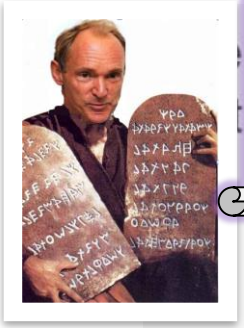
WEB SÉMANTIQUE ET ONTOLOGIES
WEB DES DONNÉES
DONNÉES LIÉES (LINKED DATA)

3 – LE MODÈLE DE DONNEES RDF (RESOURCE DESCRIPTION FRAMEWORK)

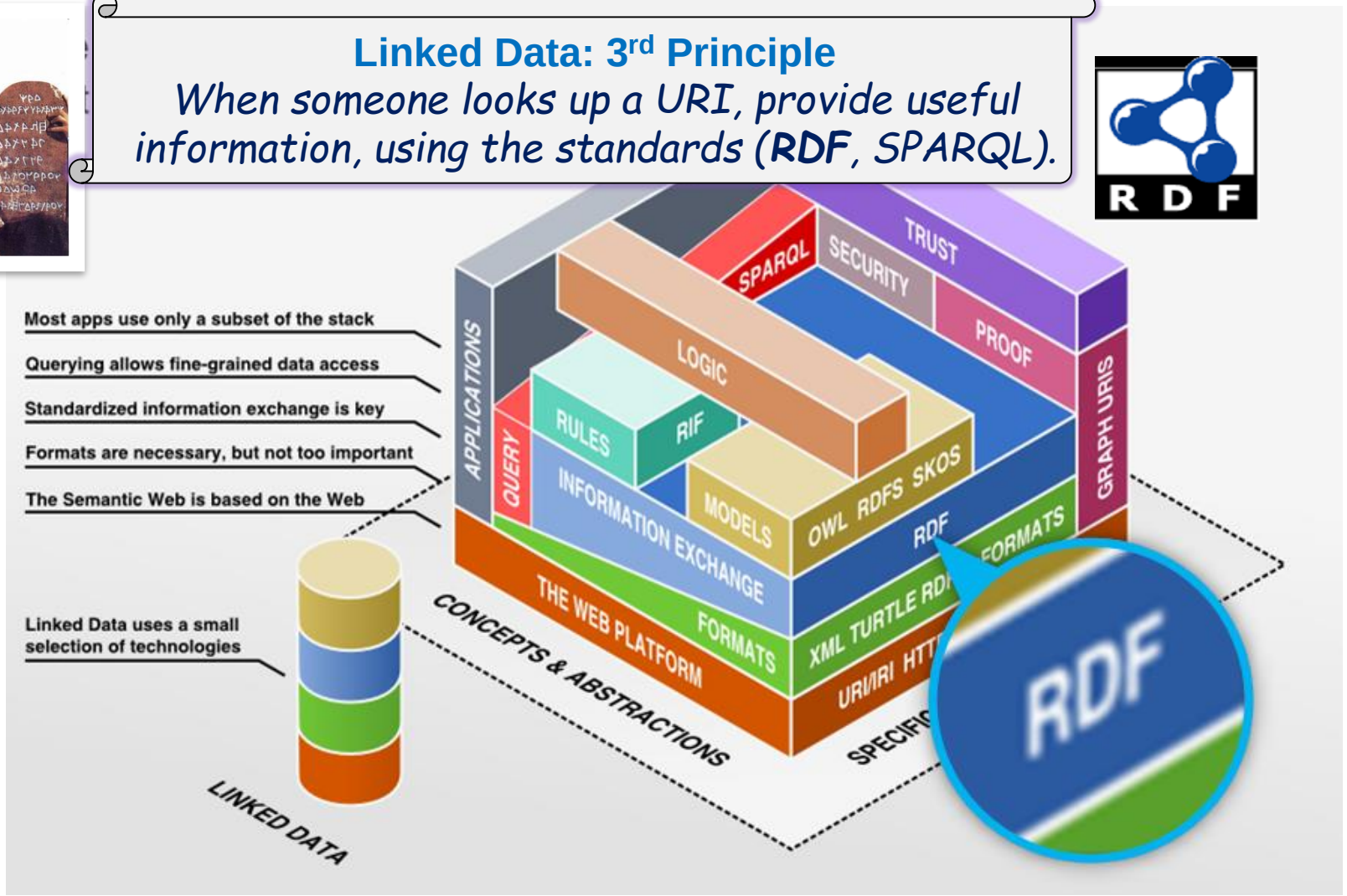
Philippe GENOUD – Danielle ZIEBELIN - LIG-Steamer
Prénom.Nom@imag.fr

Resource Description Framework (RDF)



Linked Data: 3rd Principle

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



(Crédit Benjamin Nowack)

Resource Description Framework

- RDF Resource Description Framework
  - a framework for describing resources on the web
    - "The Resource Description Framework (RDF) is a framework for representing information in the Web." [1]
  - is designed to be read and understood by computers
  - RDF is a part of the W3C's Semantic Web Activity
    - became a W3C recommendation 10. February 2004
    - Updated February 2014 (RDF 1.1)

http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ [1]
http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/ [2]

"The Resource Description Framework (RDF) is a framework for representing information in the Web." [1]

Resource ?

- *"The Resource Description Framework (RDF) is a language for* ***representing information about resources in the World Wide Web****. It is particularly intended for representing metadata about Web resources, such as the title, author, and modification date of a Web page, copyright and licensing information about a Web document, or the availability schedule for some shared resource.* ***However, by generalizing the concept of a "Web resource", RDF can also be used to represent information about things that can be identified on the Web, even when they cannot be directly retrieved on the Web****."*

  http://www.w3.org/TR/rdf-primer/

- *"To publish data on the Web, the* ***items in a domain of interest must first be identified****. These are the things whose properties and relationships will be described in the data, and may include Web documents as well as real-world entities and abstract concepts. As Linked Data builds directly on Web architecture , the Web architecture term* ***resource*** *is used to refer* ***to these things of interest****, which are, in turn, identified by HTTP URIs."*

  Tom Heath, Christian Bizer *: Linked Data: Evolving the Web into a Global DataSpace*

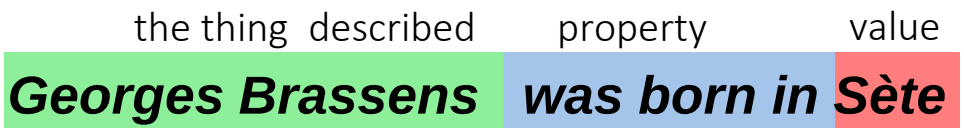
  http://linkeddatabook.com/editions/1.0/

RDF outline

- **RDF Data Model**
- RDF formats
- Blank nodes
- Typed literals
- Resources definition
- RDF and data integration
- Persisting RDF
- References

RDF Data Model

- With RDF, knowledge is represented by a set of assertions (statements)
- All RDF statements follow a simple structure composed of three parts :
  - the thing the statement describes
  - the properties of the thing the statement describes
  - the values of those properties the statement describe

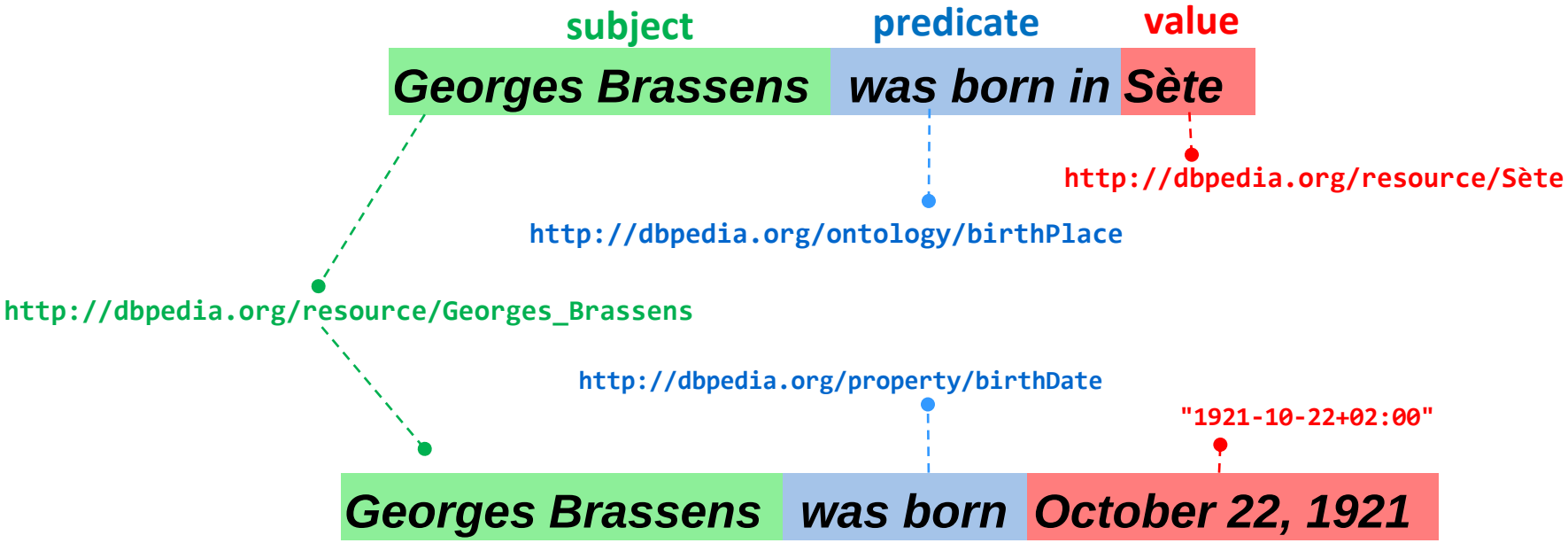


RDF Data Model

- RDF Statements are *triples*

Subject	Predicate	Object
URI	URI	URI/Literal

- the subject and the predicate are resources : RDF uses **URIs** (Universal Resource Identifiers) for **uniquely identifying** them
- object can be a **resource** (URI) or a **literal** (constants that don't have other attributes that describe them)

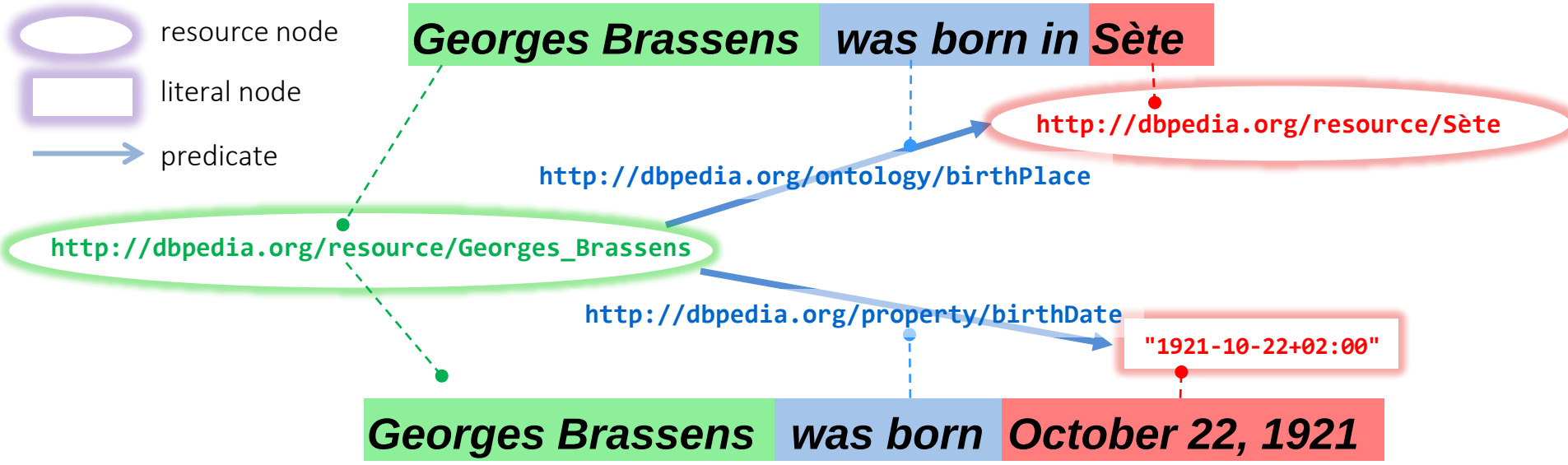


RDF Data Model

- RDF Statements are *triples*

Subject	Predicate	Object
URI	URI	URI/Literal

- RDF data can be viewed as a directed labeled graph
  - subjects and objects are nodes (vertices)
  - predicates are oriented edges (arcs)

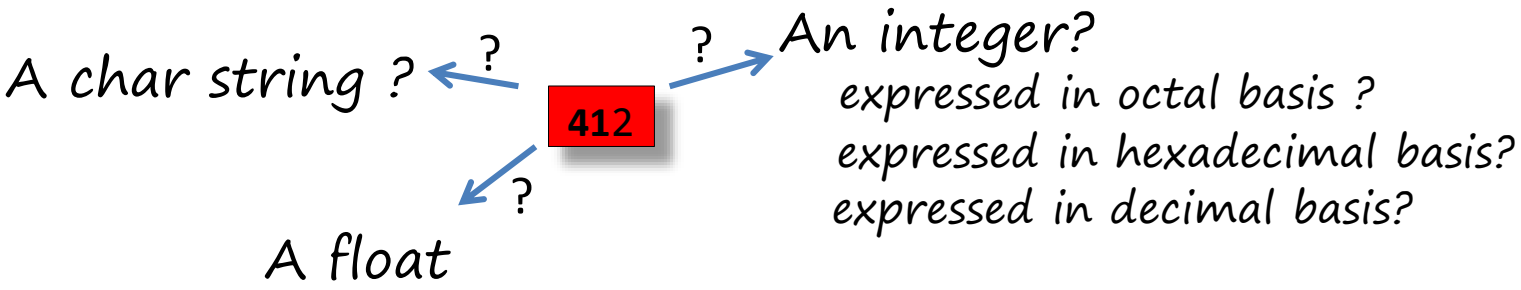


Typed Literals

- Literal are not resources : they are values

When looking at that description a human can easily realize that 412 is an integer.

But what about a computer program ?



You must provide **some context** if you intend to use the value in any other way than to just view it on a web page **typed literals**

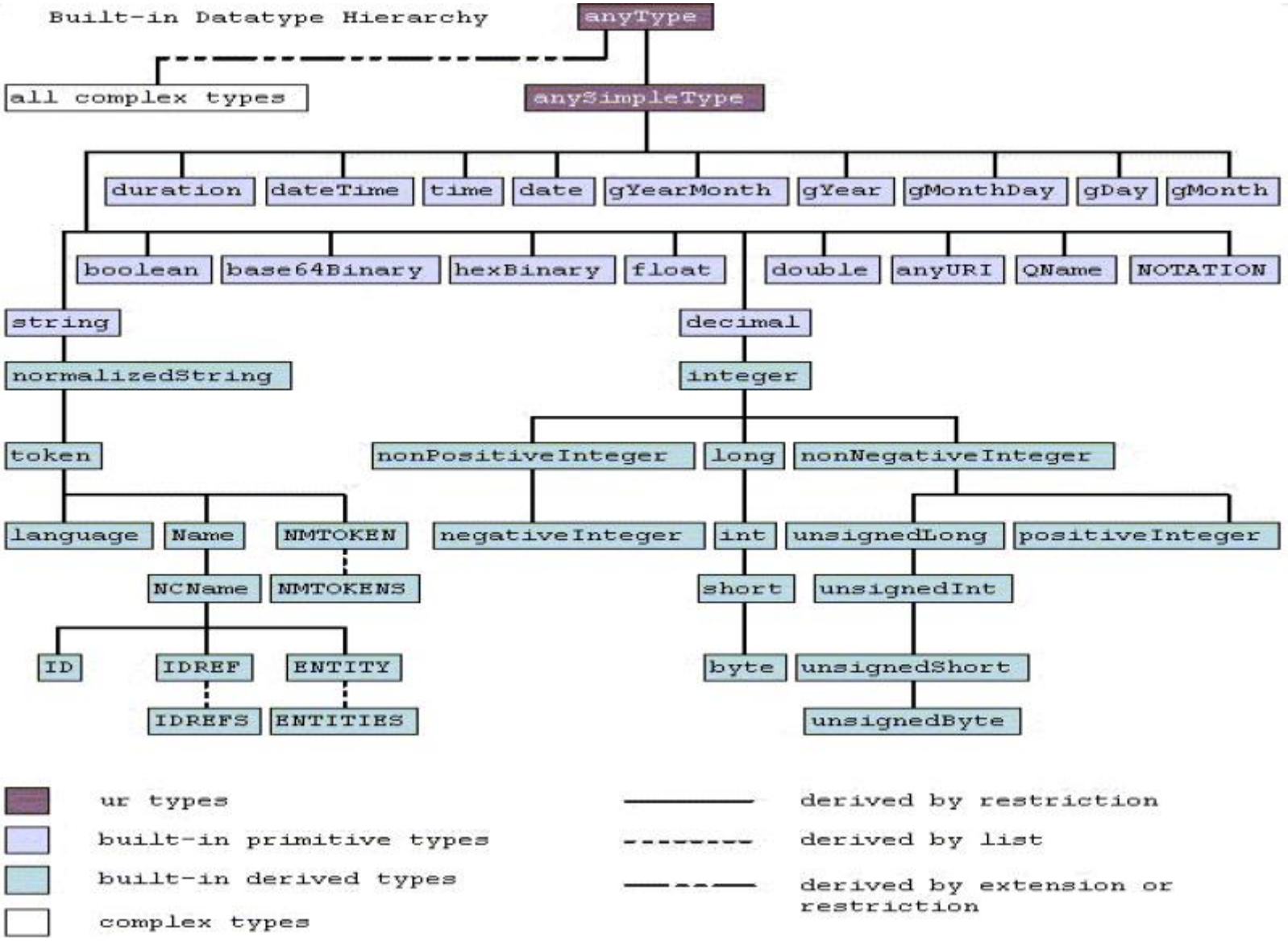
Typed Literals

- How to define a type (datatype) ?
  - **value space** : the set of values represented by the type
    - - e.g. an integer interval , dates, ….
  - **lexical space** : the set of char strings defining the representations of the values
    - eg. dates : `yy-mm-dd` or `dd-mm-yy`
  - a **mapping** between the lexical space and the value space
    - associating a concrete value with each eligible literal

Typed Literals : XSD

- To associate type to literals RDF uses XSD (**X**ML **S**chema **D**efinition)
  - W3C recommendations :
    - http://www.w3.org/TR/xmlschema-2/
    - http://www.w3.org/TR/rdf-mt/
  - XSD defines a predefined datatype hierarchy (see next slide)
    - primitive types (string, float, decimal, etc.)
    - derived types (integer, long, etc.)
  - new types can be defined by derivation
    - restriction
    - lists
    - union
    - extension

