

MODÈLE DE DONNÉES RDF

→ LIER DES RESSOURCES DU WEB PAR DES PROPRIÉTÉS

1. Modèle de triplets et graphe
2. Syntaxe de sérialisation
3. Groupes

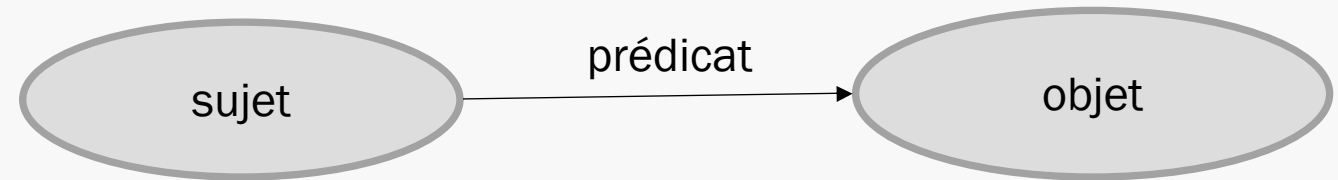
RDF (Resource Description Framework)



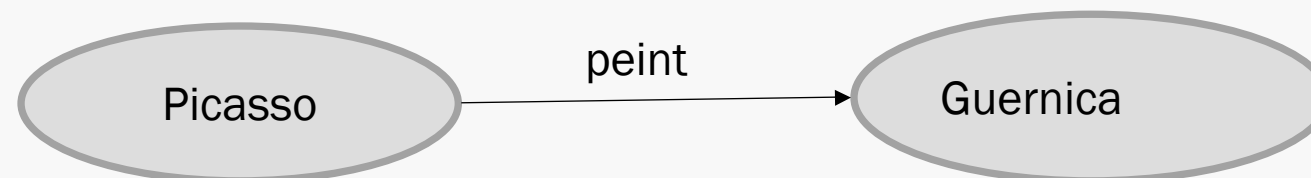
- Standard recommandation du W3C (1999) pour la description des ressources
- Objectif: permettre de structurer l'information accessible sur le web et de l'indexer efficacement.
- Recommandations W3C:
 - 1999: *RDF Model and Syntax Specification Recommendation*
 - 2004: *RDF Vocabulary Description Language 1.0: RDF Schema*
 - 2004: *RDF Semantics*

Modèle RDF

- RDF décompose chaque élément d'information en triplets reliant une ressource à une propriété et une valeur.
(sujet, prédicat, objet):
 - *sujet* : une ressource qui peut être identifiée par un URI (ou ressource anonyme)
 - *prédicat* : une propriété ;
 - *objet* : une ressource (identifiée par un URI ou anonyme) ou constante (littéral) à laquelle le sujet est lié.



Ex: Picasso a peint Guernica



Modèle RDF

– Exemple

« *http://www.example.org/index.html* a pour auteur John Smith et a pour thème le langage RDF »

Combien de triplets?

http://www.example.org/index.html a pour auteur John Smith

http://www.example.org/index.html a pour thème le langage RDF

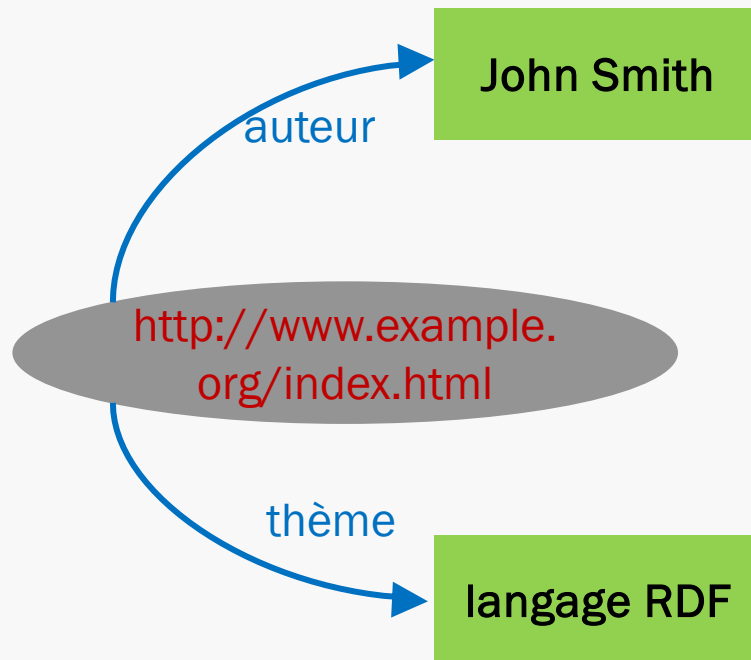
(*http://www.example.org/index.html* , auteur , John Smith)

(*http://www.example.org/index.html* , thème , langage RDF)

Modèle RDF: graphe

- Le triplet est représenté sous la forme d'un graphe:

(sujet, prédicat, objet) → (sujet, arc, objet)



(<http://www.example.org/index.html>, auteur, John Smith)

(<http://www.example.org/index.html>, thème, langage RDF)

Notation Graphique:

- Nœuds des ressources représentés par des ellipses
- Arcs orientés étiquetés représentés par des flèches avec un libellé
- Littéraux sont représentés comme des nœuds rectangulaires contenant la valeur littérale.

Modèle RDF: graphe

- Un graphe RDF est:
 - *Un multi-graphe: un graphe contenant plusieurs arcs et même des boucles entre deux mêmes sommets.*
 - *Un graphe orienté: chaque arc est orienté, allant du sommet représentant le sujet au sommet représentant l'objet.*
 - *Un graphe étiqueté: RDF attribue une étiquette à chaque arc et à chaque sommet du graphe. Les arcs sont étiquetés avec des URI identifiant le type de la relation décrite entre deux sommets. Les sommets sont étiquetés avec des URI, des valeurs littérales ou des « identificateurs » anonymes.*

Modèle RDF: graphe

■ Exercice:

Soit le texte suivant:

« le cours webDonnees.html a pour auteur Prof, a pour sujet le Web des données, a pour type text et a été écrit en 2023 ».

