (3.692764, 43.393794), 1)) .
}
```

SPARQL Query Language: SELECT

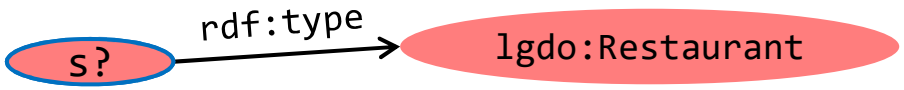
- SPARQL based on :
  - RDF serialization with Turtle
  - Graph Pattern Matching

```
Prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix ogc:  <http://www.opengis.net/ont/geosparql#>
Prefix geom: <http://geovocab.org/geometry#>
Prefix lgdo: <http://linkedgeodata.org/ontology/>

Select ?s, ?l From <http://linkedgeodata.org>
Where
{
      ?s    a lgdo:Restaurant ;
            rdfs:label ?l ;
            geom:geometry [
               ogc:asWKT ?g
            ] .

      Filter(bif:st_intersects (?g, bif:st_point (3.692764, 43.393794), 1)) .
}
```

variable



Graph Pattern: RDF triple containing one or more variables at any position (subject, predicate, object)

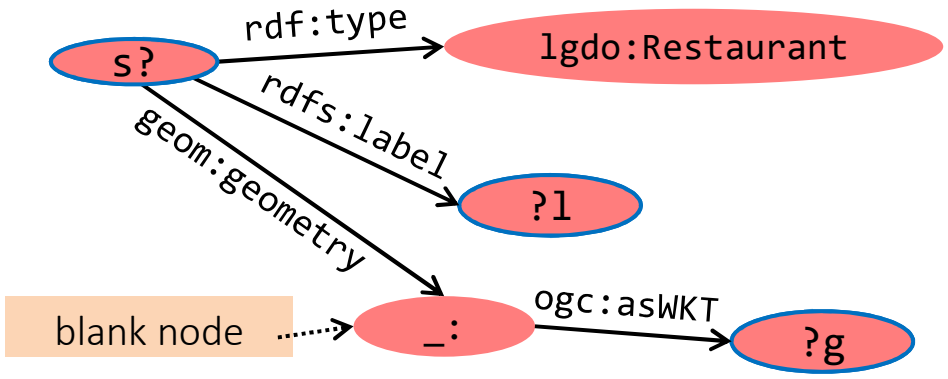
SPARQL Query Language
Select

- Graph patterns can be combined to construct complex (conjunctive) requests.

```
Prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix ogc:  <http://www.opengis.net/ont/geosparql#>
Prefix geom: <http://geovocab.org/geometry#>
Prefix lgdo: <http://linkedgeodata.org/ontology/>

Select ?s, ?l From <http://linkedgeodata.org>
Where
{
      ?s    a lgdo:Restaurant ;
            rdfs:label ?l ;
            geom:geometry [
                ogc:asWKT ?g
            ] .
      Filter(bif:st_intersects (?g, bif:st_point (3.692764, 43.393794), 1)) .
}
```

variable



SPARQL Query Language
Select - Filters

- Possibility to filter results
  - A filter allows to constrain solution values

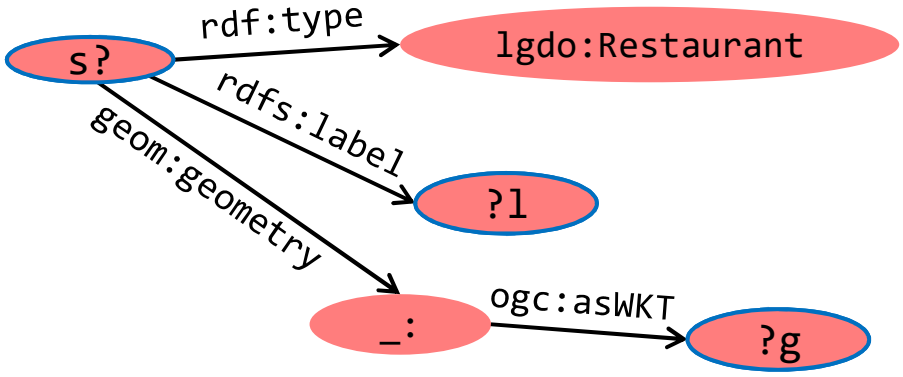
```
Prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix ogc:  <http://www.opengis.net/ont/geosparql#>
Prefix geom: <http://geovocab.org/geometry#>
Prefix lgdo: <http://linkedgeodata.org/ontology/>