Typed Literals: XSD



Typed Literals: XSD

- examples of definition of new data types

new type derived by **restrictions**

```
<xsd:schema ...>
  <xsd:simpleType name="humanAge">
    <xsd:restriction base="integer">
     <xsd:minInclusive value="0">
     <xsd:maxExclusive value="150">
    </xsd:restriction>
  </xsd:simpleType>
  ...
</xsd:schema>
```

the "super" type

constraints to express the restriction

new type derived by **list**

```
<simpleType name="listOfFloat">
  <list itemType="float"/>
</simpleType>
```

type of the list elements

+ constraints about the list
`length, maxLength, minLength`

Typed Literals : XSD

new type derived by **union** and **extension**

```
<xsd:simpleType name="fontsize">
  <xsd:union>
    <xsd:simpleType>
       <xsd:restriction  base="xsd:positiveInteger">
         <xsd:minInclusive  value="8"/>
         <xsd:maxInclusive  value="72"/>
       </xsd:restriction>
    </xsd:simpleType>
    <xsd:simpleType>
       <xsd:restriction  base="xsd:NMTOKEN">
         <xsd:enumeration  value="small"/>
         <xsd:enumeration  value="medium"/>
         <xsd:enumeration  value="large"/>
       </xsd:restriction>
    </xsd:simpleType>
  </xsd:union>
</xsd:simpleType>
```

type defined by union

type defined by restrictions

type defined by extension

fontsize : `8-72` or `small`, `medium`, `large`

Typed Literals

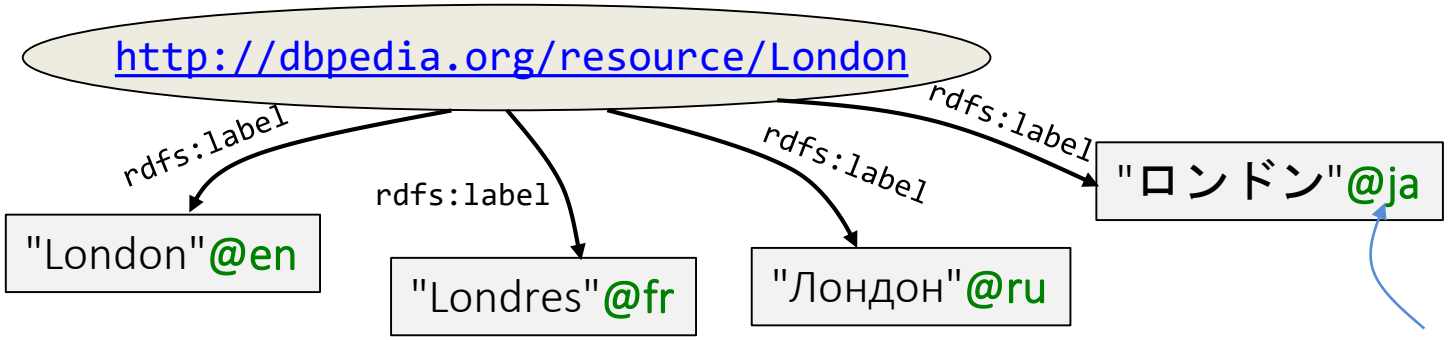
- Turtle : **literalvalue^^dataTypeURI**

```
@prefix dbo:   <http://dbpedia.org/ontology/> .
@prefix dbr:   <http://dbpedia.org/resource/> .
@prefix ns9:   <http://dbpedia.org/ontology/PopulatedPlace/> .
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .

dbr:Grenoble  a                    dbo:Place , <http://schema.org/Place> ,
        rdfs:comment               "Grenoble (IPA: gʁə.ˈnɔbl ; Grenoblo in ..."
        rdfs:label                 "Grenoble"@fr , "Grenoble"@pl , "
        dbo:abstract               "Grenoble (prononcé [gʁə.ˈnɔbl ] ,..."
        ns9:area                   "18.44"^^<http://dbpedia.org/datatype/squareKilometre> ;
        dbo:country                dbr:France ;
        dbo:department             dbr:Isère ;
        dbo:elevation              "398.0"^^xsd:double ;
        dbo:inseeCode              "38185" ;
        dbo:intercommunality       dbr:Grenoble-Alpes_Métropole ;
        dbo:maximumElevation       "500.0"^^xsd:double ;
        dbo:mayor                  dbr:Éric_Piolle ;
        dbo:minimumElevation       "212.0"^^xsd:double ;
        dbo:populationTotal        "156659"^^xsd:nonNegativeInteger ;
        dbo:postalCode             "38000, 38100" ;
        dbo:region                 dbr:Rhône-Alpes ;
        ...
```

Localized textual literals

- A language can be associated to textual literals



étiquette d'identification de langues(standard BCP47 de l'IETF)

RDF/XML

```
<rdf:RDF (…)>
<rdf:Description rdf:about="http://dbpedia.org/resource/London">
      <rdfs:label xml:lang='en'>London</rdfs:label>
      <rdfs:label xml:lang='fr'>Londres</rdfs:label>
      ...
</rdf:Description>
</rdf:RDF>
```

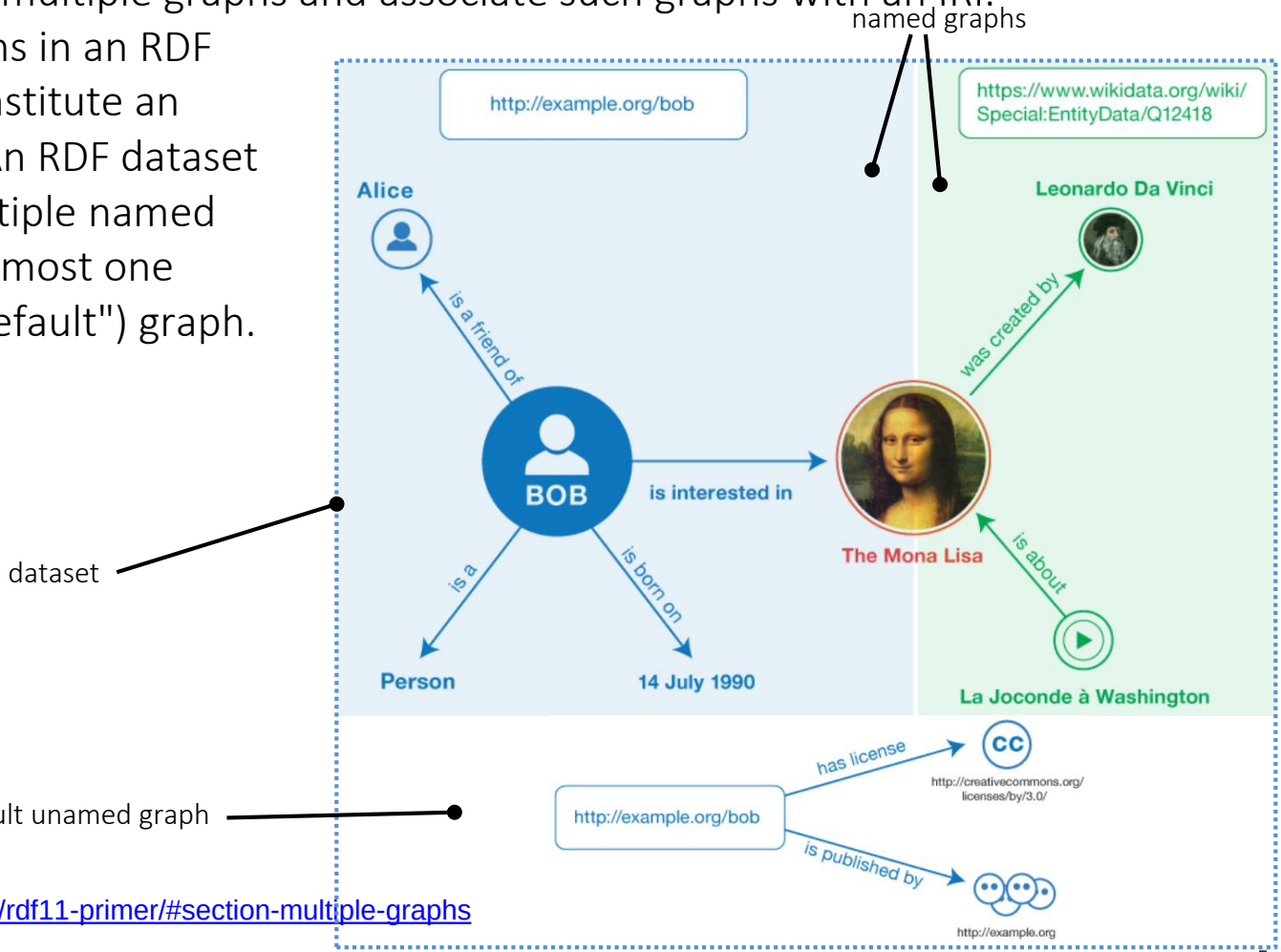
Turtle

```
@prefix rdfs:< http://www.w3.org/2000/01/rdf-schema#>.

<http://dbpedia.org/resource/London>
       rdfs:label "Londres"@fr, ... , "London"@en.
```

RDF Datasets

- In practice, RDF tool builders and data managers needed a mechanism to talk about subsets of a collection of triples RDF 1.1 a mechanism to group RDF statements in multiple graphs and associate such graphs with an IRI.
- Multiple graphs in an RDF document constitute an **RDF dataset**. An RDF dataset may have multiple named graphs and at most one unnamed ("default") graph.



https://www.w3.org/TR/rdf11-primer/#section-multiple-graphs

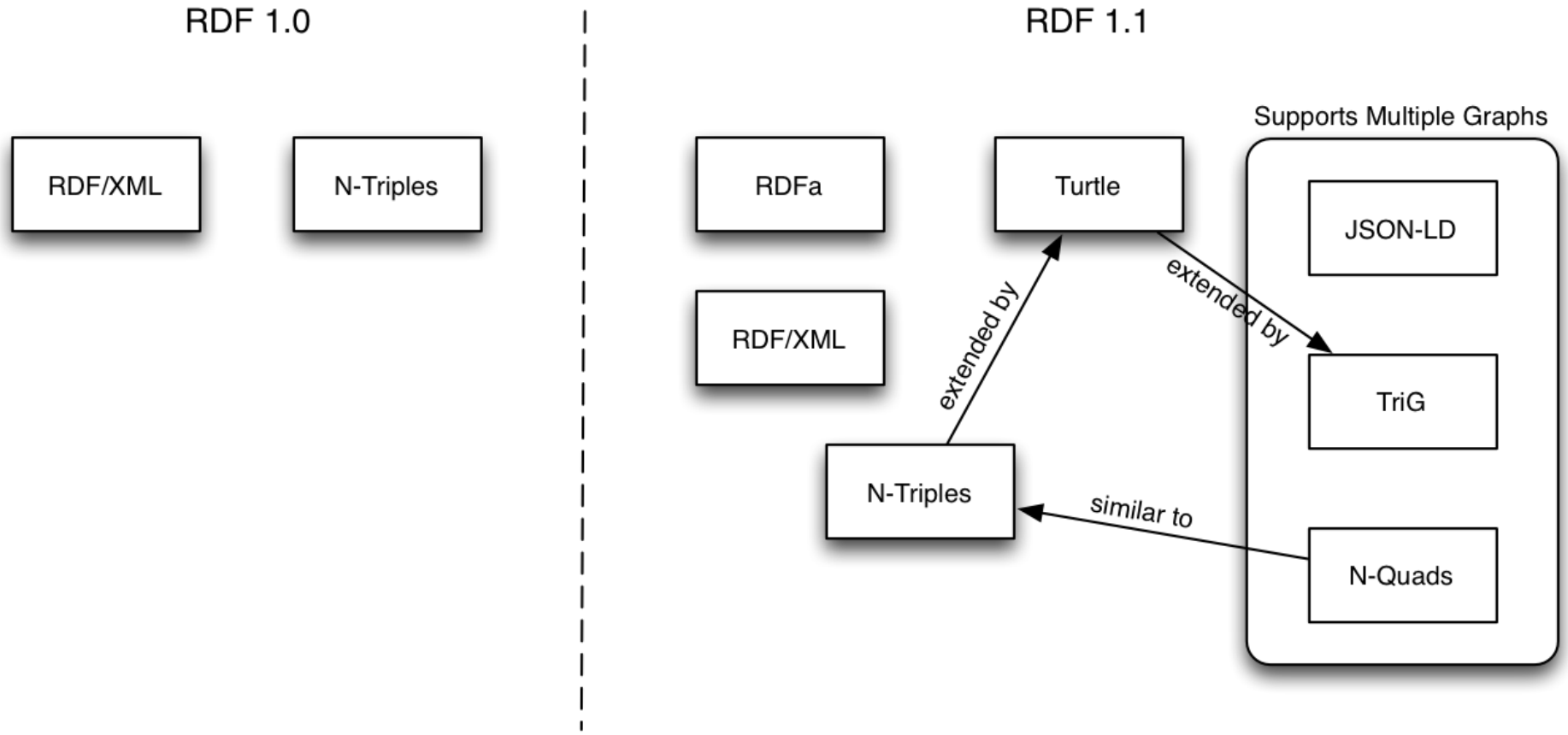
RDF outline

- RDF Model
- Typed and localized literals
- RDF formats
- Blank nodes
- Resources definition
- RDF and data integration
- Persisting RDF
- References

RDF Serializations

- RDF Graphs
  - good for human analysis but unsuitable for application exchange
- RDF serialization
  - provides a way to convert between the abstract model and a concrete syntax (format for files or other byte stream)
  - several equally expressive serialization formats
    - XML/RDF (normative (standard) exchange format for serialization)
    - N-Triples
    - Turtle (Terse RDF Triple Language)
    - Nquads
    - TriG
    - JSON-LD
    - RDFa

RDF Serializations

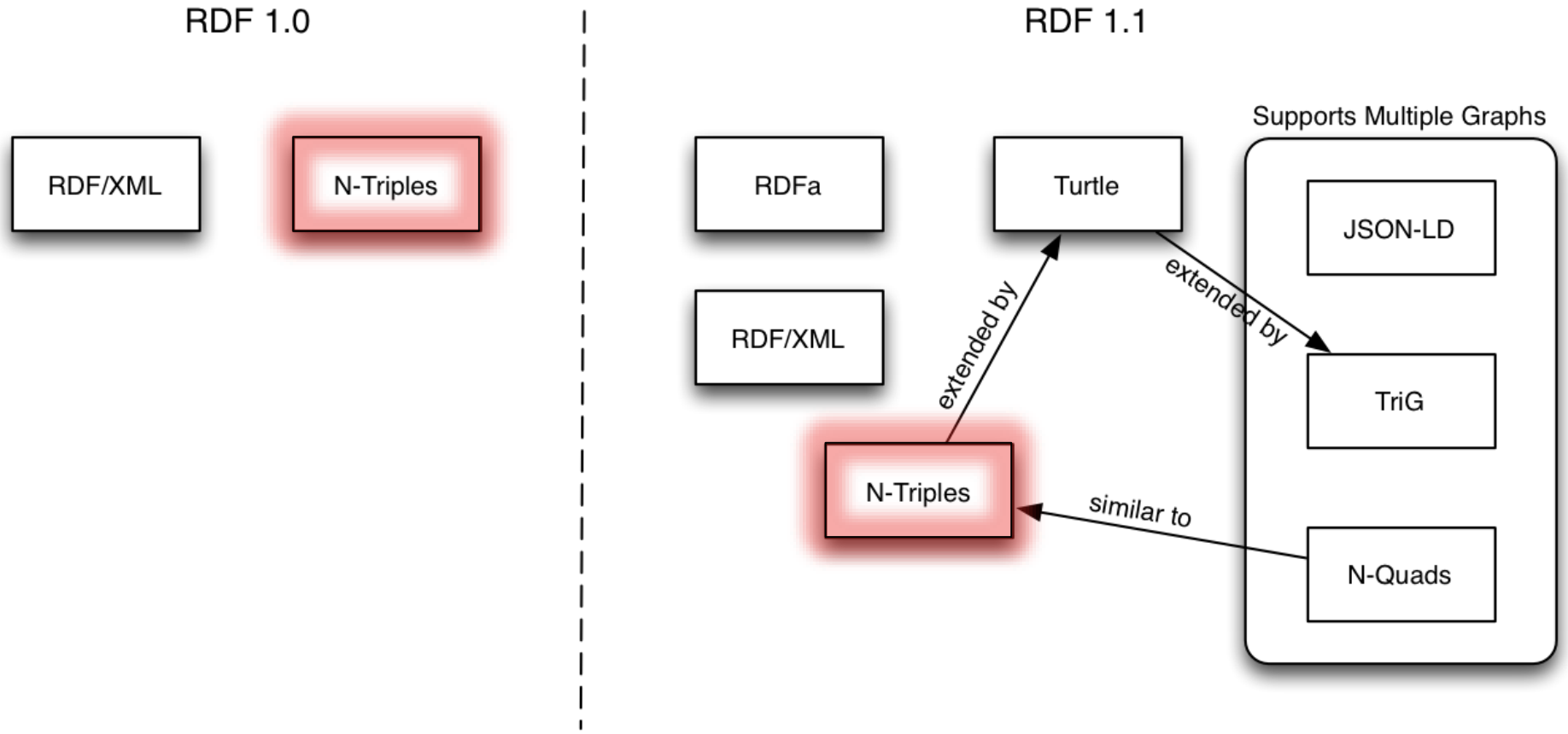


https://www.w3.org/TR/rdf11-new/#section-serializations

RDF Serializations - N-Triples

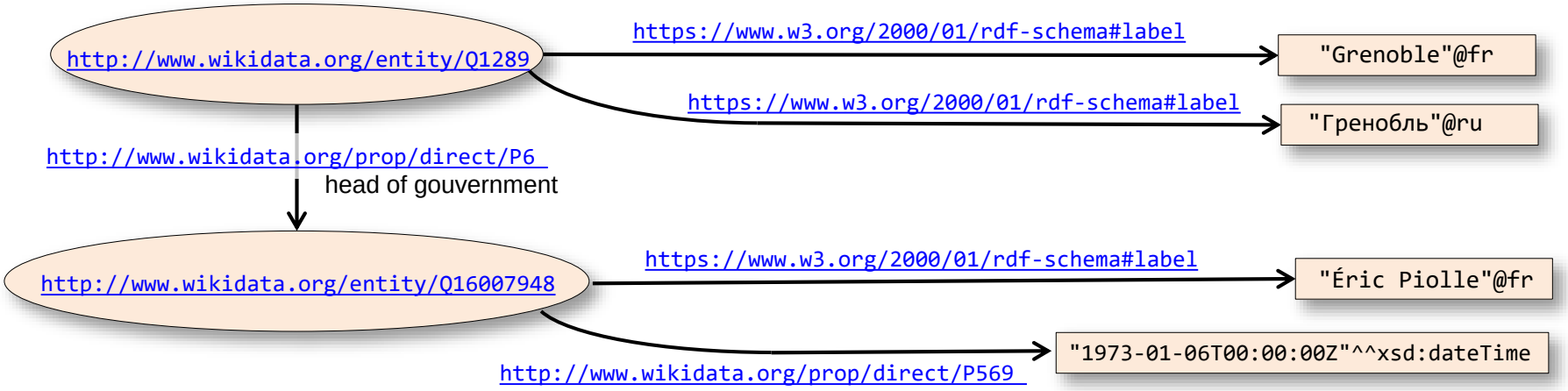
media type*: `application/n-triples`

*media type (formerly known as MIME, Multipurpose Internet Mail Extensions, type) : a two-part identifier for file formats and format contents transmitted on the Internet https://www.iana.org/assignments/media-types/media-types.xhtml



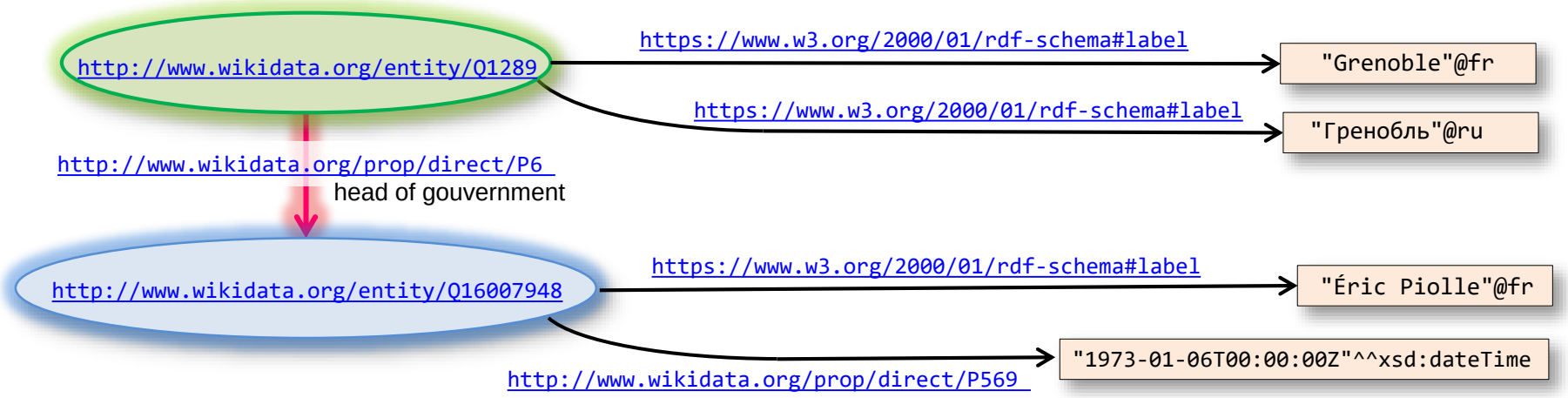
https://www.w3.org/TR/rdf11-new/#section-serializations

RDF Serialization - N-Triples



- the simplest notation
  - each line of output represents a single statement followed by '.'