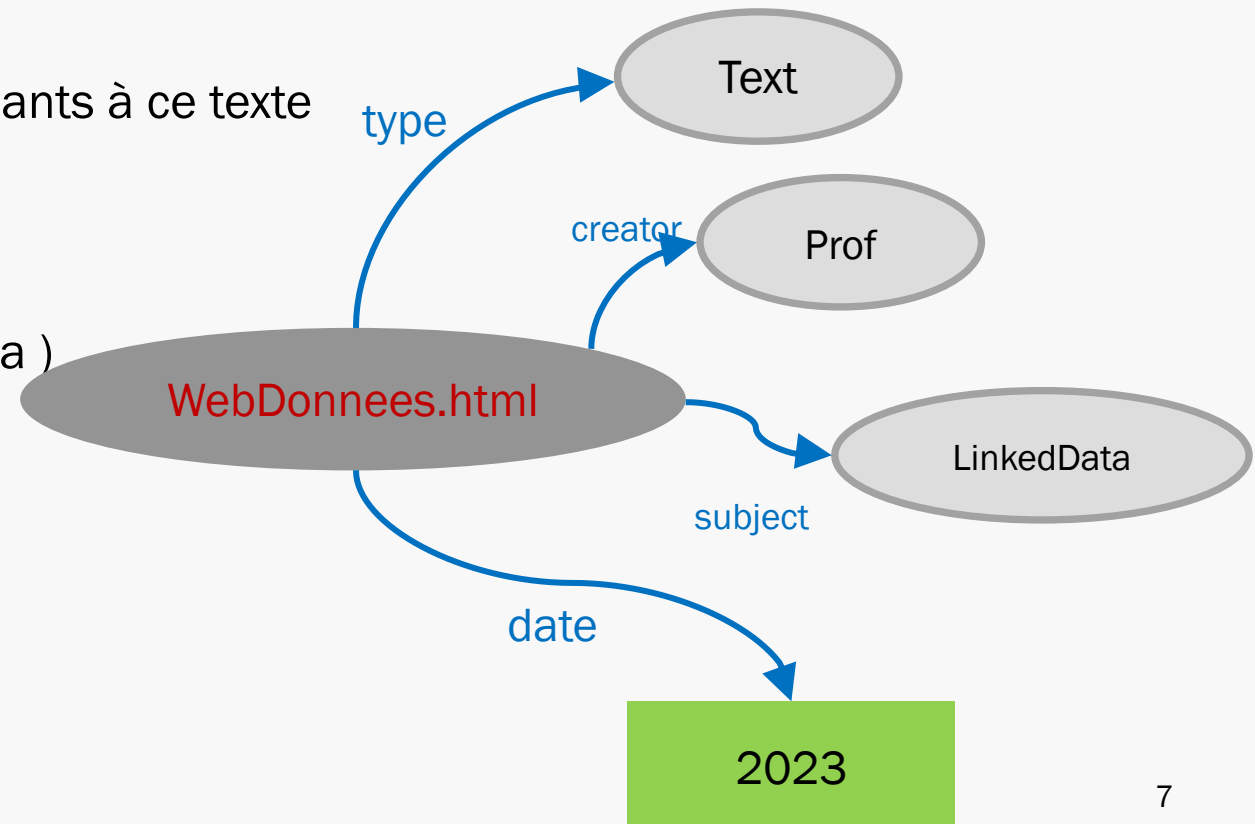
1. Déterminer les triplets correspondants à ce texte

(webDonnees.html, type, Text)

(webDonnees.html, creator, Prof)

(webDonnees.html, subject, LinkedData)

(webDonnees.html, date, "2023")



Modèle RDF: graphe

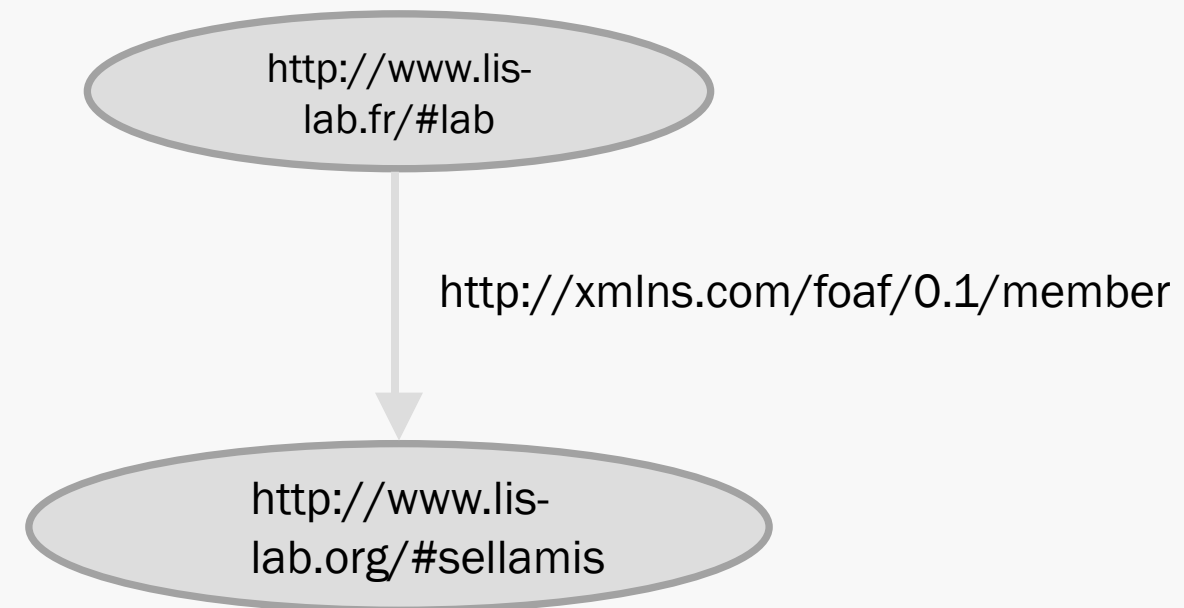
- Les **sujets** et les **prédicats** sont des URI, les objets sont soit des URI, soit des valeurs littérales (chaînes de caractères)

- Exemple

- *Le laboratoire LIS (sujet)*
- *a pour membre (prédicat)*
- *Sana Sellami (objet)*

- Nommage par des URIs:

- <http://www.lis-lab.fr/#lab>
- <http://xmlns.com/foaf/0.1/member>
- <http://www.lsis.org/#sellamis>