Select ?s, ?l From <http://linkedgeodata.org>
Where
{
      ?s    a lgdo:Restaurant ;
            rdfs:label ?l ;
            geom:geometry [
               ogc:asWKT ?g
            ] .

      Filter(bif:st_intersects (?g, bif:st_point (3.692764, 43.393794), 1)) .
}
```

Filter on `?g` variable

Graph Pattern



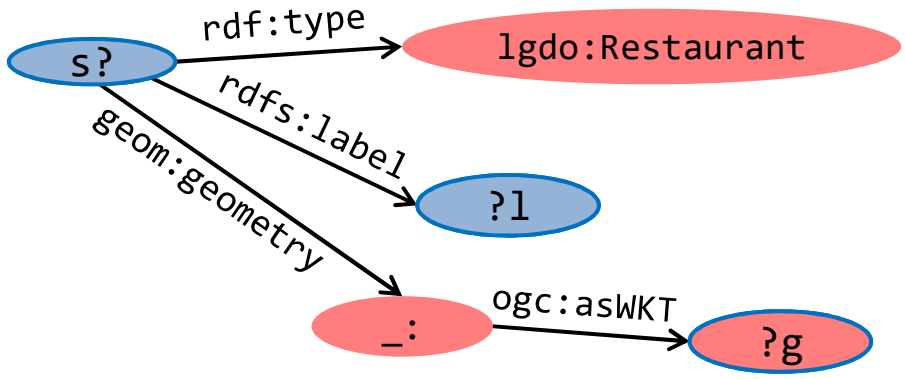
SPARQL Query Language
Select

- The SELECT clause indicates which variables to consider in the result
- * all the variables (like SQL)

```
Prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix ogc:  <http://www.opengis.net/ont/geosparql#>
Prefix geom: <http://geovocab.org/geometry#>
Prefix lgdo: <http://linkedgeodata.org/ontology/>

Select ?s, ?l From <http://linkedgeodata.org>
Where
{
     ?s    a lgdo:Restaurant ;
           rdfs:label ?l ;
           geom:geometry [
              ogc:asWKT ?g
           ] .

     Filter(bif:st_intersects (?g, bif:st_point (3.692764, 43.393794), 1)) .
}
```



SPARQL Query Language
Select clause - DataSets

- The RDF data-store service can handle one or more RDF graphs, the SPARQL query is executed against a data set (RDF Dataset) that represents a collection of one or more graphs.

```
Prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix ogc:  <http://www.opengis.net/ont/geosparql#>
Prefix geom: <http://geovocab.org/geometry#>
Prefix lgdo: <http://linkedgeodata.org/ontology/>

Select ?s, ?l From <http://linkedgeodata.org>
Where
{
     ?s    a lgdo:Restaurant ;
           rdfs:label ?l ;
           geom:geometry [
              ogc:asWKT ?g
           ] .

     Filter(bif:st_intersects (?g, bif:st_point (3.692764, 43.393794), 1)) .
}
```

graph on which the query is executed

SPARQL Query Language
Select - DataSets

- a SPARQL request can associate different graph patterns to different named graphs.
  - FROM defines the default graph (this may be merging several graphs)
  - FROM NAMED defines graphs which then can be explicitly named in the WHERE part of the query through the GRAPH keyword

```
PREFIX p1: ...
SELECT ...
FROM p1:g1
FROM p2:g2
FROM NAMED p1:g3
FROM NAMED p1:g4
WHERE {
   ......
   GRAPH p1:g3 {
        ......
   }
   GRAPH ?g {
     ......
   }
}
```