RDF Serialization - N-Triples

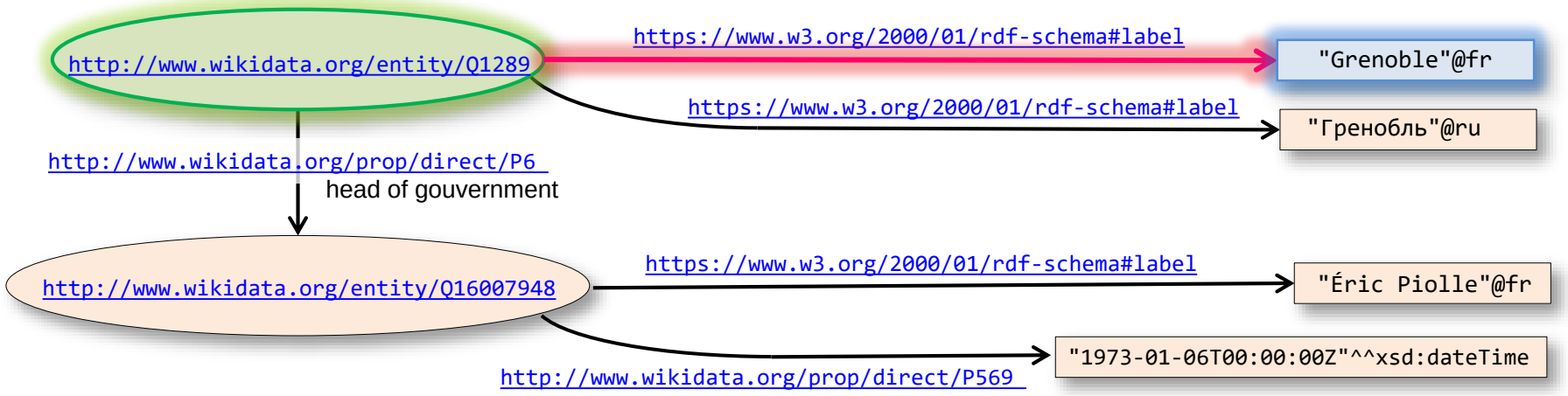


- the simplest notation
  - each line of output represents a single statement followed by '.'
  - resources (subject, predicate, resource object) expressed as absolute URI enclosed in angle brackets

subject | predicate | object

```
<http://www.wikidata.org/entity/Q1289> <http://www.wikidata.org/prop/direct/P6> <http://www.wikidata.org/entity/Q16007948> .
```

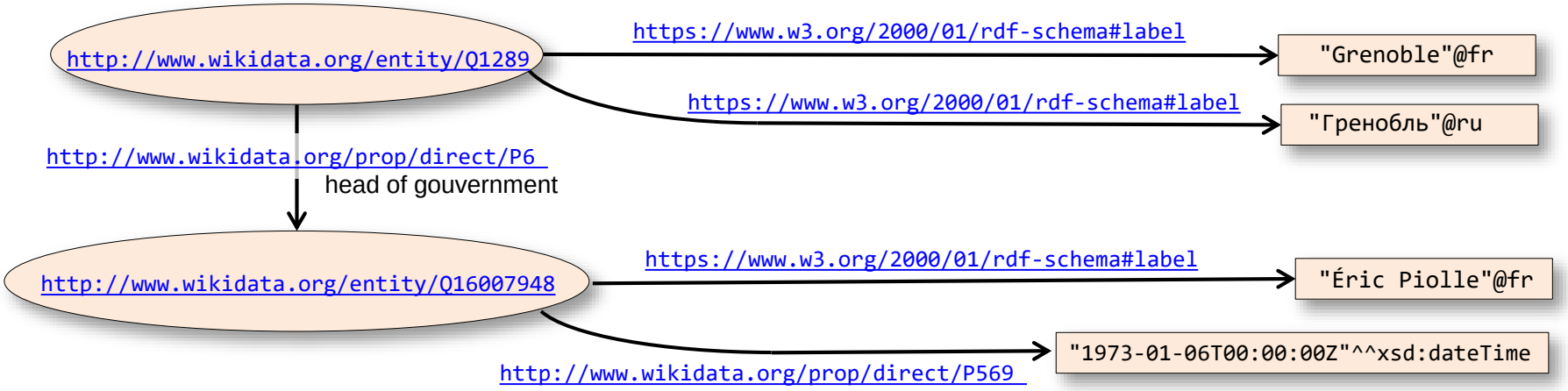
RDF Serialization - N-Triples



- the simplest notation
  - each line of output represents a single statement followed by '.'
  - resources (subject, predicate, resource object) expressed as absolute URI enclosed in angle brackets
  - object literals are double-quoted strings

```
<http://www.wikidata.org/entity/Q1289> <http://www.wikidata.org/prop/direct/P6> <http://www.wikidata.org/entity/Q16007948>.
<http://www.wikidata.org/entity/Q1289> <https://www.w3.org/2000/01/rdf-schema#label> "Grenoble"@fr .
```

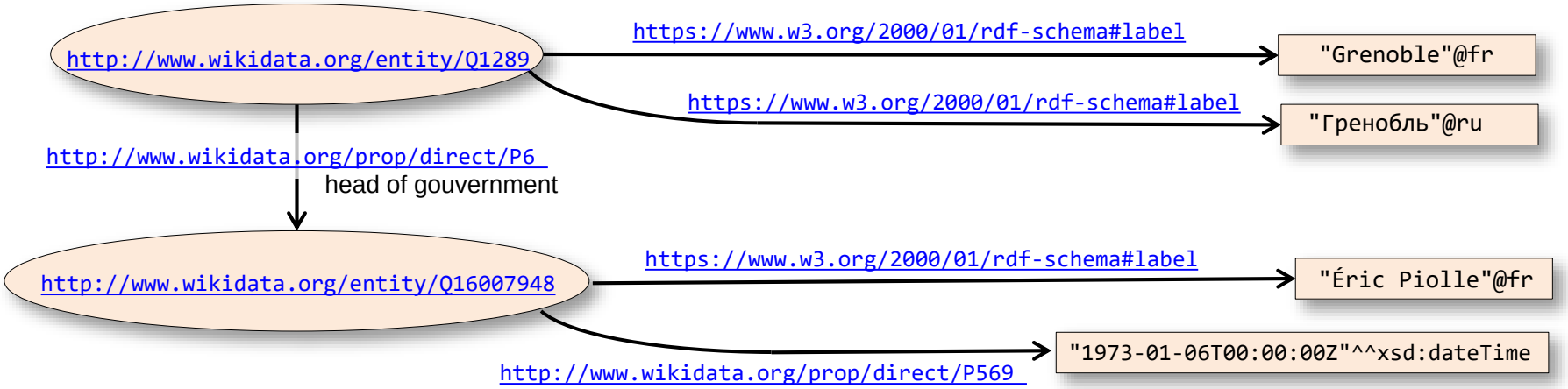
RDF Serialization - N-Triples



- the simplest notation
  - each line of output represents a single statement followed by '.'
  - resources (subject, predicate, resource object) expressed as absolute URI enclosed in angle brackets
  - object literals are double-quoted strings

```
<http://www.wikidata.org/entity/Q1289>  <http://www.wikidata.org/prop/direct/P6>  <http://www.wikidata.org/entity/Q16007948>.
<http://www.wikidata.org/entity/Q1289>   <https://www.w3.org/2000/01/rdf-schema#label>  "Grenoble"@fr   .
<http://www.wikidata.org/entity/Q1289>   <https://www.w3.org/2000/01/rdf-schema#label>  "Гренобль"@ru   .
<http://www.wikidata.org/entity/Q16007948> <https://www.w3.org/2000/01/rdf-schema#label>  "Éric Piolle"@fr .
<http://www.wikidata.org/entity/Q16007948> <http://www.wikidata.org/prop/direct/P569>
                                          "1973-01-06T00:00:00Z"^^<http://www.w3.org/2001/XMLSchema#dateTime> .
```

RDF Serialization - N-Triples



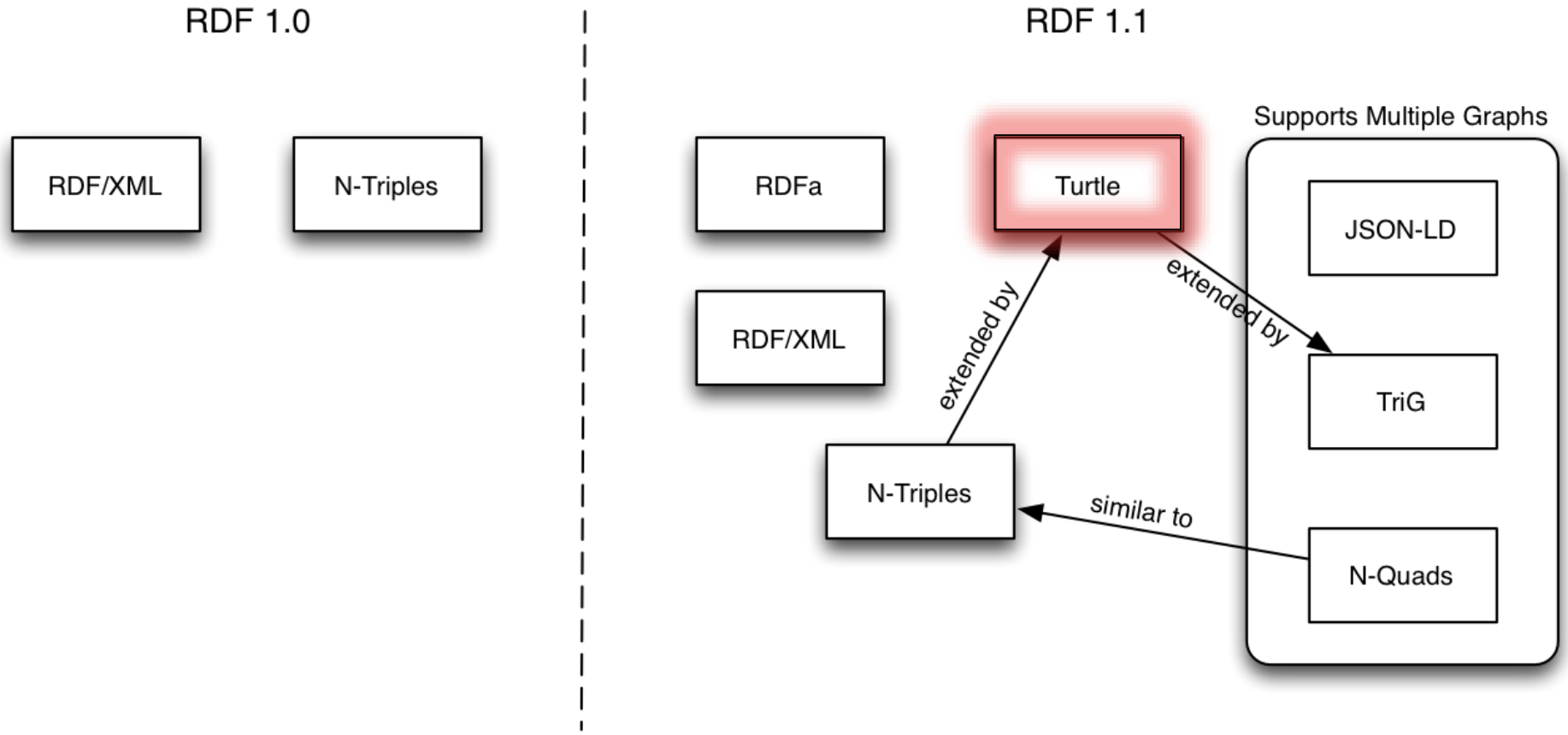
- the simplest notation
  - each line of output represents a single statement followed by '.'
  - resources (subject, predicate, resource object) expressed as absolute URI enclosed in angle brackets
  - object literals are double-quoted strings

```
<http://www.wikidata.org/entity/Q1289> <http://www.wikidata.org/prop/direct/P6> <http://www.wikidata.org/entity/Q16007948>.
<http://www.wikidata.org/entity/Q1289> <https://www.w3.org/2000/01/rdf-schema#label> "Grenoble"@fr .
<http://www.wikidata.org/entity/Q1289> <https://www.w3.org/2000/01/rdf-schema#label> "Гренобль"@ru .
<http://www.wikidata.org/entity/Q16007948> <https://www.w3.org/2000/01/rdf-schema#label> "Éric Piolle"@fr .
<http://www.wikidata.org/entity/Q16007948> <http://www.wikidata.org/prop/direct/P569> "1973-01-06T00:00:00Z"^^<http://www.w3.org/2001/XMLSchema#dateTime> .
```

- useful when hand-crafting data sets for application testing and debugging
- … but verbose (redundant information takes additional time to transmit and parse)

RDF Serializations - Turtle

media (MIME) type: `text/turtle`



https://www.w3.org/TR/rdf11-new/#section-serializations

RDF Serialization – N3 – Turtle

- Notation3 (N3) more compact format than N-Triples.
  - has several absolute features that go beyond a serialization for RDF models (e.g. support for RDF-based rules).
- Turtle (Terse RDF Triple Language)
  - a simplified, RDF-only subset of N3.
- Both condense much of the repetitions of N-Triples
  - URIs can be shortened by using a prefix declared at the beginning of the document

```
<http://www.wikidata.org/entity/Q1289> <http://www.wikidata.org/prop/direct/P6>
                                                <http://www.wikidata.org/entity/Q16007948> .
<http://www.wikidata.org/entity/Q1289> <https://www.w3.org/2000/01/rdf-schema#label> "Grenoble"@fr .
```

```
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/>
@prefix rdfs: <https://www.w3.org/2000/01/rdf-schema#>
```

```
wd:Q1289 wdt:P6 wd:Q16007948 .
wd:Q1289 rdfs:label  "Grenoble"@fr .
```

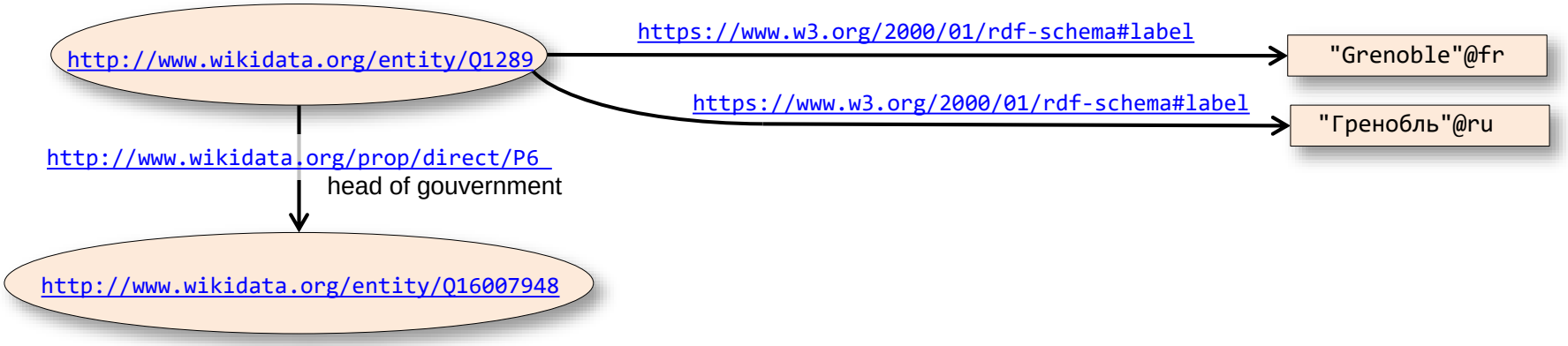
RDF Serialization - N3 - Turtle

– possibility to combine multiple statements about the same subject using a semicolon (';')

```
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .
@prefix rdfs: <https://www.w3.org/2000/01/rdf-schema#> .

wd:Q1289 wdt:P6 wd:Q16007948 ;
         rdfs:label  "Grenoble"@fr .
```

RDF Serialization - N3 - Turtle



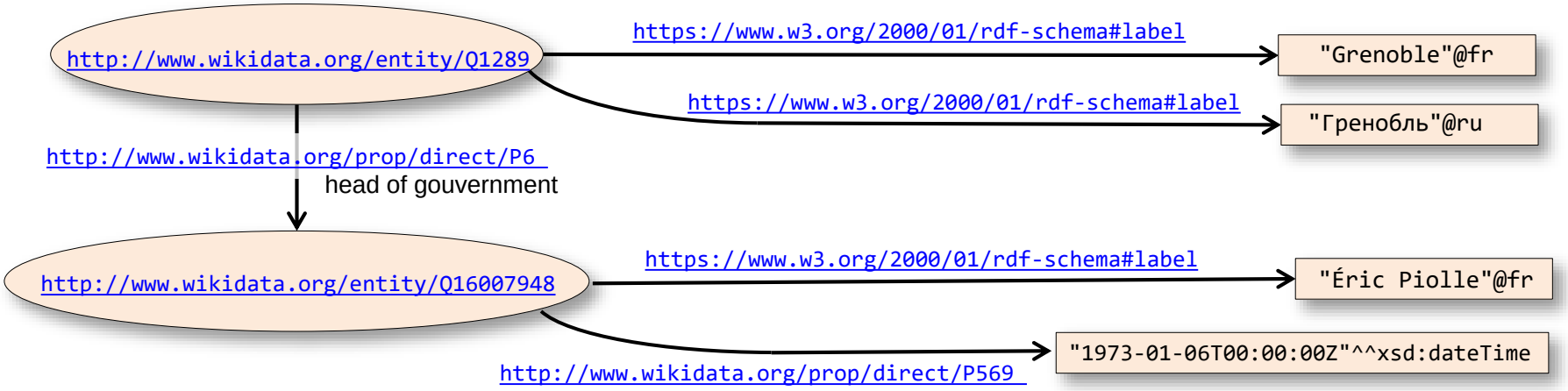
– possibility to combine multiple statements involving the same subject and predicate using a coma (',')

```
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .
@prefix rdfs: <https://www.w3.org/2000/01/rdf-schema#> .

wd:Q1289 wdt:P6 wd:Q16007948 ;
         rdfs:label  "Grenoble"@fr , "Гренобль"@ru .
```

ne pas oublier le point terminal

RDF Serialization - N3 - Turtle




```
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .
@prefix rdfs: <https://www.w3.org/2000/01/rdf-schema#> .

wd:Q1289 wdt:P6 wd:Q16007948 ;
         rdfs:label  "Grenoble"@fr , "Гренобль"@ru .

wd:Q16007948 rdfs:label "Éric Piolle"@fr ;
         wdt:P569 "1973-01-06T00:00:00Z"^^xsd:dateTime .
```

Un peu de pratique

- Est-il possible de récupérer la description RDF de cette ressource http://dbpedia.org/resource/Grenoble dans le format de sérialisation Turtle ?
  - Oui / Non
- Comment ?
  - Si oui décrire le processus sans le faire

Un peu de pratique : solution

- Oui,
- pour récupérer une description en Turtle il suffit de faire une requête HTTP de type GET avec l'url de la ressource (http://dbpedia.org/resource/Grenoble_Alpes_University) mais en précisant dans l'en tête Accept le type MIME text/turle.