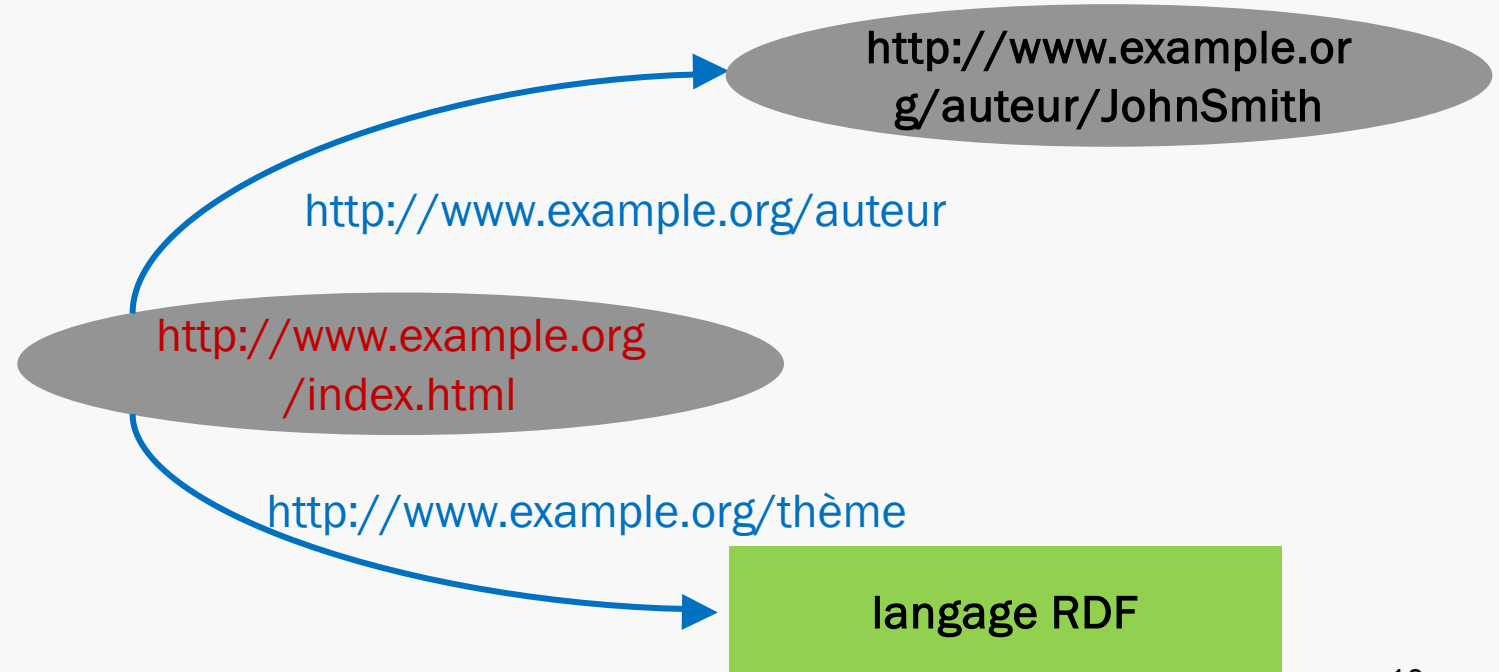
les sujets sont des URI,
les propriétés sont des URI
les objets peuvent être des URI

URI (Uniform Resource Identifier)

- Une URI représente un objet unique (une personne, un lieu, un livre, etc).
- Chaque URI identifie une ressource `http://api.domain.com/users/hugo`
- Exemple: Représenter une personne sur le Web
 - la personne elle-même : *`http://www.example.org/john#Jo` ; **URI***
 - le document RDF de cette personne : *`http://www.example.org/jo` ; **URI***
 - le document HTML de cette personne : *`http://www.example.org/jo.html`. **URL***
- **Remarque:** même si les URIs ont la forme `http://www..` Cela ne signifie pas qu'il s'agisse d'une URL mais tout simplement d'une convention utilisée pour uniformiser les URI (il n'est pas nécessaire d'avoir un document sur le Web pour chaque URI).

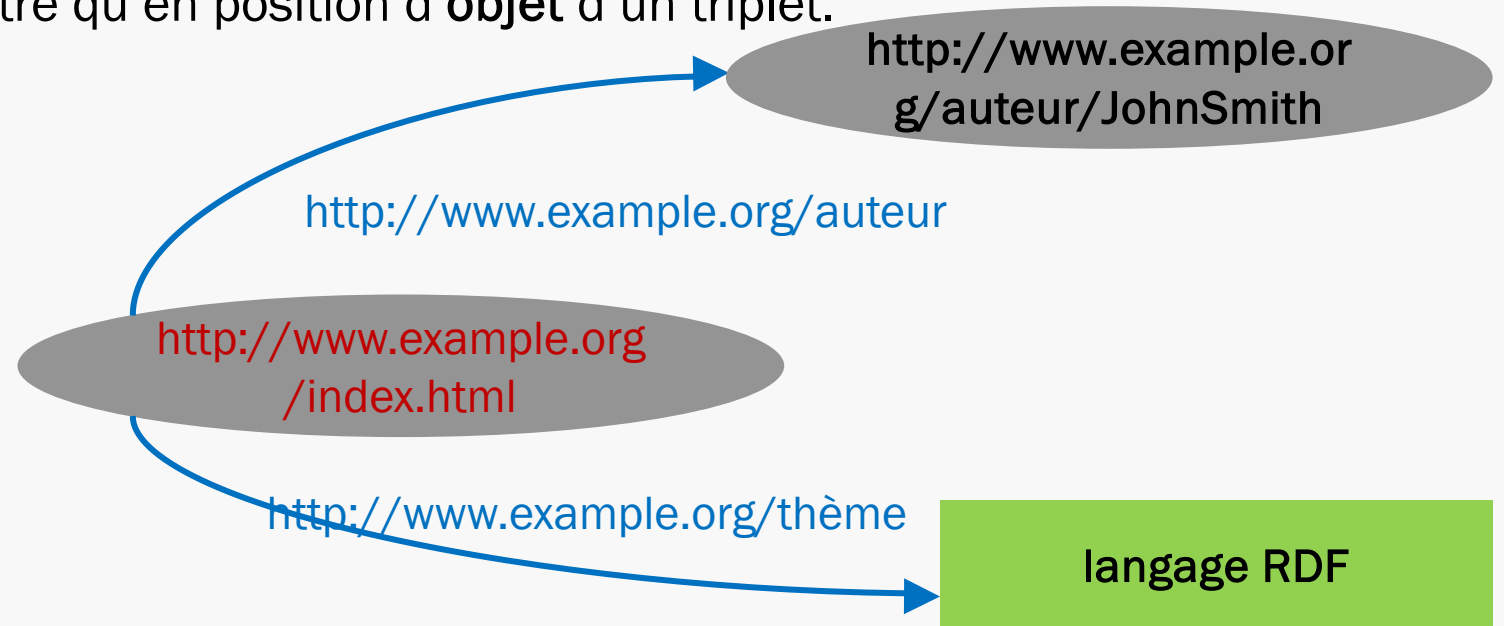
RDF graphes étiquetés par des URI et des littéraux

- (*<http://www.example.org/index.html>* , *<http://www.example.org/auteur>* , *<http://www.example.org/auteur/John Smith>*)
- (*<http://www.example.org/index.html>* , *<http://www.example.org/theme>* , *langage RDF*)



Les littéraux

- Un littéral consiste en des valeurs de base qui ne sont pas des IRI.
- Exemple:
les chaînes de caractères telles que « langage RDF », des dates telles que « 2023 » et des nombres tel que « 3.14159 ».
- Les littéraux sont associés à un type de données permettant à ces valeurs d'être traitées et interprétées correctement.
- Les littéraux ne peuvent apparaître qu'en position d'**objet** d'un triplet.



Les littéraux

- Pour typer les valeurs littérales, RDF repose sur les **datatypes** de XML Schema (`xmlns:xsd='http://www.w3.org/2001/XMLSchema#'`)

- Les datatypes sont :
 - *xsd:integer*
 - *xsd:float*
 - *xsd:string*
 - *xsd:dateTime*
 - *xsd:boolean*
 - *rdf:XMLLiteral*
 - *rdfs:Literal*

Espaces de noms

- Les URI sont utilisés pour identifier les ressources
- Cas particulier: l'URI utilisé pour identifier un ensemble de termes, un vocabulaire, un schéma → espaces de noms
- **Exemple:** la relation *type*:
 - permet de déclarer qu'une ressource appartient à une certaine classe.
 - Le prédicat *type* appartient au vocabulaire RDF qui est identifié par l'URI: <https://www.w3.org/1999/02/22-rdf-syntax-ns#> qui est souvent associé au préfixe *rdf*.
 - Le prédicat *type* peut être identifié par *rdf:type* au lieu de <https://www.w3.org/1999/02/22-rdf-syntax-ns#>

Espaces de noms

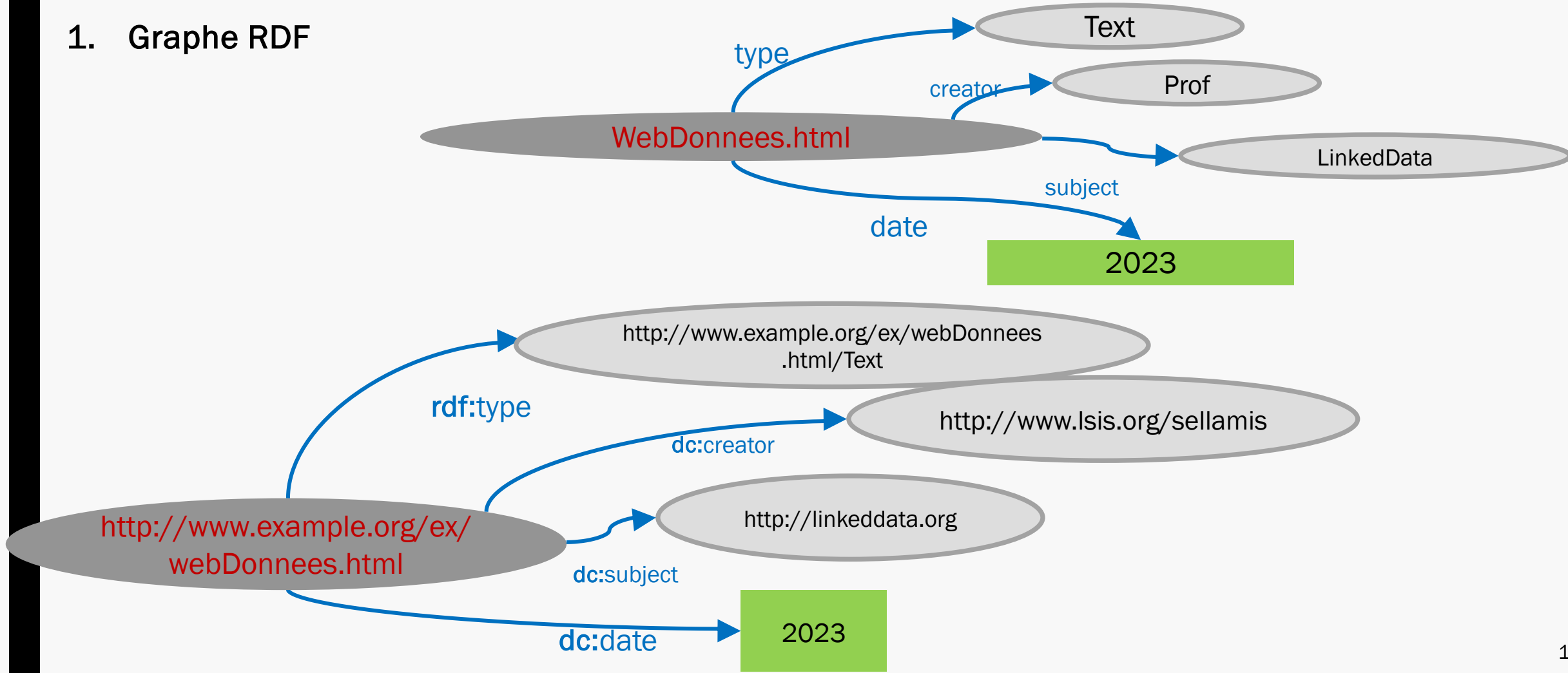
- Espaces de noms utilisés dans ce cours et leurs préfixes

Préfixe	URI de l'espace de noms
rdf	https://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	https://www.w3.org/2000/01/rdf-schema#
foaf	http://xmlns.com/foaf/0.1/ (un vocabulaire pour décrire les personnes)
dc	http://purl.org/dc/elements/1.1 (métadonnées de documents)
dt	http://purl.org/dc/dcmitype/ (types de documents)
xsd	http://www.w3.org/2001/XMLSchema#
data	http://www.example.org/ex/ (données d'exemples)

Espaces de noms

- Exercice: Modifier l'exercice précédent en ajoutant l'espace de noms et les préfixes