Un peu de pratique

Le 3 juin 1950 Maurice Herzog et Louis Lachenal réussirent la première ascension de l'Annapurna. Parmi les alpinistes qui participèrent à l'expédition ayant permis cette victoire se trouve Jean Couzy. Lui même a été le vainqueur de plusieurs sommets de l'Himalaya, mais savez-vous lesquels ? A quelles dates ? Et avec qui ?

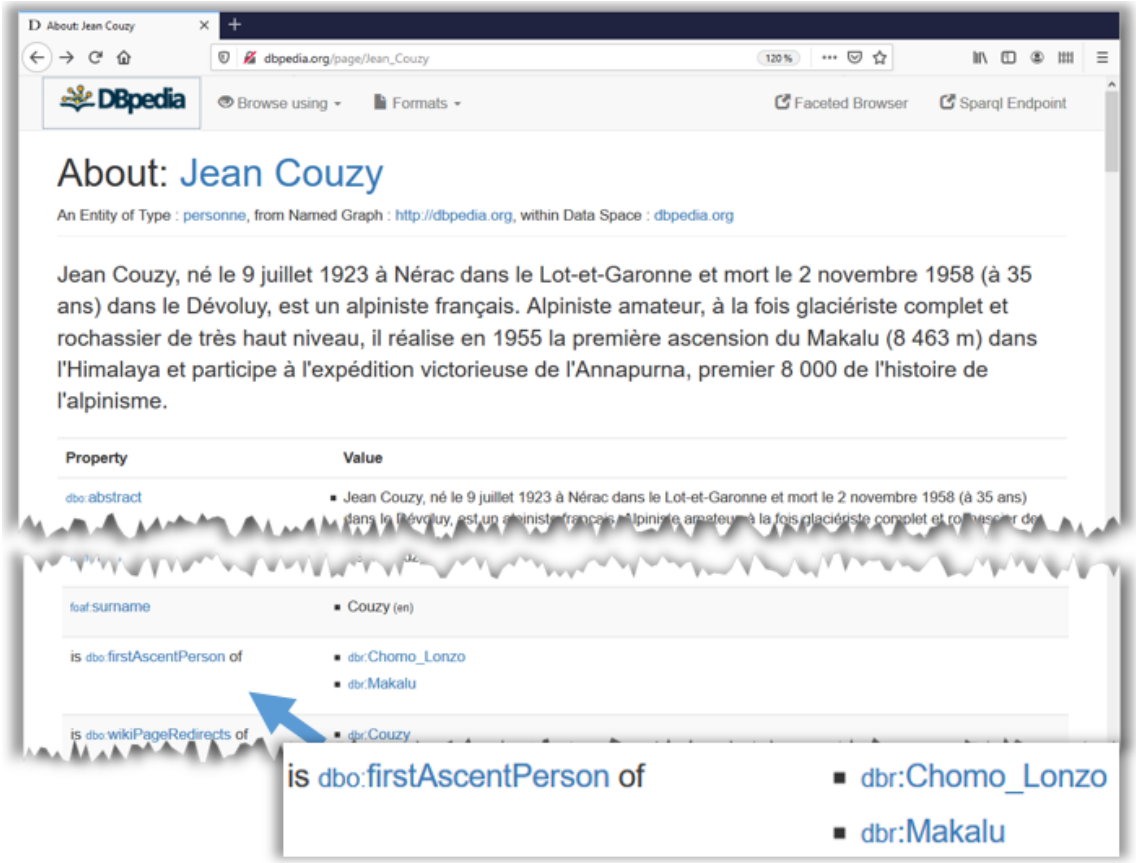
Question 1: Faites une recherche sur le web pour trouver les réponses à ces questions.

Les informations précédentes sont présentes dans Wikipedia et ont été extraites pour être intégrées au graphe de données de DBpedia. On pourrait facilement les retrouver à l'aide d'une simple requête SPARQL, mais pour le moment nous allons nous contenter de les obtenir en parcourant manuellement DBpedia

Question 2: En partant de la ressource représentant Jean Couzy dans DBpedia retrouvez les informations de la question 1. Dessinez le graphe présentant les resources de DBpedia que vous avez parcourues et les prédicats les reliant.

Un peu de pratique : solution

Dans DBpedia, la ressource qui représente Jean Couzy a pour URI http://dbpedia.org/page/Jean_Couzy. Si on explore cette ressource avec un navigateur on voit qu'il y a une propriété dbo:firstAscentPerson qui la relie au Chomo Lonzo et au Makalu.

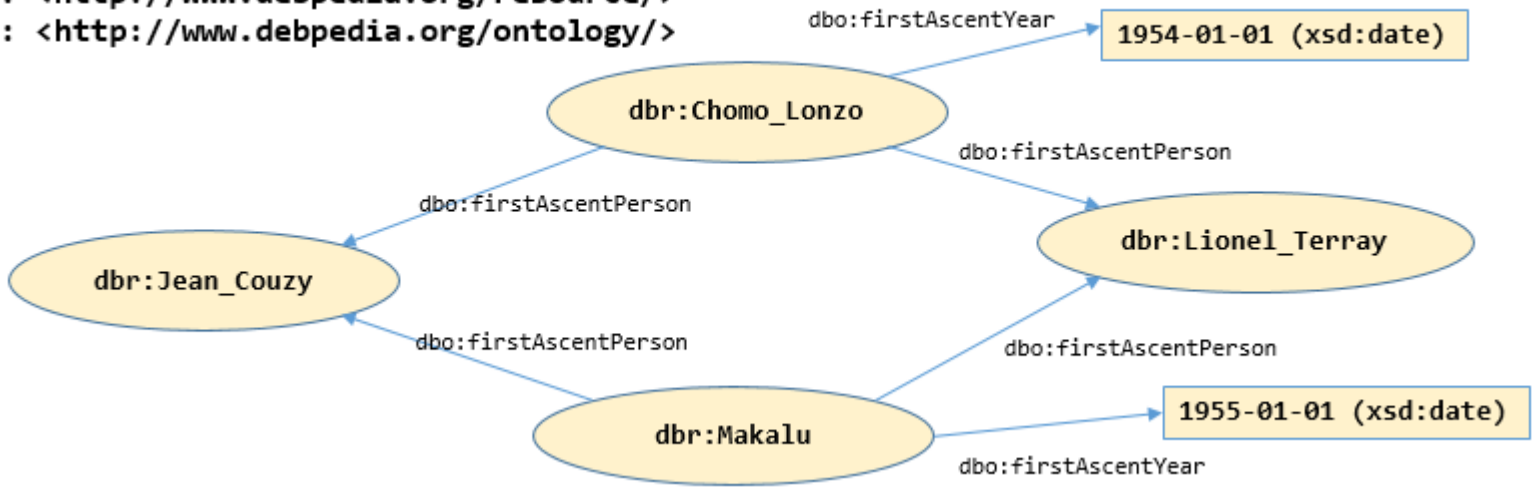


La propriété **is dbo:firstAscentPerson of dbr:Chomo_Lonzo** signifie que **dbr:Jean_Cousy** est **l'objet** d'un triplet dont **dbr:Chomo_Lonzo** est le sujet et **dbo:firstAscentPerson** le prédicat

Un peu de pratique : solution

Le graphe de données dans DBpedia.
ensuite en suivant les liens dbr:Chomo_Lonzo et dbr:Makalu on peut facilement compléter le graphe
dbr : resource
dbo : ontology

préfixes
dbr: <http://www.debpedia.org/resource/>
dbo: <http://www.debpedia.org/ontology/>



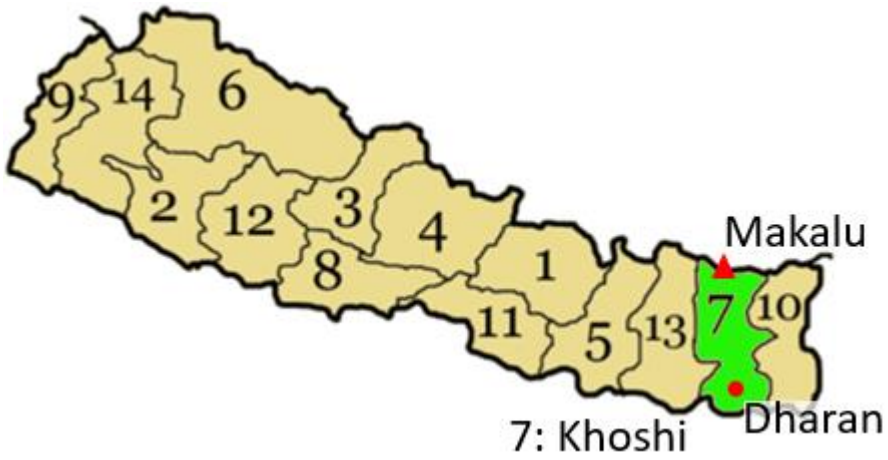
Un peu de pratique

Toujours en explorant le graphe RDF de DBpedia répondez à cette question :
Quelle est la région administrative où se trouve le versant népalais du Makalu et capitale de celle-ci et complétez le dessin de votre graphe RDF.

Un peu de pratique : solution

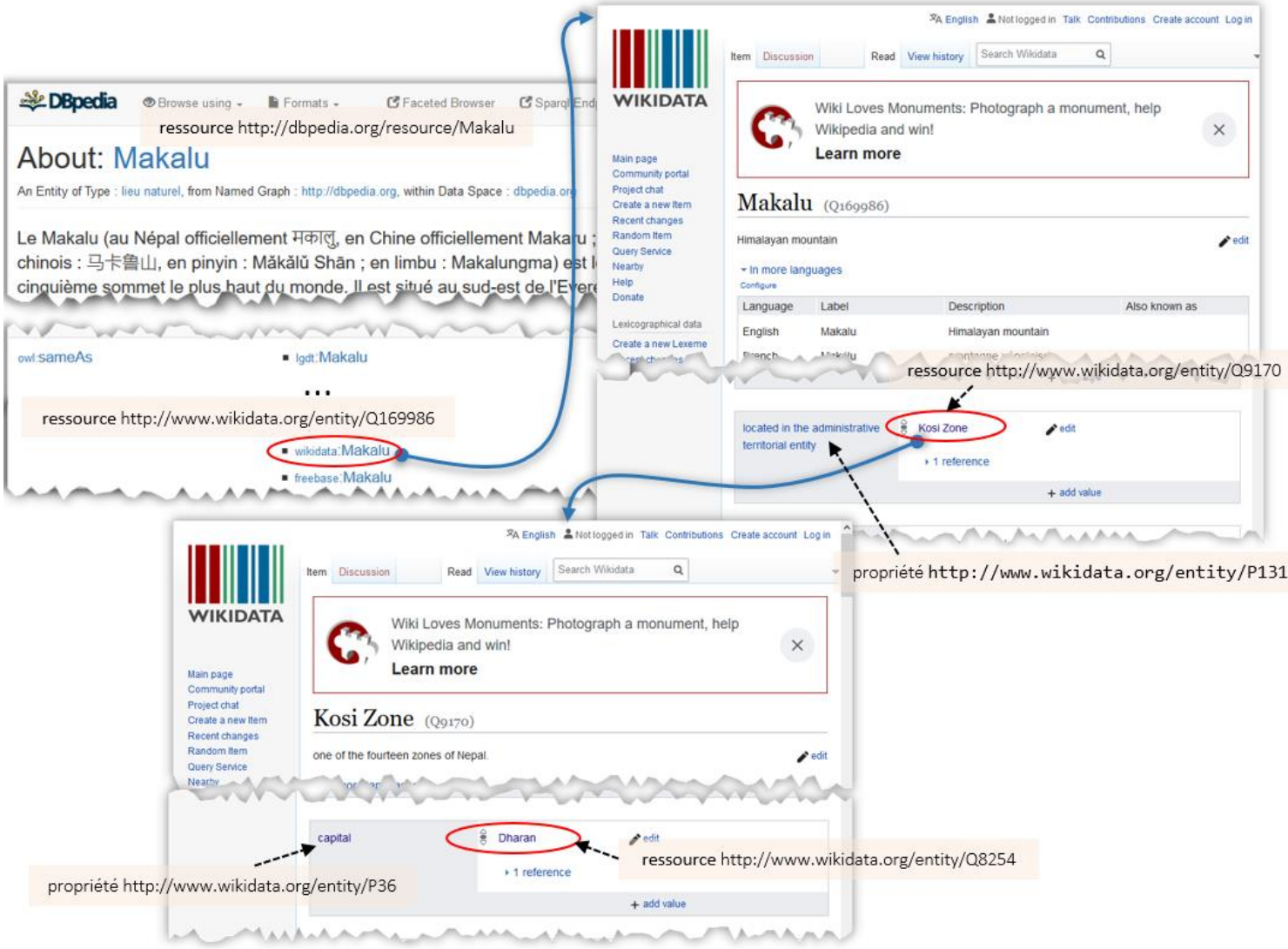
Zone administrative du Népal où se trouve le Makalu ?

La bonne réponse est la zone administrative du Kosi (zone 7 sur la carte) et sa capitale est Darhan.



Un peu de pratique : solution

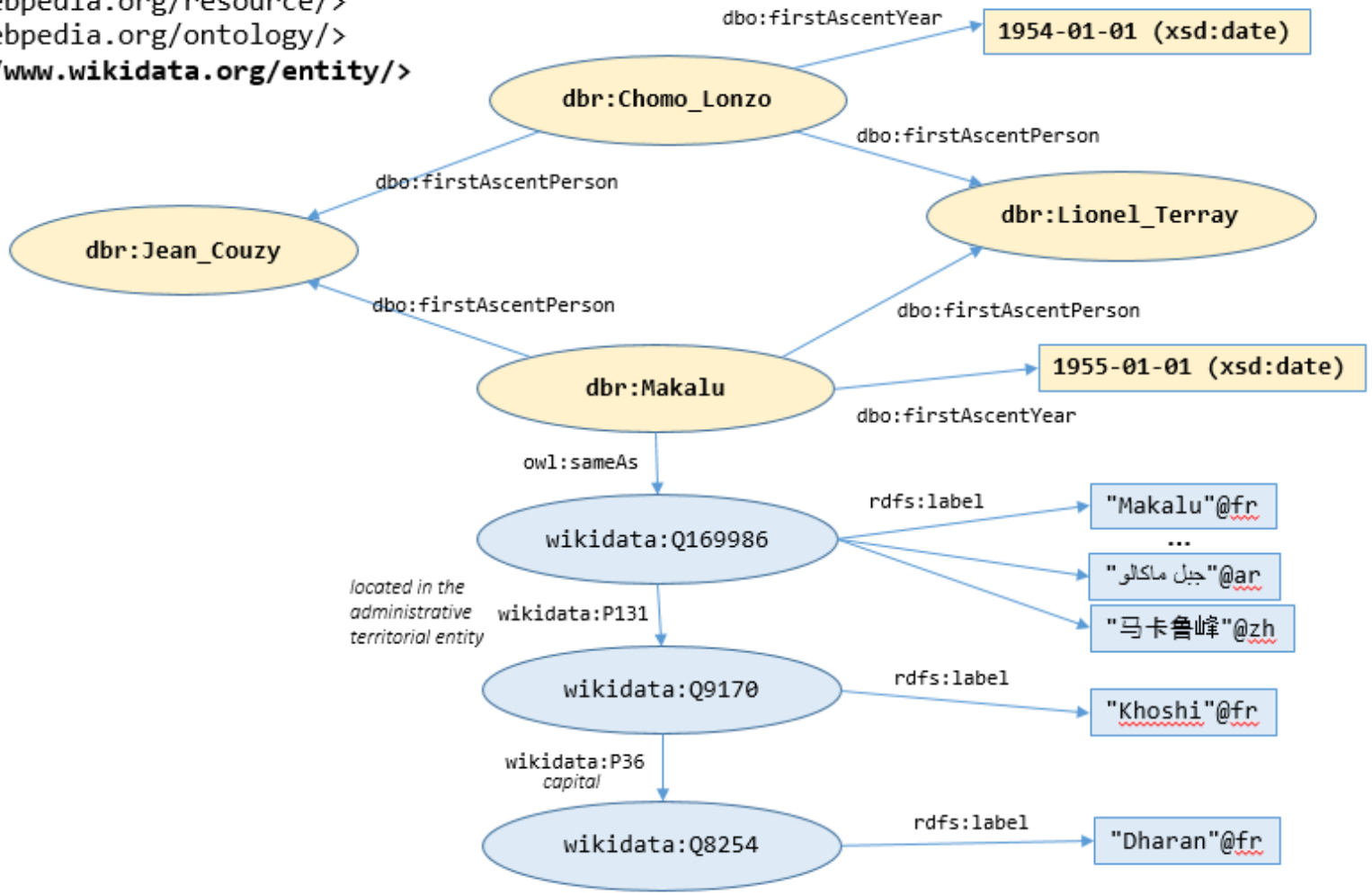
Cette information a pu être trouvée en suivant les liens suivants : Recherche des informations dans Wikidata



Un peu de pratique : solution

ce qui correspond au graphe RDF ci-dessous

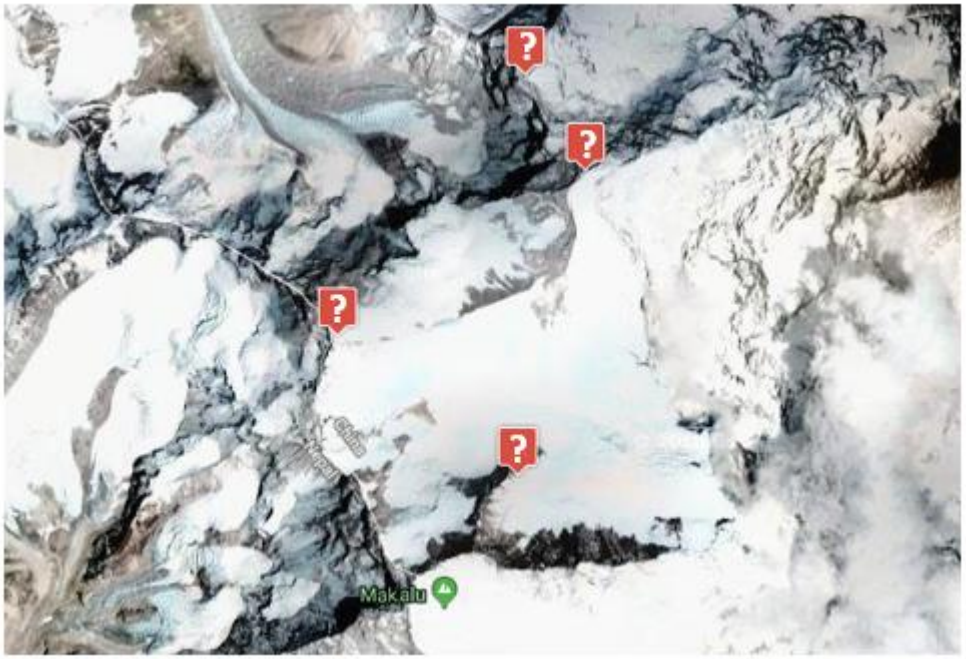
```
dbr: <http://www.debpedia.org/resource/>
dbo: <http://www.debpedia.org/ontology/>
wikidata: <https://www.wikidata.org/entity/>
```



Le graphe de données dans DBpedia.