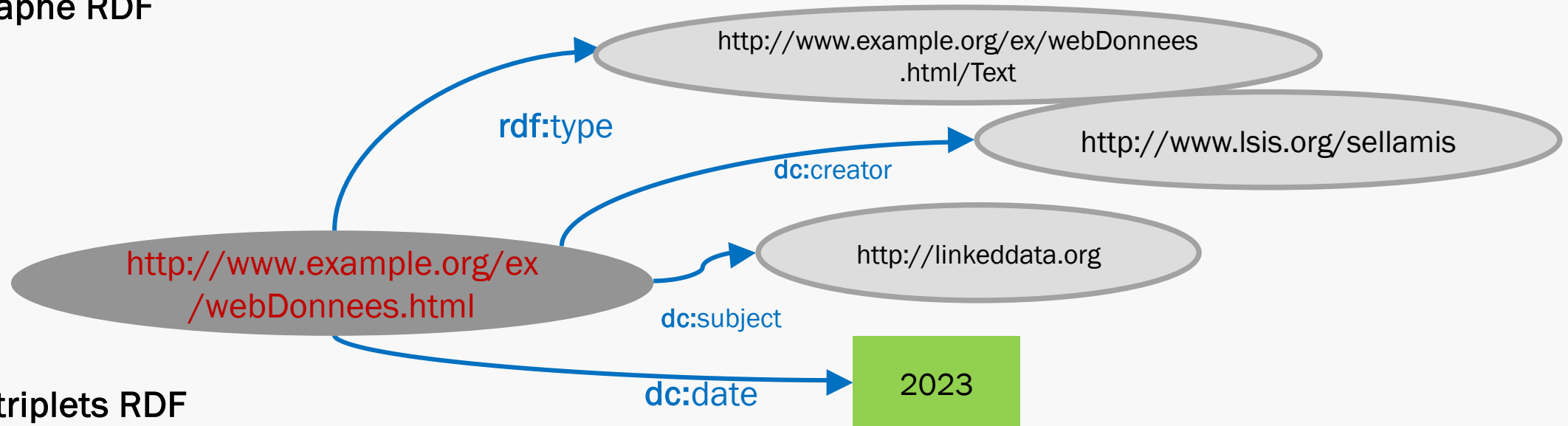
1. Graphe RDF



Espaces de noms

- Exercice: Modifier l'exercice précédent en ajoutant l'espace de noms et les préfixes

1. Graphe RDF



2. Les triplets RDF

(`http://www.example.org/ex/webDonnees.html`, `rdf:type`, `http://www.example.org/ex/webDonnees.html/Text`)

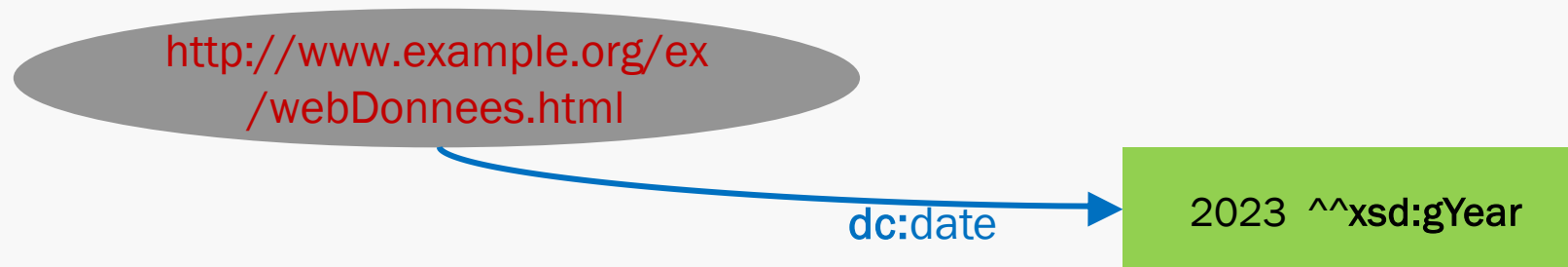
(`http://www.example.org/ex/webDonnees.html`, `dc:creator`, `http://www.lsis.org/sellamis`)

(`http://www.example.org/ex/webDonnees.html`, `dc:subject`, `http://linkeddata.org`)

(`http://www.example.org/ex/webDonnees.html`, `dc:date`, "2023")

Espaces de noms

- La date 2023 est de type xsd:gYear issu de la hiérarchie des types XML Schemas



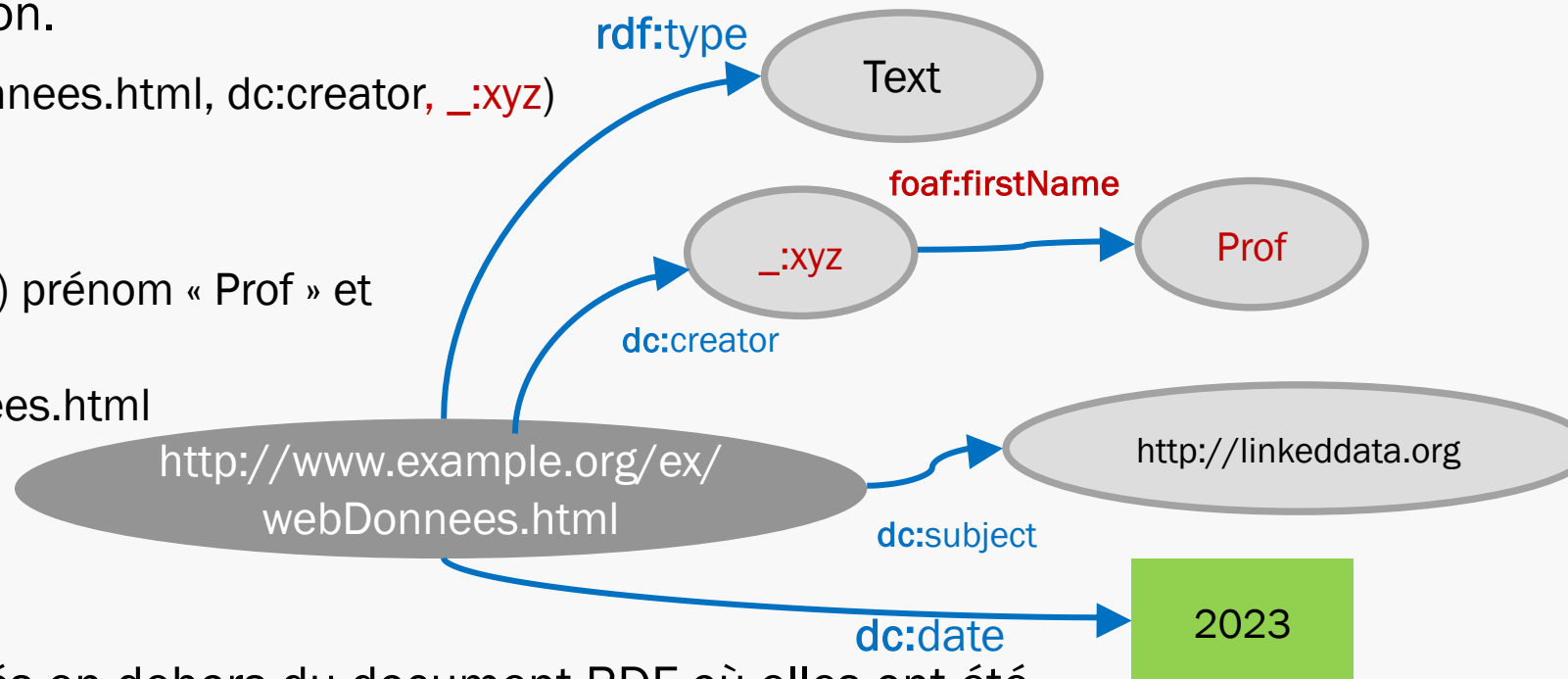
Nœuds anonymes

- Dans certains cas, une telle identification unifiée des ressources n'est pas possible
- Un nœud anonyme (ou nœud vide ou blank node bnode) représente, dans un graphe RDF, une ressource anonyme ou, plus simplement, une URI qui n'existe pas ou qui n'a pas d'identification.

(<http://www.example.org/ex/webDonnees.html>, dc:creator, **_:xyz**)
(**_:xyz**, foaf:firstName, "Prof")

Il existe une ressource qui a pour (foaf) prénom « Prof » et qui est un (dc) auteur de la ressource

<http://www.example.org/ex/webDonnees.html>

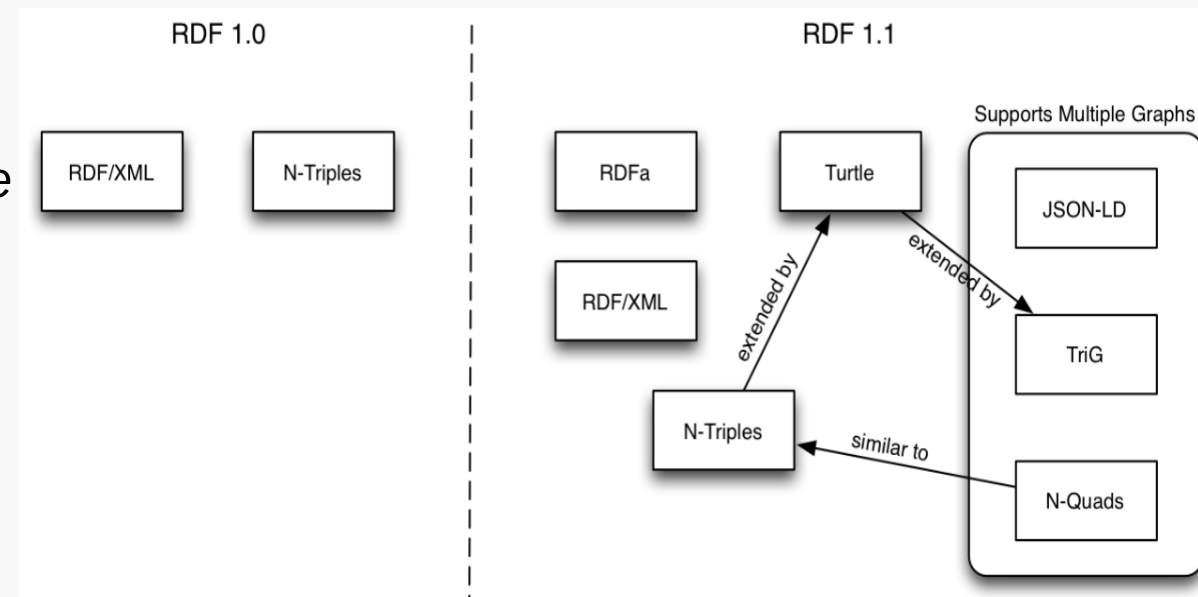


☹ Ne pouvant pas être réutilisés en dehors du document RDF où elles ont été déclarées, les ressources anonymes empêchent quiconque d'étendre leur description et de même simplement s'y référer.

RDF: Syntaxe

■ Plusieurs syntaxes pour RDF

- *RDF/XML: format standard, conçu pour une utilisation par des machines*
- *N3: Notation3 (triplets)*
- *N-triples*
- *Turtle: syntaxe plus concise et compacte*
- *JSON-LD*



RDF/XML

- Première syntaxe recommandée par le W3C (1999)
- basée sur XML
- relativement complexe et verbeuse

Syntaxe: <http://www.w3.org/TR/rdf-syntax-grammar/>

Valideur: <http://www.w3.org/RDF/Validator/>

RDF: RDF/XML Example

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:ex="http://example.org/stuff/1.0/" >
```

racine

```
<rdf:Description rdf:about="http://www.example.org/index.html">
```

Description des ressources

```
<ex:auteur rdf:resource="http://www.example.org/auteur/JohnSmith"/>
```

```
<ex:theme>langage RDF</ex:theme>
```

lien entre ressources

```
</rdf:Description>
```

Valeurs littérales

```
</rdf:RDF>
```

http://www.example.org/auteur/JohnSmith

http://www.example.org/auteur

http://www.example.org/index.html

http://www.example.org/theme

langage RDF

RDF: RDF/XML Exemple

- Élément `rdf:RDF` contenant les déclarations d'espaces de noms.
- Élément `rdf:Description` contient l'URI du sujet dans l'attribut `rdf:about`. Un élément RDF peut contenir plusieurs Descriptions.
- Élément prédicat ayant pour nom le nom du prédicat et contenant :
 - *un attribut `rdf:resource` objet*
- Plusieurs prédicats possibles dans une Description

RDF: RDF/XML Example

Validation Results

Your RDF document validated successfully.

Triples of the Data Model

Number	Subject	Predicate	Object
1	http://www.example.org/index.html	http://example.org/stuff/1.0/auteur	http://www.example.org/auteur/JohnSmith
2	http://www.example.org/index.html	http://example.org/stuff/1.0/theme	"langage RDF"

The original RDF/XML document