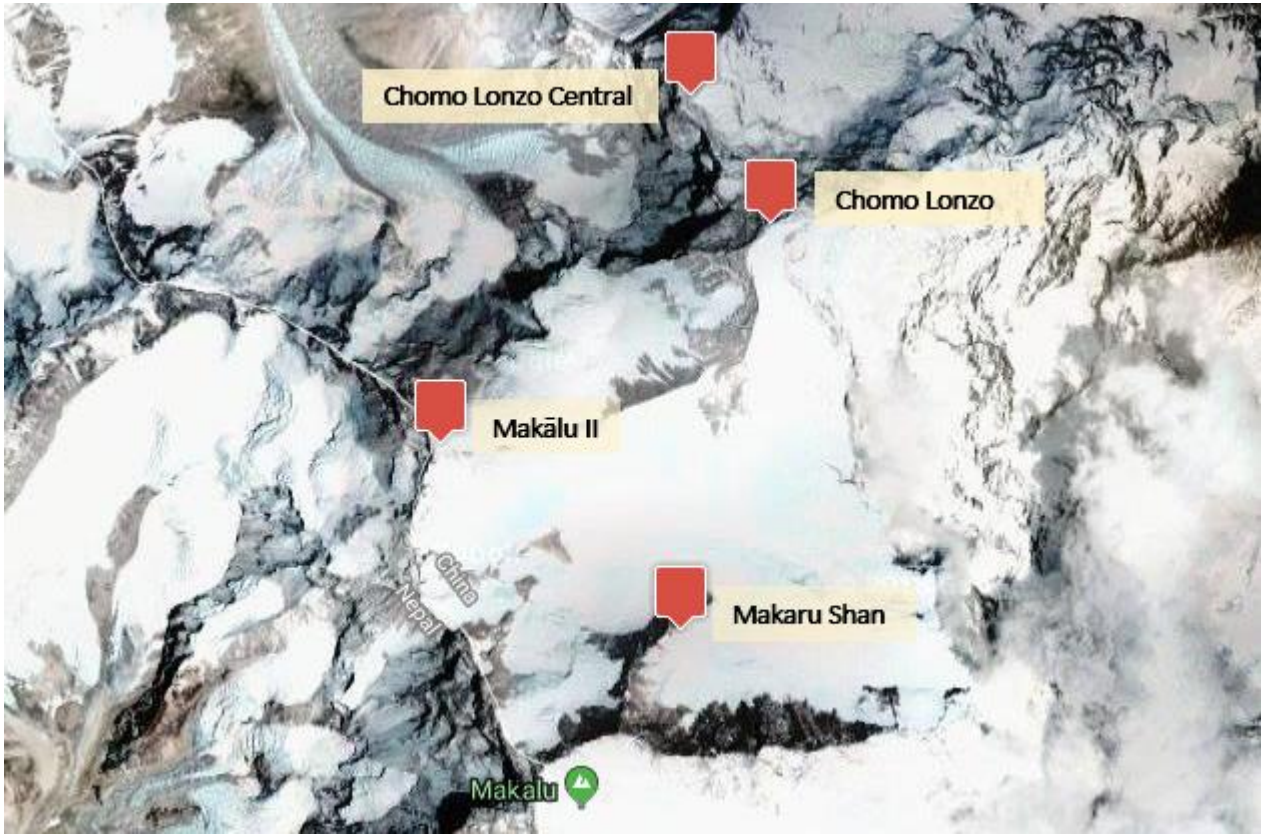
Un peu de pratique

Toujours en vous servant des données Dbpedia, pouvez-vous trouver des sommets proches du Makalu ? Par exemple pourriez-vous nommer ces 4 sommets ?



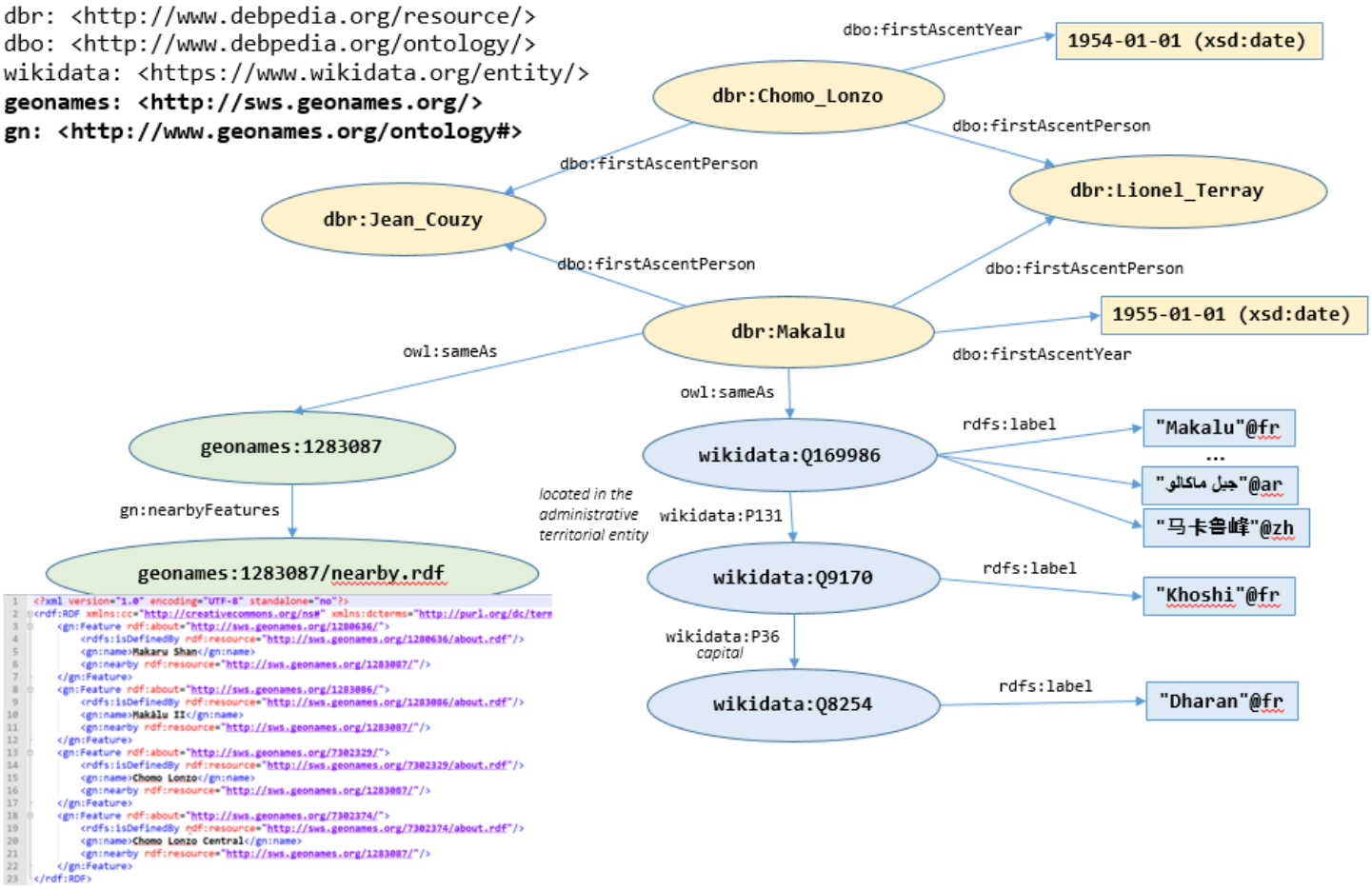
Un peu de pratique : solution

Les sommets proches du Makalu repertoriés dans geonames sont : le Makaru Shan, le Makalu II, le Chomo Lonzo et le Chomo Lonzo Central.



Un peu de pratique : solution

Ces informations ont pu être trouvées en suivant les liens suivants owl:sameAs vers geodata:Makalu (geodata:Makalu est un alias pour la ressource geonames:1283087) puis depuis cette ressouce le lien gn:nearByFeatures vers geonames:1283087/nearby.rdf, ce qui correspond au graphe RDF ci-dessous



Le graphe de données dans geonames.

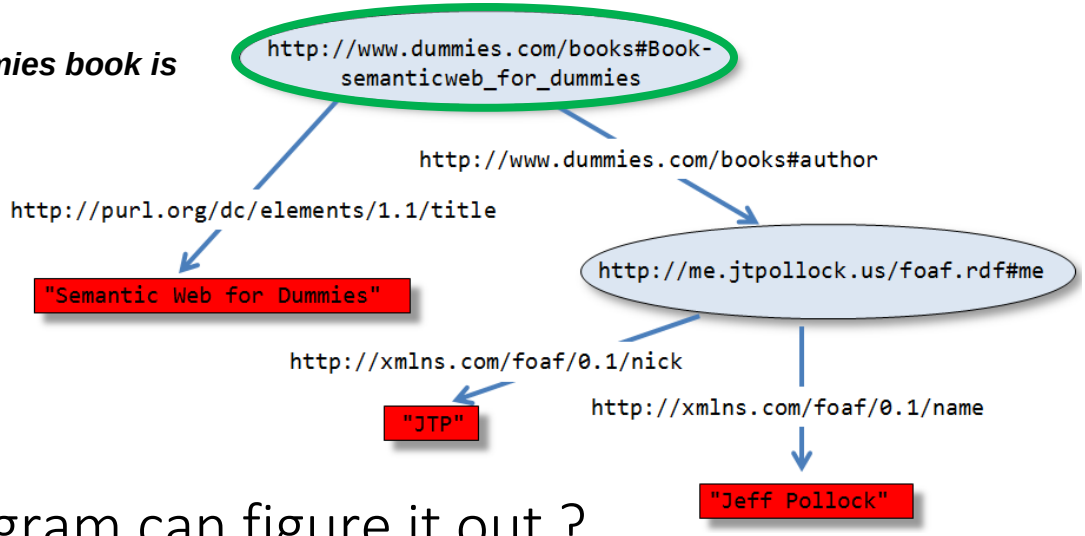
RDF outline

- RDF Model
- Typed and localized literals
- RDF formats
- Blank nodes
- **Resources definition**
- RDF an data integration
- Persisting RDF
- References

Identifying the type of a resource

- the same way literals can be typed, it's possible to associate a type to a resource

The Semantic Web For Dummies book is authored by Jeff Pollock

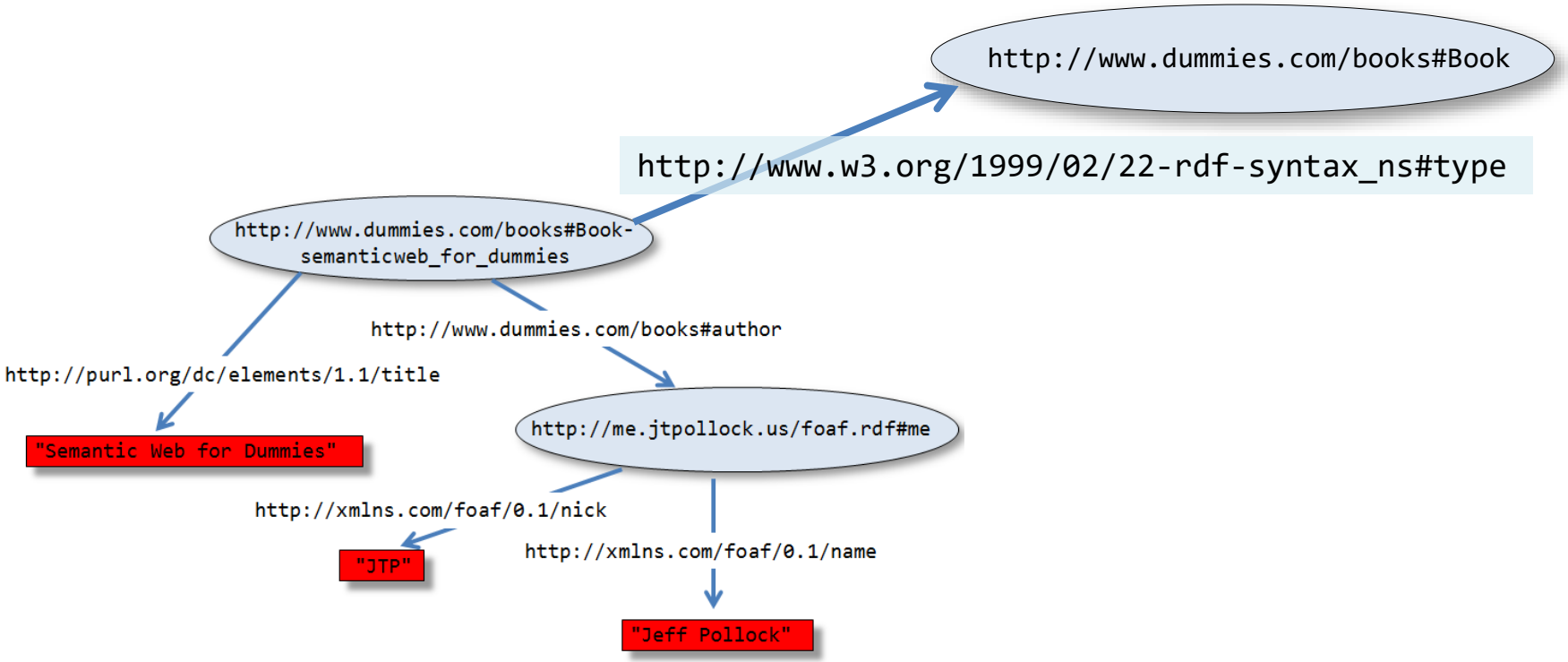


We know it's a book…
but how a computer program can figure it out ?

To solve the problem of classifying resources a way the software can understand, RDF vocabulary has a predefined predicate : **`rdf:type`**

Identifying the type of a resource

- `rdf:type` predicate's semantics
  - the value of this predicate is a resource and represents a **class** of things
  - the subject of this predicate is also an instance of that class



Identifying the type of a resource

- N3 - Turtle

```
@prefix swbook: <http://www.dummies.com/books#>.

swbook:Book-semanticweb_for_dummies
        swbook:author <http://me.jtpollock.us/foaf.rdf#me>;
        a swbook:Book.
```

shortcut for rdf:type

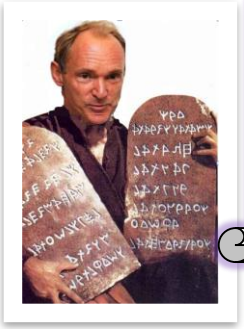
- XML/RDF

```
<?xml version="1.0"?>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:books="http://www.dummies.com/books#">
   <rdf:Description
      rdf:about="http://www.dummies.com/books#Book-semanticweb_for_dummies">
      <rdf:type rdf:resource="http://www.dummies.com/books#Book"/>
      <books:author
           rdf:resource="http://me.jtpollock.us/foaf.rdf#me" />
   </rdf:Description>
</rdf:RDF>
```

RDF outline

- RDF Model
- Typed and localized literals
- RDF formats
- Blank nodes
- Resources definition
- RDF and data integration
- Persisting RDF
- References

RDF and Data Integration



Linked Data : 4th Principle

Include links to other URIs, so that they can discover more things.

DBpedia

`http://dbpedia.org/resource/grenoble`

GeoNames

`http://sws.geonames.org/6454071`

`dbpedia:Stendhal`

`dbpedia-owl:birthPlace`

`dbpedia:Grenoble`

`dbpedia-owl:elevation`

`"398"^^xsd:double`

`geonames:3014728`

`insee:38185`

INSEE

`gn:parentFeature`

`gn:name`

`geonames:6454071`

`"Grenoble"`

RDF and Data Integration

adaptation of presentations by Ivan Herman (W3C) ivan@w3.org) at
Semantic Technology Conferences 2009 et 2011
(San Jose, CA. USA, June, 2009) (San Francisco, CA. USA, June, 20
http://www.w3.org/2009/Talks/0615-SanJose-tutorial-IH/
http://www.w3.org/2011/Talks/0606-SemTech-Tut-IH/

- Dataset "A": a simplified bookstore data base

BOOKS

ID	Author	Title	Publisher	Year
ISBN 0-00-6511409-X	id_xyz	The Glass Palace	id_qpr	2000

FK

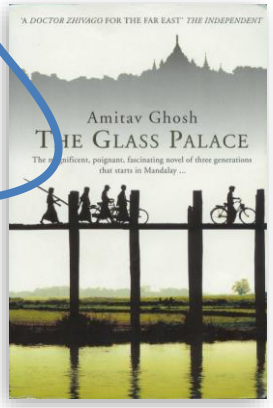
FK

AUTHORS

ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

PUBLISHERS

ID	Publisher's name	City
id_qpr	Harper Collins	London



RDF and Data Integration

BOOKS

ID	Author	Title	Publisher	Year
ISBN 0-00-6511409-X	id_xyz	The Glass Palace	id_qpr	2000

FK FK

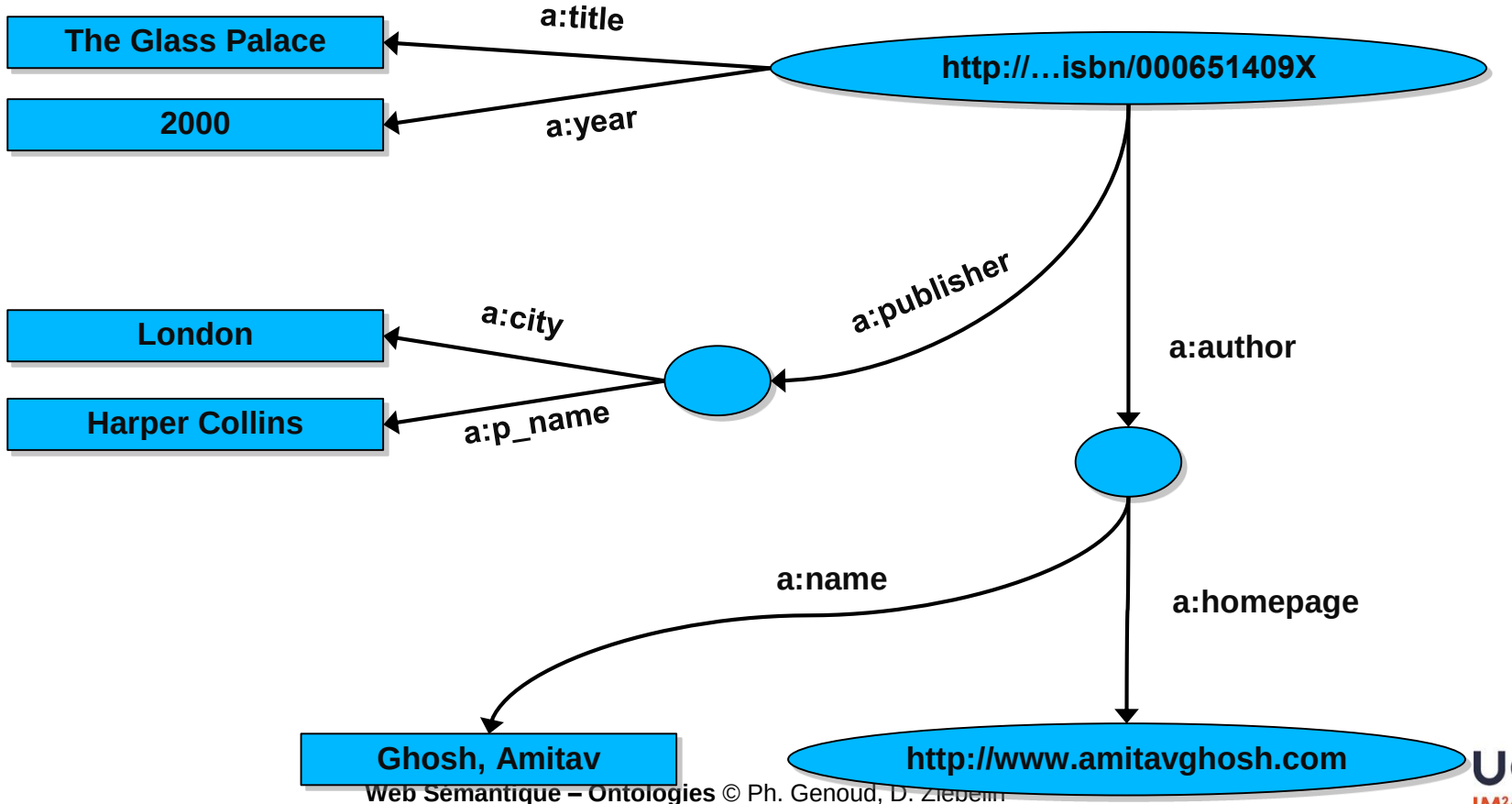
AUTHORS

ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

PUBLISHERS

ID	Publisher's name	City
id_qpr	Harper Collins	London

- 1st: export your data as a RDF graph



RDF and Data Integration

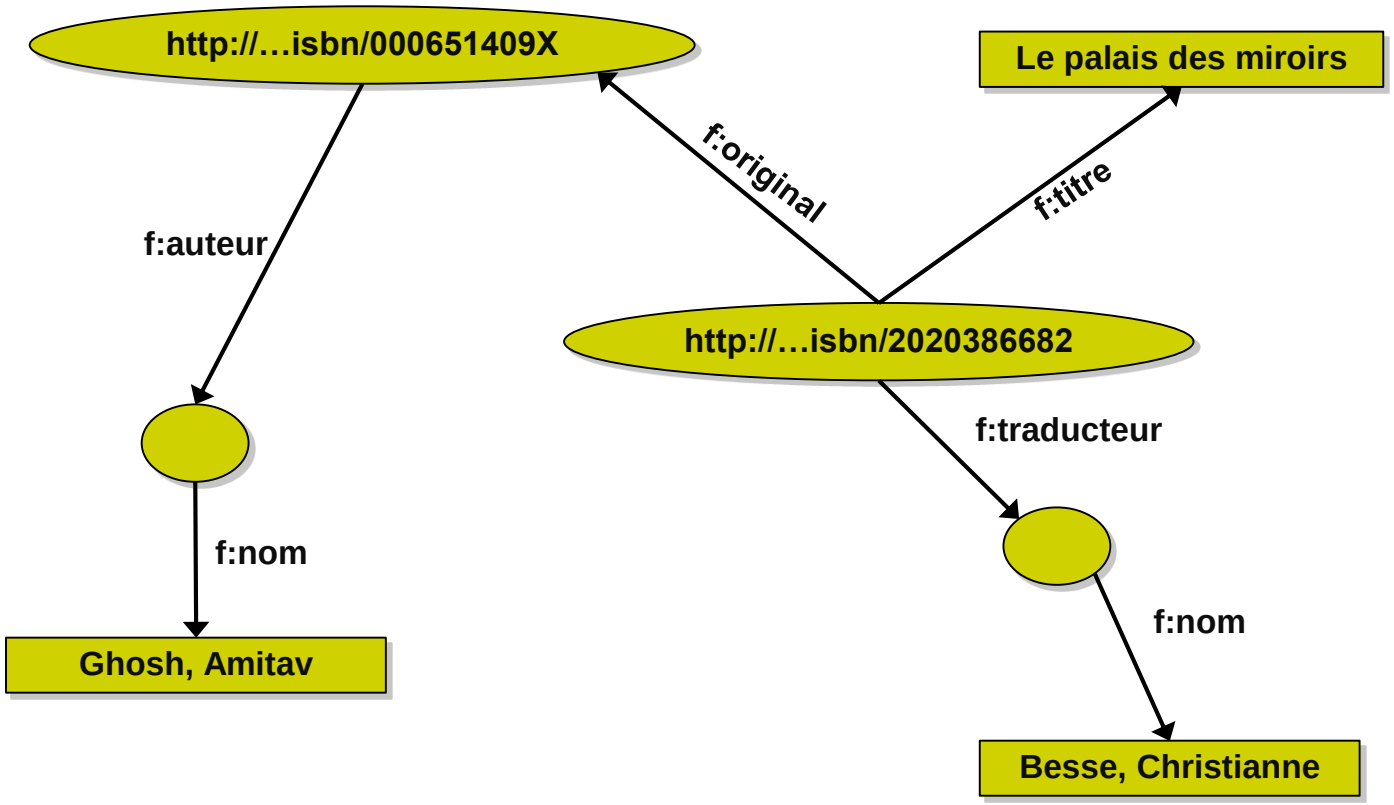
- Another dataset "F" : a google docs spreadsheet bookstore data

	A	B	C	D
1	**ID**	**Titre**	**Traducteur**	**Original**
2	ISBN 2020286682	Le Palais des Miroirs	$A12$	ISBN 0-00-6511409-X
3				
4				
5				
6	**ID**	**Auteur**		
7	ISBN 0-00-6511409-X	$A11$		
8				
9				
10	**Nom**			
11	Ghosh, Amitav			
12	Besse, Christianne			

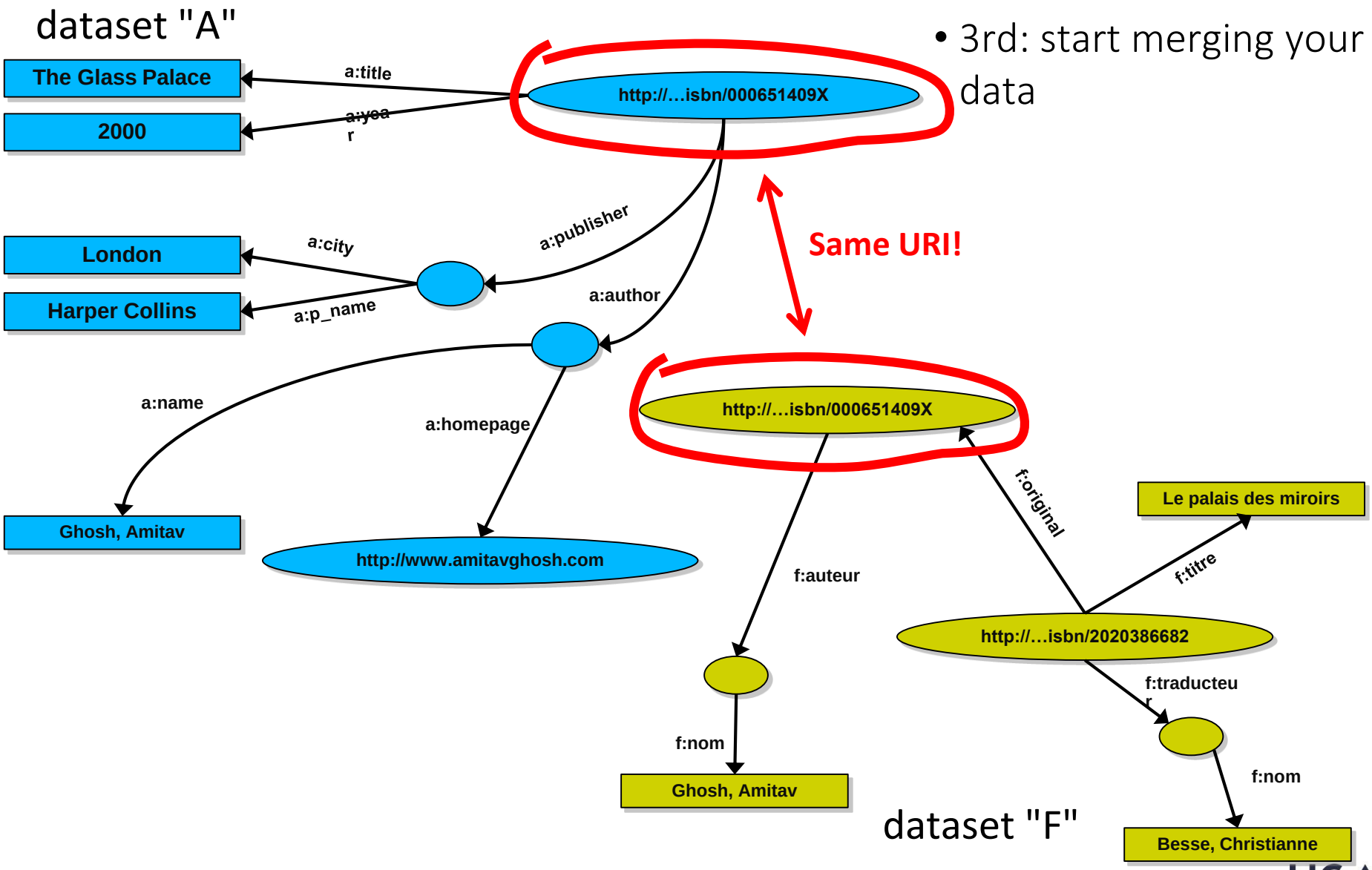
RDF and Data Integration

	A	B	C	D
1	**ID**	**Titre**	**Traducteur**	**Original**
2	ISBN 2020286682	Le Palais des Miroirs	$A12$	ISBN 0-00-6511409-X
3				
4				
5				
6	**ID**	**Auteur**		
7	ISBN 0-00-6511409-X	$A11$		
8				
9				
10	**Nom**			
11	Ghosh, Amitav			
12	Besse, Christianne			