```
1: <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ex="http://example.org/stuff/1.0/" >
2:
3: <rdf:Description rdf:about="http://www.example.org/index.html">
4:
5: <ex:auteur rdf:resource= "http://www.example.org/auteur/JohnSmith"/>
6:
7: <ex:theme>langage RDF</ex:theme>
8:
9: </rdf:Description>
10:
11: </rdf:RDF>
```

Les nœuds vides

- Un nœud vide est représenté par une balise `rdf:Description` ne contenant aucun attribut `rdf:about` → une description qui ne précise pas la ressource décrite
- Intérêt: décrire des ressources sans les nommer
- Un nœud vide ne peut être que le sujet ou l'objet d'un triplet; le prédicat est toujours désigné par une URI.
- Quantification existentielle: Il existe une ressource telle que.....

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:ex="http://example.org/stuff/1.0/" >
```

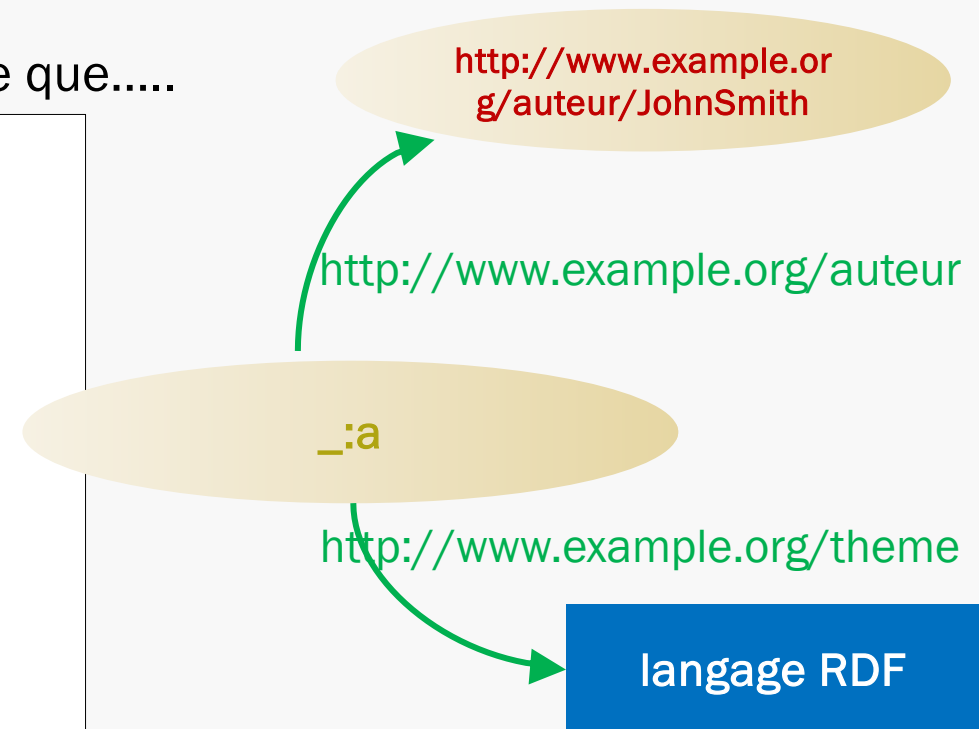
```
<rdf:Description>
```

```
<ex:auteur rdf:resource="http://www.example.org/auteur/JohnSmith"/>
```

```
<ex:theme>langage RDF</ex:theme>
```

```
</rdf:Description>
```

```
</rdf:RDF>
```



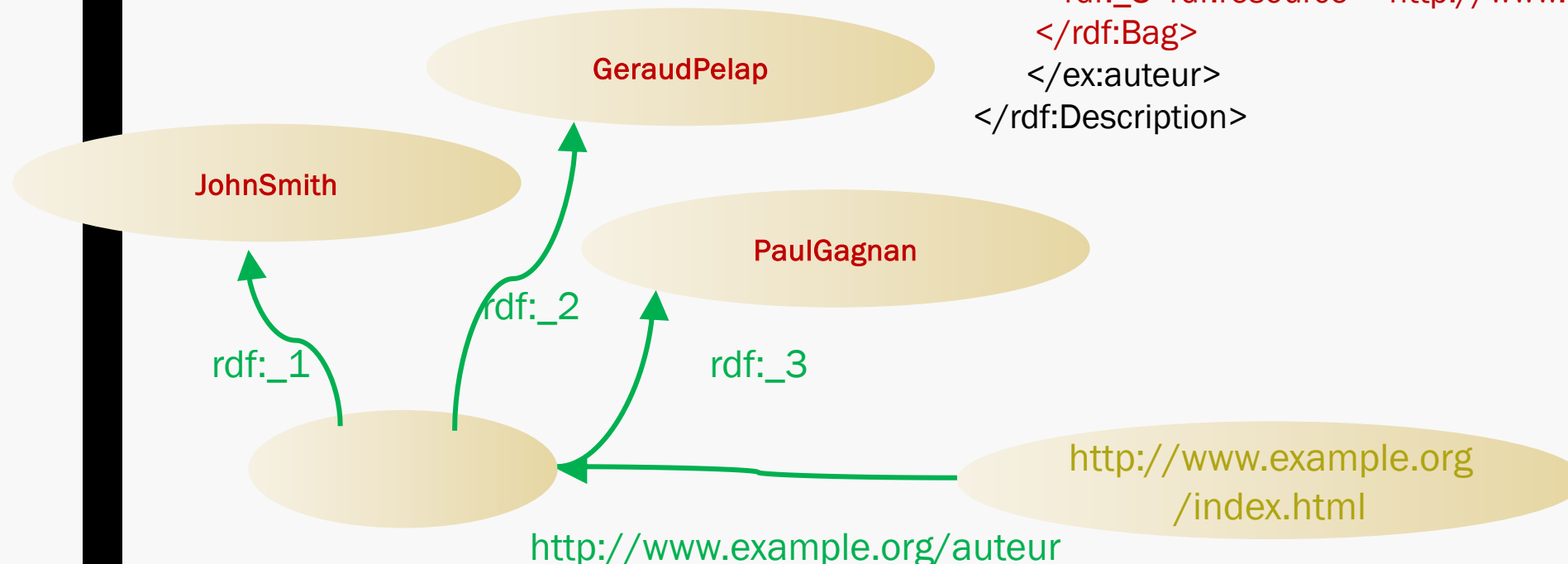
Les groupes (containers)