- 2nd: export your second set of data to another RDF graph

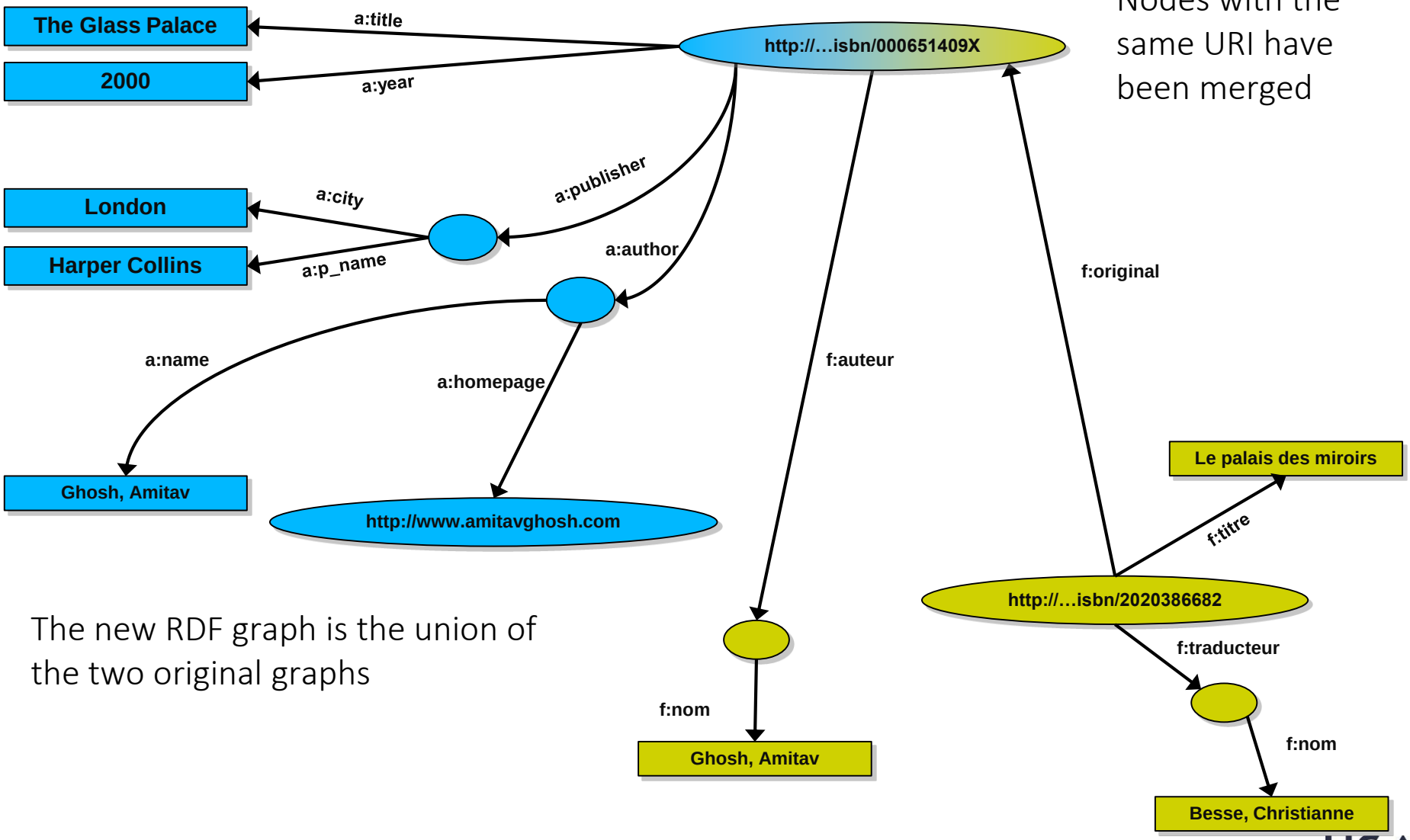


RDF and Data Integration



- 3rd: start merging your data

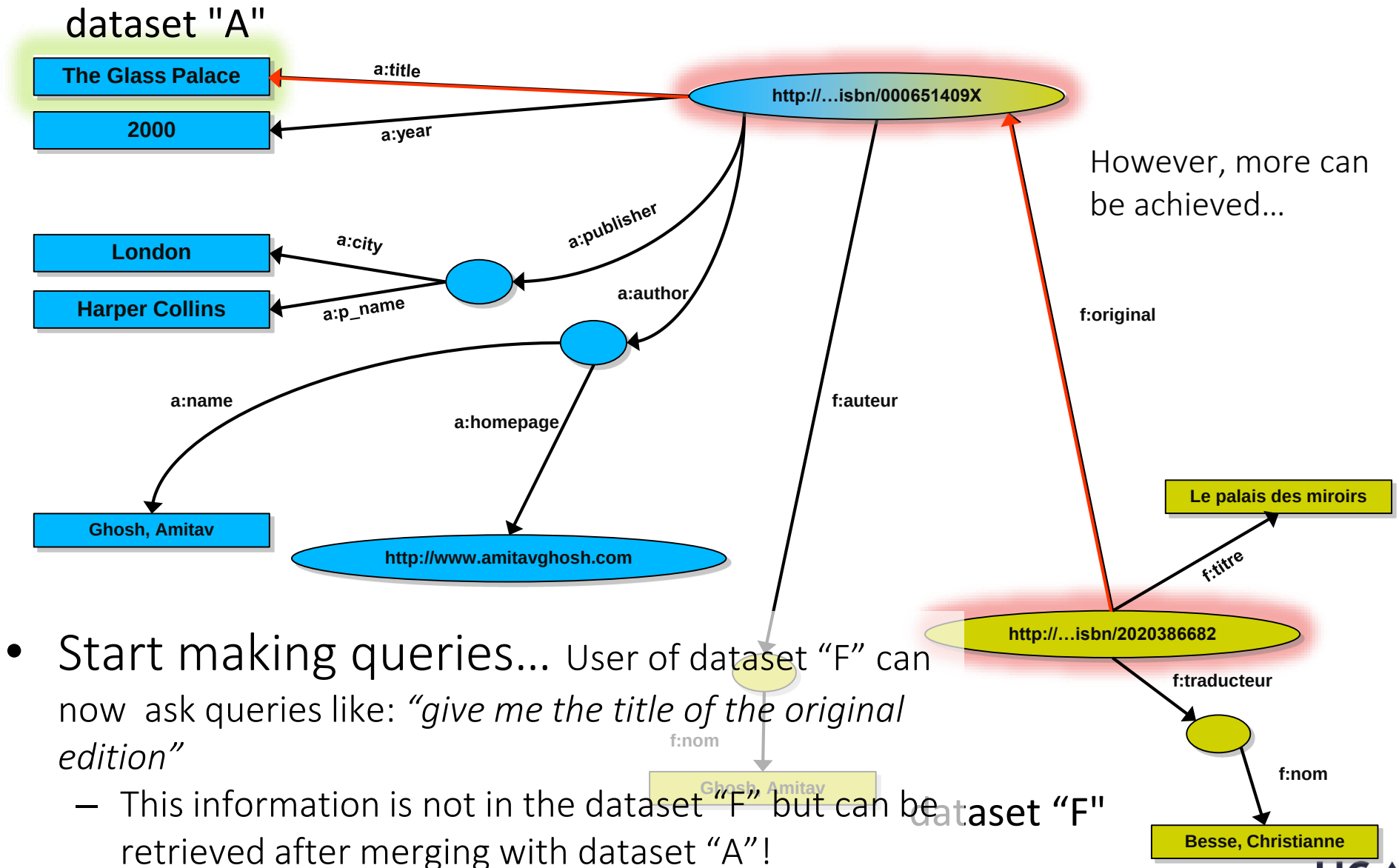
RDF and Data Integration



Nodes with the same URI have been merged

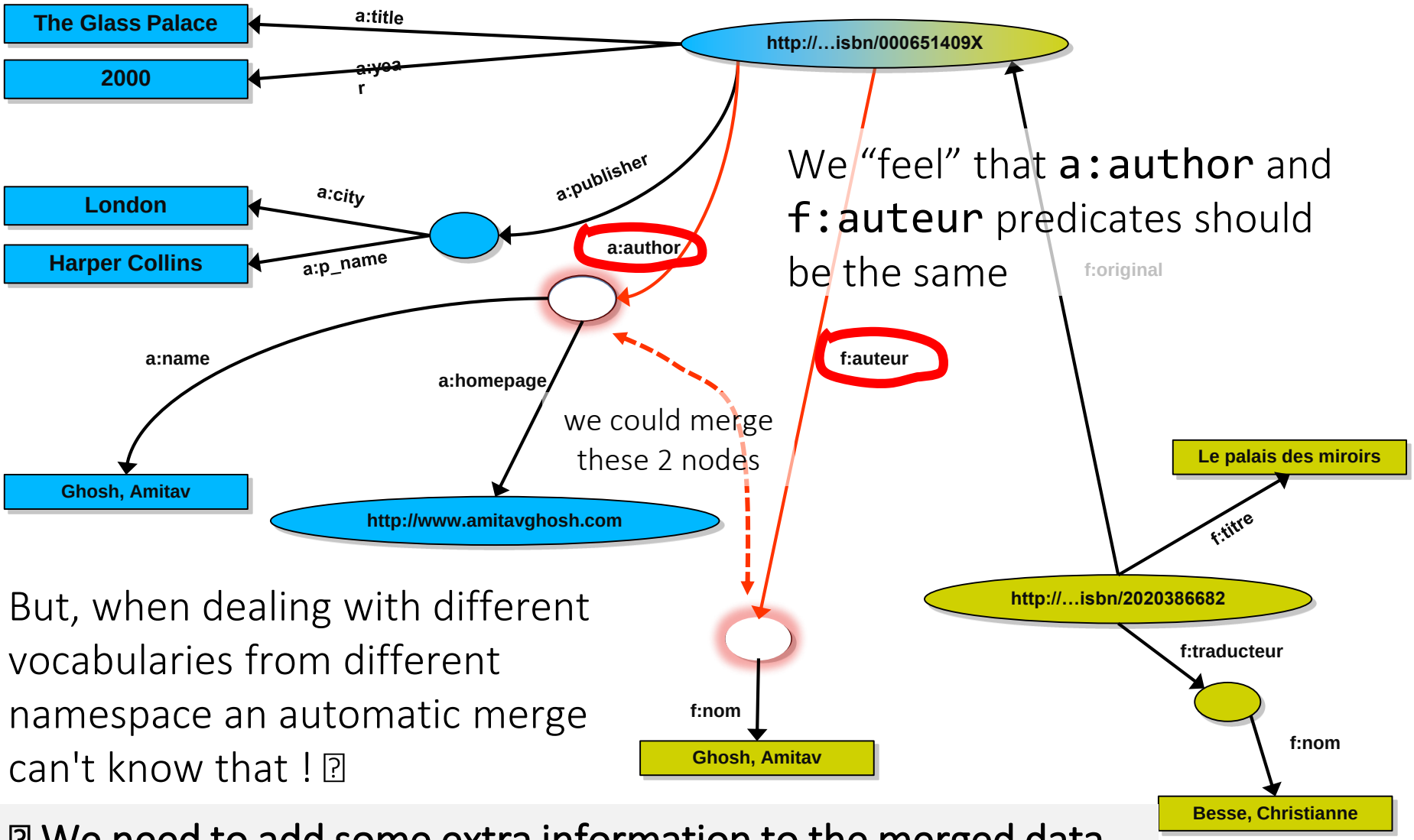
The new RDF graph is the union of the two original graphs

RDF and Data Integration



- Start making queries… User of dataset “F” can now ask queries like: *“give me the title of the original edition”*
  - This information is not in the dataset “F” but can be retrieved after merging with dataset “A”!

RDF and Data Integration



But, when dealing with different vocabularies from different namespace an automatic merge can't know that !

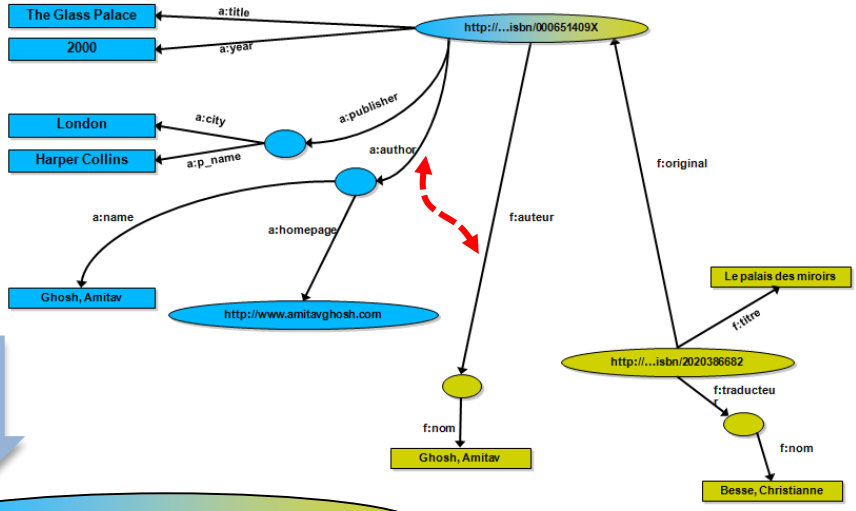
 We need to add some extra information to the merged data

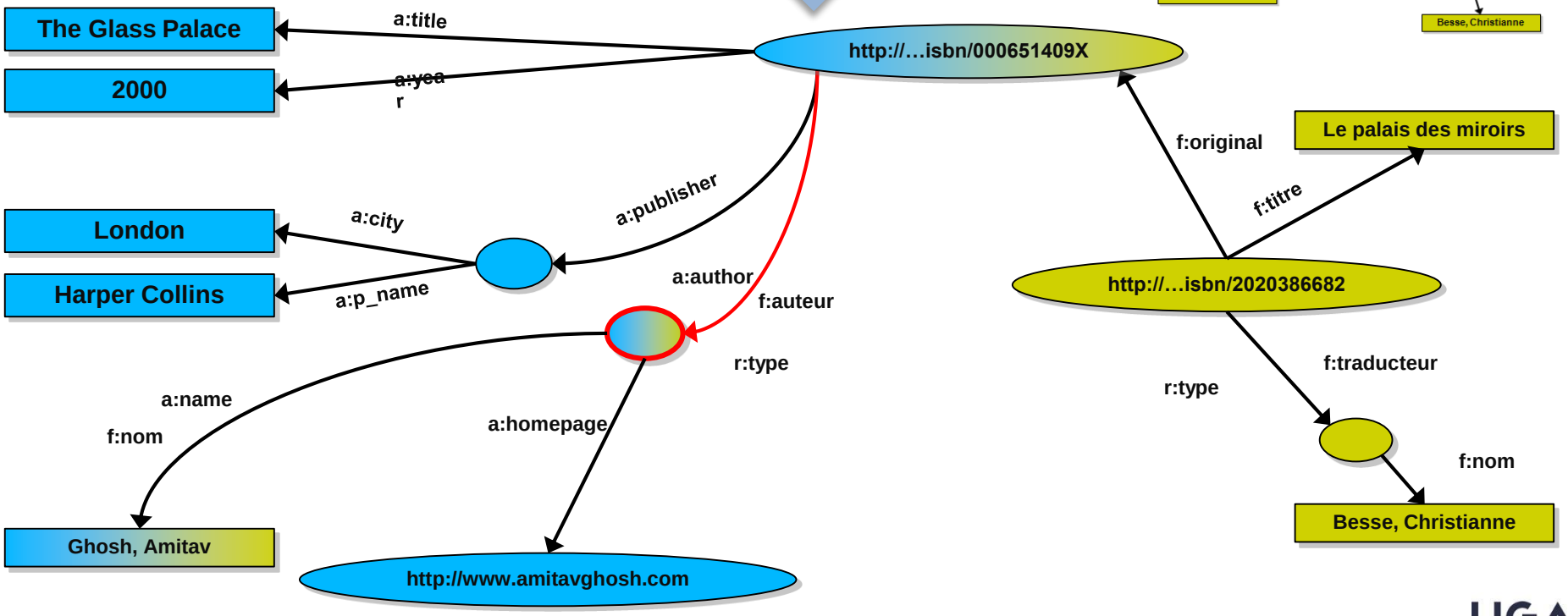
RDF and Data Integration

`a:author` and `f:auteur` are URIs identifying resources in different namespaces. We can add RDF statement about them in our RDF graph

```
a:author owl:equivalentProperty f:auteur
```

The well defined meaning (semantic) of this property allows to performs the merging



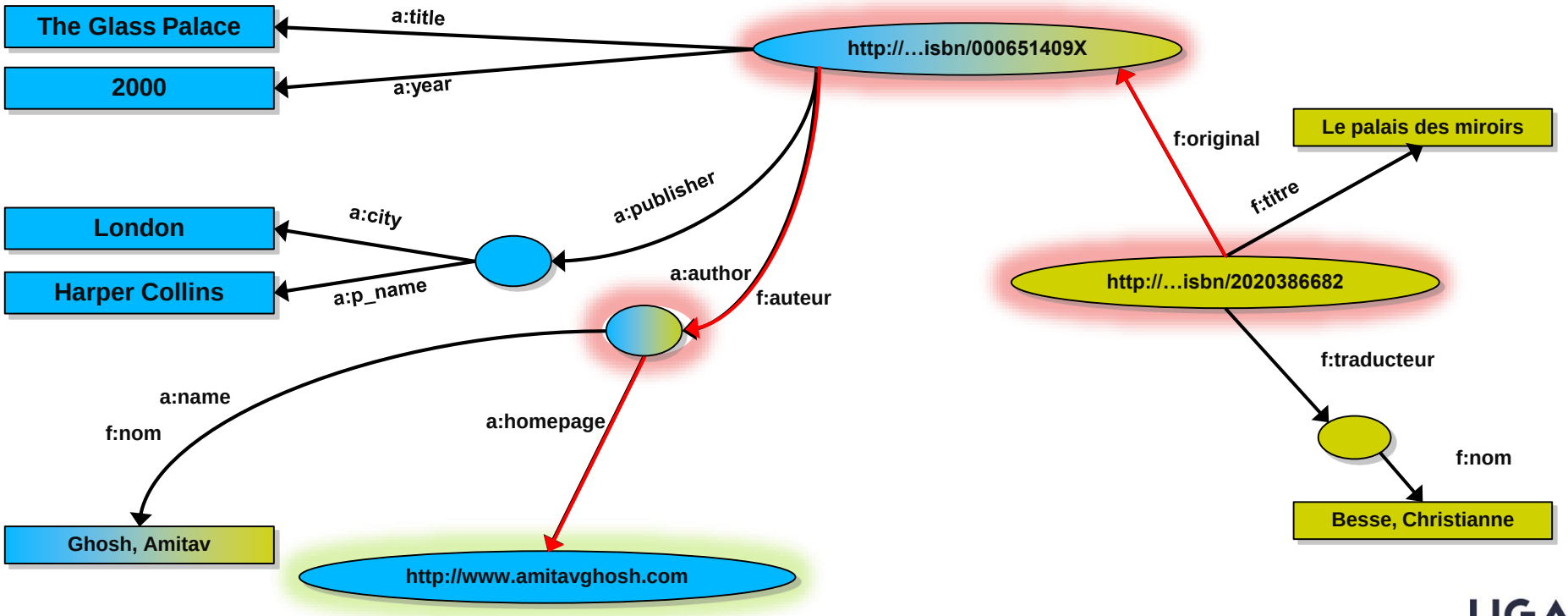


RDF and Data Integration

- By merging datasets “A” and datasets “F”
- By adding simple extra statements (owl:sameAs) as an extra “glue”
- It’s now possible to make richer queries

“donnes-moi la page d’accueil de l’auteur de l’édition originale”

“give me the home page of the original’s ‘auteur’”



Examples of Links between datasets

Prefixes

```
dbpedia:     "http://dbpedia.org/resource/"
dbo:         "http://dbpedia.org/ontology/"
geonames:    "http://sws.geonames.org/"
gn:          "http://www.geonames.org/ontology#"
insee:       "http://id.insee.fr/geo/commune/"
inseePop:     http://id.insee.fr/demo/populationLegale/commune/
inseeDef:     http://rdf.insee.fr/def/demo#populationTotale
```

Different URIs (URNs) in different namespaces can represents the same thing

INSEE

DBpedia

GeoNames

http://dbpedia.org/resource/Toulouse

http://sws.geonames.org/2972315/

"449328"^^xsd:integer

inseeDef:populationTotal

dbpedia:David_Skrela

geonames:2972315

inseePop:31555/2010

dbpedia-owl:birthPlace

gn:parentFeature

gn:name

owl:sameAs

dbpedia:Toulouse

inseeDef:population

"Toulouse"

dbo:populationTotal

geonames:6453974

insee:31555

"461190"^^xsd:integer

owl:sameAs

Examples of Links between datasets

In the RDF representation of this Dbpedia resource there is a triple **that links it** to Geonames

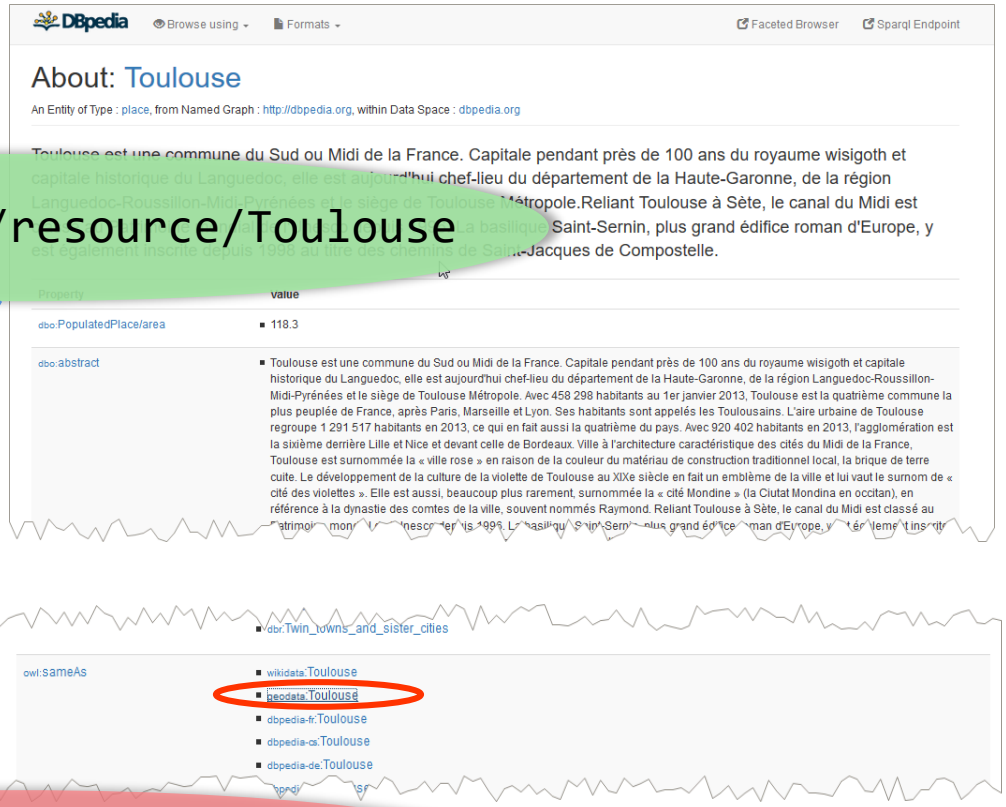
subject

http://dbpedia.org/resource/Toulouse

owl:sameAs

object

http://sws.geonames.org/2972315/



URN http://sws.geonames.org/2972315/ One resource

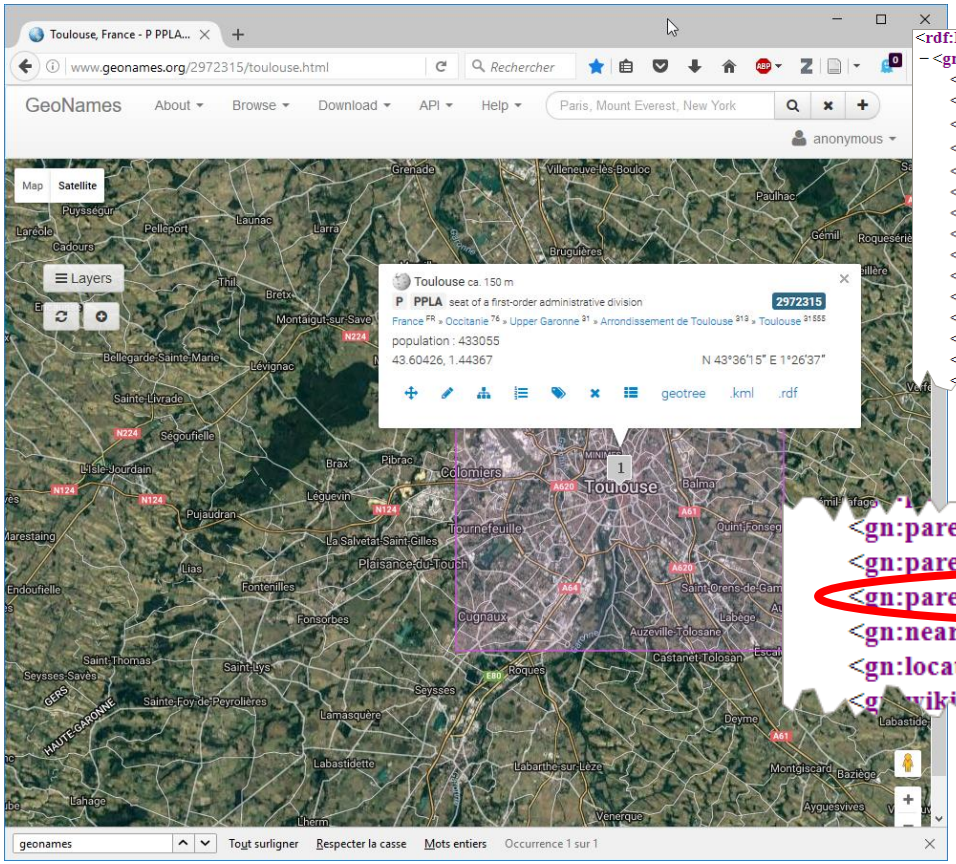
Multiple representations

URL

http://www.geonames.org/2972315/toulouse.html

URL

http://sws.geonames.org/2972315/about.rdf

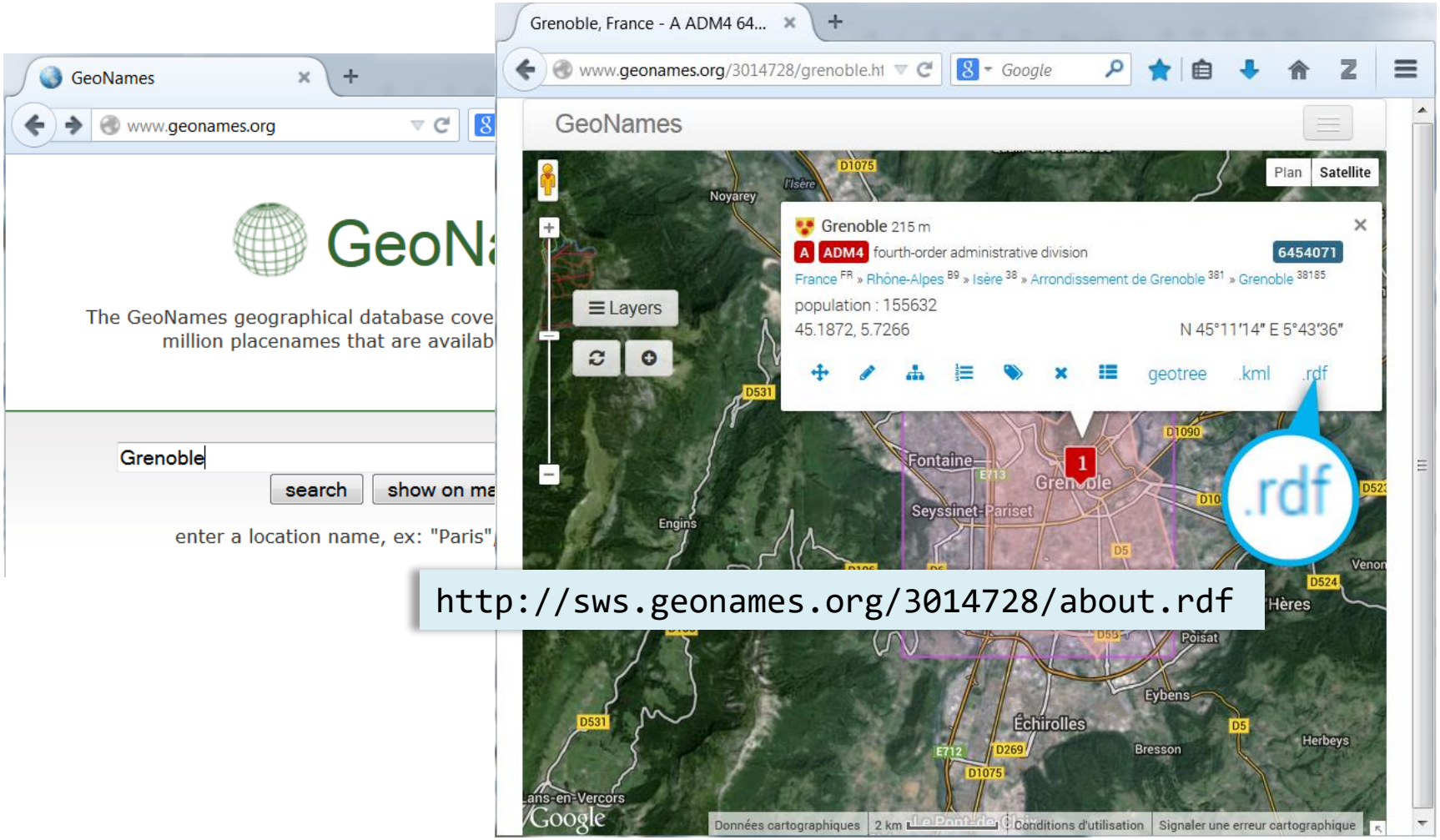



```
<rdf:RDF>
– <gn:Feature rdf:about="http://sws.geonames.org/2972315/">
    <rdfs:isDefinedBy rdf:resource="http://sws.geonames.org/2972315/about.rdf"/>
    <gn:name>Toulouse</gn:name>
    <gn:alternateName xml:lang="am">ቱሉዝ</gn:alternateName>
    <gn:alternateName xml:lang="ko">툴루즈</gn:alternateName>
    <gn:alternateName xml:lang="th">ตูลูซ</gn:alternateName>
    <gn:alternateName xml:lang="ja">トゥールーズ</gn:alternateName>
    <gn:alternateName xml:lang="ms">Lapangan Terbang Blagnac</gn:alternateName>
    <gn:alternateName xml:lang="an">Tolosa</gn:alternateName>
    <gn:alternateName xml:lang="br">Tolosa</gn:alternateName>
    <gn:alternateName xml:lang="co">Tolosa</gn:alternateName>
    <gn:alternateName xml:lang="es">Tolosa</gn:alternateName>
    <gn:alternateName xml:lang="frp">Tolosa</gn:alternateName>
    <gn:alternateName xml:lang="gl">Tolosa</gn:alternateName>
    <gn:alternateName xml:lang="it">Tolosa</gn:alternateName>
    <gn:alternateName xml:lang="la">Tolosa</gn:alternateName>
```

Toulouse's parent feature URI in Geonames

```
<gn:parentADM2 rdf:resour
<gn:parentADM3 rdf:resour
<gn:parentADM4 rdf:resource="http://sws.geonames.org/6453974/"/>
<gn:nearbyFeatures rdf:resource="http://sws.geonames.org/2972315/nearby.rdf"/>
<gn:locationMap rdf:resource="http://www.geonames.org/2972315/toulouse.html"/>
<gn:wikipedi
```

Examples of Links between datasets



Examples of Links between datasets

Geonames resource Toulouse (ADM 4)

```
<gn:parentADM... rdf:resource="http://sws.geonames.org/...157.../"/>
<gn:parentADM3 rdf:resource="http://sws.geonames.org/2972314/"/>
<gn:childrenFeatures rdf:resource="http://sws.geonames.org/6453974/contains.rdf"/>
<gn:locationMap rdf:resource="http://www.geonames.org/6453974/toulouse.html"/>
<owl:sameAs rdf:resource="http://id.insee.fr/geo/commune/31555"/>
</gn:Feature>
– <foaf:Document rdf:about=...rdf">
<foaf:primaryTopic rdf:r...74/"/>
...licen... rdf:reso...ce="http://creativecommons.org/licenses/by/3.0/"/>
```

Un lien **owl:sameAs** permet de relier les deux jeux de données

http://sws.geonames.org/6453974

owl:sameAs

http://id.insee.fr/geo/commune/31555

DataLift — Insee — Institut national de la statistique et des études économiques — Mesurer pour comprendre

INSEE resource for Toulouse

Description of node <http://id.insee.fr/geo/commune/31555>

Search for [Search] in [All columns]

Subject	Predicate	Object
http://id.insee.fr/geo/commune/31555	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://rdf.insee.fr/def/geo#Commune
http://id.insee.fr/geo/commune/31555	http://www.w3.org/2002/07/owl#sameAs	http://data.ign.fr/id/geofla/commune/31555
http://id.insee.fr/geo/commune/31555	http://rdf.insee.fr/def/demo#population	http://id.insee.fr/demo/populationLegale/commune/31555/2010
http://id.insee.fr/geo/commune/31555	http://rdf.insee.fr/def/demo#population	http://id.insee.fr/demo/populationLegale/commune/31555/2011
http://id.insee.fr/geo/commune/31555	http://rdf.insee.fr/def/demo#population	http://id.insee.fr/demo/populationLegale/commune/3155...
http://id.insee.fr/geo/commune/31555	http://rdf.insee.fr/def/demo#population	http://id.insee.fr/demo/populationLegale/commune/31555/2011
http://id.insee.fr/geo/commune/31555	http://rdf.insee.fr/def/demo#population	http://id.insee.fr/demo/populationLegale/commune/3155...
http://id.insee.fr/geo/commune/31555		
http://id.insee.fr/geo/commune/31555	http://rdf.insee.fr/def/geo#codeCommune	
http://id.insee.fr/geo/commune/31555	http://rdf.insee.fr/def/geo#nom	"Toulouse"

It could become even more powerful

- We could add extra knowledge to the merged datasets
  - geographical information
  - a full classification of various types of library data (novel, fiction, travel, history…)
  - etc.
- This is where ontologies, extra rules, etc, come in
  - ontologies/rule sets can be relatively simple and small, or huge, or anything in between…
- Even more powerful queries can be asked as a result

TP RDF

- Modéliser en RDF
- La table suivante montre un échantillon des données que la librairie ARTEMIS Bookstore conserve sur les livres qu'elle possède en stock. Ces données sont pour le moment stockées dans un tableur, il s'agit de les représenter en utilisant le modèle de données RDF.
- données artemis excel

Un peu de pratique

LastName	FirstName	BookTitle	PagesNb	Isbn	PublisherName
Abbey	Edward	Black Sun (Edward Abbey novel)	159	978-1-55566-286-8	Simon & Schuster
Abbey	Edward	Fire on the Mountain (Abbey novel	211	0-8263-0457-5	Dial Press
Abbey	Edward	Hayduke Lives	352	ISBN 0-316-00411-1 (first edition,	Little, Brown and Company
Abbey	Edward	Jonathan Troy	374	1-131-40684-2	Dodd, Mead and Company
Abbey	Edward	The Monkey Wrench Gang	352	ISBN 0-397-01084-2 (hardback ed	Lippincott Williams & Wilkins
Abbey	Edward	The Brave Cowboy	277	0-8263-0448-6	Dodd, Mead and Company
Abbey	Edward	The Fool's Progress	485	0-8050-0921-3	Henry Holt and Company
Acito	Marc	How I Paid for College: A Novel of	288	ISBN 0-7679-1841-X (first edition,	Broadway Books
Acker	Kathy	Blood and Guts in High School	165	ISBN 0-8021-3193-X (Paperback e	Grove Press
Adler	Warren	The War of the Roses (novel)	263	0-446-51220-6	Hachette Book Group
[illegible]	Will	[illegible]	[illegible]	[illegible]	[illegible]

Question 1: dessinez un graphe RDF qui reflète le contenu de la table (vous ne vous intéresserez qu'aux données contenues dans les deux premières lignes de la table).

Pour les URIs vous pouvez utiliser les préfixes suivants:

http://www.artemisbookstore.com/ontology# pour les ressources qui correspondent à une propriété ou une classe,

http://www.artemisbookstore.com/resource/ pour les autres ressources.

Un peu de pratique

- Question 2: Créez un fichier texte artemis.ttl dans le lequel vous écrirez une sérialisation de votre graphe en Turtle. Vérifiez que la syntaxe est correcte à l'aide de l'outil https://www.uuidgenerator.net/

RDF outline

- RDF Model
- Typed and localized literals
- RDF formats
- Blank nodes
- Resources definition
- RDF an data integration
- Persisting RDF
- References

Persisting RDF Data

- Small datasets (few triples)
  - RDF files stored locally or published on the web
  - Use any serialization format : `.nt`, `.rdf`, .ttl, etc …
- Large datasets (thousands to millions of triples
  - Database solutions better
  - Usually in form of RDF storage (triplestore) or Graph database
- Legacy data
  - Keep in original form but expose as RDF to the outer world
  - Provide a mapping to RDF.

Persiting RDF Data

- **Native RDF stores** (Triple stores)
  - implement their own database engine without reusing the storage and retrieval functionalities of other database management systems.

- **AllegroGraph** (commercial) RDF graph database and application framework developed byFranz Inc.
  - https://allegrograph.com/allegrograph/ , http://franz.com/agraph/allegrograph/
  - the free RDFStore server edition (< 50 million triples)
  - developer edition (< 600 million triples)
  - enterprise edition storage capacity only limited by the underlying server infrastructure (1+Trillion).
  - Clients connectors are available for Java, Python, Lisp, Clojure, Ruby, Perl, C#, and Scala.

Persiting RDF Data

- Native RDF stores (continued…)
- **GraphDB™** (commercial formerly OWLIM) - An Enterprise Triplestore with Meaning (GNU LGPL license and commercial) provided by Ontotext
  - http://ontotext.com/products/graphdb/
  - GraphDB™Lite, (Free)
  - GraphDB™ Standard Edition
  - GraphDB™ Enterprise.
- **Stardog** (commercial) an enterprise data unification platform built on smart graph technology: query, search, inference, and data virtualization.
  - http://stardog.com/
- **Apache Jena TDB** (open-source) - a component of the Jena Semantic Web framework (open-source java software – Apache foundation).
  - http://jena.apache.org/
- …

Persiting RDF Data

- DBMS-backed stores
  - RDF Stores that use the storage and retrieval functionality provided by another database management system.

  - Save triples into Relational Database
    - Various strategies: 1 giant table for alls triplet hexastore (create indexes for all possible combinations: spo, pos, osp, sop, pso, ops)
    - Building an Efficient RDF Store Over a Relational Database (Mihaela A. Bornea et al., SIGMOD '13 Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data) https://cs.uwaterloo.ca/~gweddell/cs848/papers/Bornea.pdf

Persiting RDF Data

- DBMS-backed stores (continued…)

- Apache Jena SDB (open-source) another component of the Jena Semantic Web framework
  - http://jena.apache.org/
  - Provides storage and query for RDF datasets using conventional relational databases: Microsoft SQL Server, Oracle 10g, IBM DB2, PostgreSQL, MySQL, HSQLDB, H2, and Apache Derby.

- Oracle Spatial and Graph (commercial, formerly Oracle Semantic Technologies) a W3C standards-based, full-featured graph store in Oracle Database for Linked Data and Social Networks applications.
  - http://www.oracle.com/technetwork/database/options/spatialandgraph/overview/rdfsemantic-graph-1902016.html
- Semantics Platform (commercial) - a family of products for building medium and large scale semantics-based applications using the Microsoft .NET framework.
  - http://www.intellidimension.com/products/semantics-platform/

- …

Persiting RDF Data

- Hybrid Stores
  - RDF Stores that supports both architectural styles (native and DBMS-backed).

- OpenLink Virtuoso Universal Server
  - http://virtuoso.openlinksw.com/
  - hybrid storage solution for a range of data models, including relational data, RDF and XML, and free text documents.
  - Through it unified storage it can be also seen as a mapping solution between RDF and other data formats, therefore it can serve as an integration point for data from different, heterogeneous sources.
  - used to host many important Linked Data sets (e.g., DBpedia),
  - open-source version – commercial

- Blazegraph (former Bigdata)(open-source and commercial license) is ultra-scalable, high-performance graph database with support for the Blueprints and RDF/SPARQL APIs. *https://www.blazegraph.com/product/*
- …

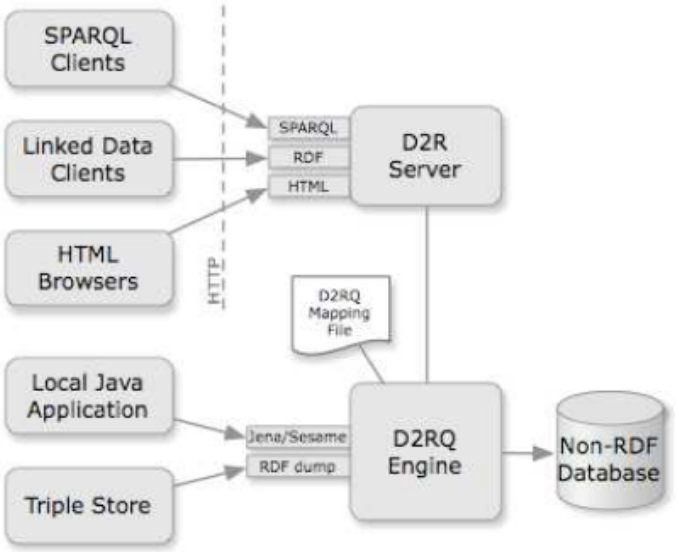
Persisting RDF Data

- Legacy systems Mapping RDF-relational databases
  - W3C RDB2RDF Working Group published two recommendations (september 2012)
    - R2RML: RDB to RDF Mapping Language, http://www.w3.org/TR/r2rml/
    - A Direct Mapping of Relational Data to RDF, http://www.w3.org/TR/rdb-direct-mapping/
  - DR2Q Accessing Relational Databases as Virtual RDF Graphs http://d2rq.org/

The D2RQ Platform consists of:

- **D2RQ Mapping Language**, a declarative mapping language for describing the relation between an ontology and an relational data model.
- **D2RQ Engine**, a plug-in for the Jena Semantic Web toolkit, which uses the mappings to rewrite Jena API calls to SQL queries against the database and passes query results up to the higher layers of the frameworks.
- **D2R Server**, an HTTP server that provides a Linked Data view, a HTML view for debugging and a SPARQL Protocol endpoint over the database.

Supported databases: Oracle, MySQL, PostgreSQL, SQL Server, HSQLDB, Interbase/Firebird



Persisting RDF Data
RDF Triple Stores

- W3C maintains a list of triplestores
  - http://www.w3.org/wiki/SemanticWebTools#RDF_Triple_Store_Systems
- Commercial:
  - Open Link Virtuoso - http://virtuoso.openlinksw.com
  - AllegroGraph - http://www.franz.com/agraph/allegrograph/
  - Ontotext GraphDB (SwiftOWLIM) : http://www.ontotext.com/products/ontotext-graphdb-owlim/
  - …
- Open source
  - Apache Jena (TDB) - http://jena.apache.org
  - Sesame - http://www.openrdf.org
  - Parliament – http://parliament.semwebcentral.org
  - …

RDF outline

- RDF Model
- Typed and localized literals
- RDF formats
- Blank nodes
- Resources definition
- RDF and data integration
- Persisting RDF
- References

References

- RDF is part of W3C Semantic Web W3C activity

"The Resource Description Framework (RDF) is a framework for representing information in the Web." [1]

  - W3C recommendation February 2004 (RDF 1.0)
  - Updaded February 2014 (RDF 1.1)

RDF Working Group

Recommandations

- RDF 1.1 Concepts and Abstract Syntax [1] } RDF model
- RDF 1.1 Semantics
- JSON-LD 1.0
- JSON-LD 1.0 Processing Algorithms and API
- RDF 1.1 Turtle
- RDF 1.1 TriG
- RDF 1.1 N-Triples
- RDF 1.1 N-Quads
- RDF 1.1 XML Syntax

} RDF serialization formats

Notes

- RDF 1.1 Primer [2]
- What's new in RDF 1.1
- RDF 1.1: On Semantics of RDF Datasets
- RDF 1.1 Test Cases
- RDF 1.1 JSON Alternate Serialization (RDF/JSON)

RDFa Working Group

Recommandations

- RDFa Core 1.1 - Second Edition
- XHTML+RDFa 1.1 - Second Edition
- HTML+RDFa 1.1
- RDFa Lite 1.1

Notes

- RDFa 1.1 Primer - Second Edition [3]
- Linked Data Glossary [4]
- HTML Data Guide

[1] http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/
[2] http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/
[3] http://www.w3.org/TR/2013/NOTE-rdfa-primer-20130822/
[4] http://www.w3.org/TR/2013/NOTE-ld-glossary-20130627/