- Conteneur: une ressource contenant d'autres ressources
- Les éléments d'un container sont appelés membres du groupe.
- Il existe 3 types de containers prédéfinis :
 - *rdf:Bag* : multi-ensemble de ressources ou littéraux
 - *rdf:Seq* : séquence de ressources ou littéraux (l'ordre est important)
 - *rdf:Alt* : alternative entre plusieurs ressources ou littéraux

Les groupes (containers)

- Supposons que notre ressource possède 3 auteurs: John Smith, Gerard Pelap et Paul Gagnan

```
<rdf:Description rdf:about="http://www.example.org/index.html">  
<ex:auteur>  
  <rdf:Bag>  
    <rdf:_1 rdf:resource="http://www.example.org/auteur/JohnSmith"/>  
    <rdf:_2 rdf:resource="http://www.example.org/auteur/GeraudPelap"/>  
    <rdf:_3 rdf:resource="http://www.example.org/auteur/PaulGagnan"/>  
  </rdf:Bag>  
</ex:auteur>  
</rdf:Description>
```

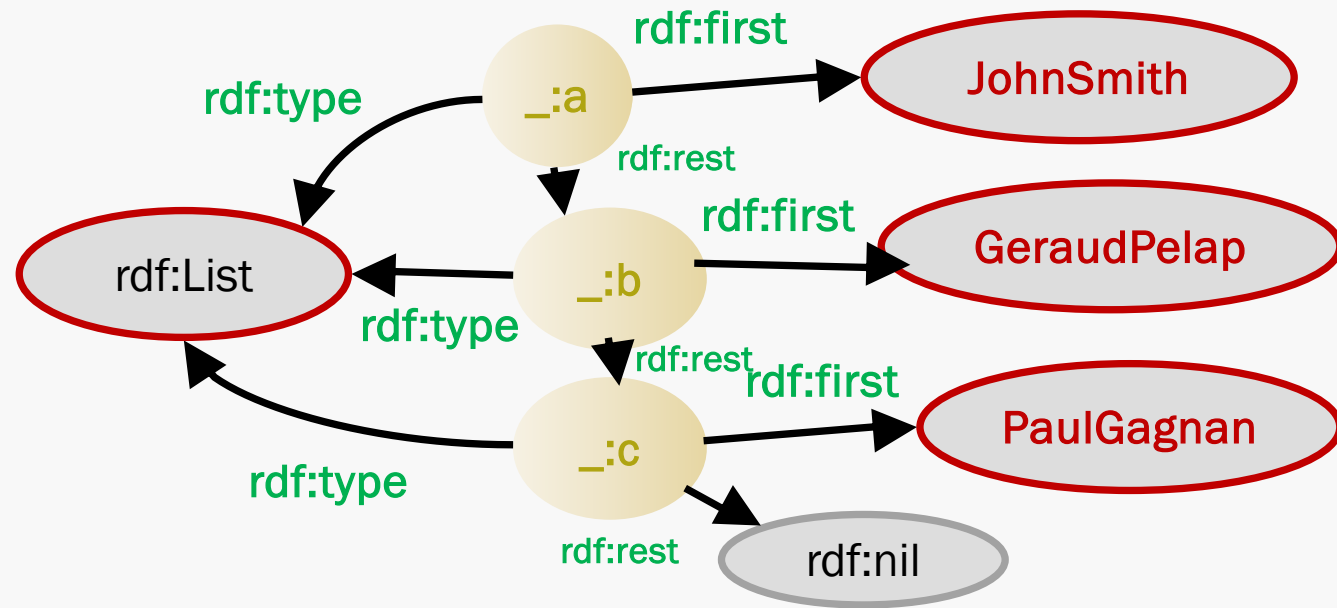


Les groupes (containers)

Construct	Description copied from [3]	Comments and examples
<code>rdf:Bag</code>	The <code>rdf:Bag</code> class represents RDF's 'Bag' container construct, and is a subclass of <code>rdfs:Container</code> .	Representing the mathematical set. $\{1, 2, 4, 8\}$ The order of its members is not important. There is so far no rule about duplicates. A multi set is therefore also a valid <code>rdf:Bag</code> . It depends on the application how to handle.
<code>rdf:Seq</code>	The <code>rdf:Seq</code> class represents RDF's 'Sequence' container construct, and is a subclass of <code>rdfs:Container</code> .	Representing a mathematical ordered list. Hence, $[1, 2, 2]$ and $[2, 2, 1]$ would not be the same.
<code>rdf:Alt</code>	The <code>rdf:Alt</code> class represents RDF's 'Alt' container construct, and is a subclass of <code>rdfs:Container</code> .	Representing an alternative, e.g., "female" or "male".

Les collections

- Une collection est une liste fermée : elle forme un groupe qui ne contient que les membres spécifiés lors de la déclaration de la collection.
- Définition des listes: balises `<rdf:list>`, `<rdf:first>`, `<rdf:rest>`
- La liste vide est définie par une ressource spéciale prédéfinie dont l'URI est `rdf:nil`
- Liste construite de manière récursive en utilisant le prédicat `rdf:first` (premier élément de la liste, `rdf:rest` (pour le reste de la liste) qui lui-même une liste



Les collections

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ex="http://example.org/stuff/1.0/" >

<rdf:Description rdf:about="http://www.example.org/index.html">
  <ex:auteur>
    <rdf:Description>
      <rdf:first rdf:resource="http://www.example.org/auteur/JohnSmith"/>
      <rdf:rest>
        <rdf:Description>
          <rdf:first rdf:resource="http://www.example.org/auteur/GeraudPelap"/>
          <rdf:rest>
            <rdf:Description>
              <rdf:rest rdf:resource="http://www.example.org/auteur/PaulGagnan"/>
            </rdf:Description>
          </rdf:rest>
        </rdf:Description>
      </rdf:rest>
    </rdf:Description>
  </rdf:rest>
</rdf:Description>
</ex:auteur>
</rdf:Description>
</rdf:RDF>
```

Les collections

- Syntaxe plus légère avec l'attribut `rdf:parseType="Collection"`
- Condition: la liste doit être l'objet d'un triplet

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:ex="http://example.org/stuff/1.0/" >

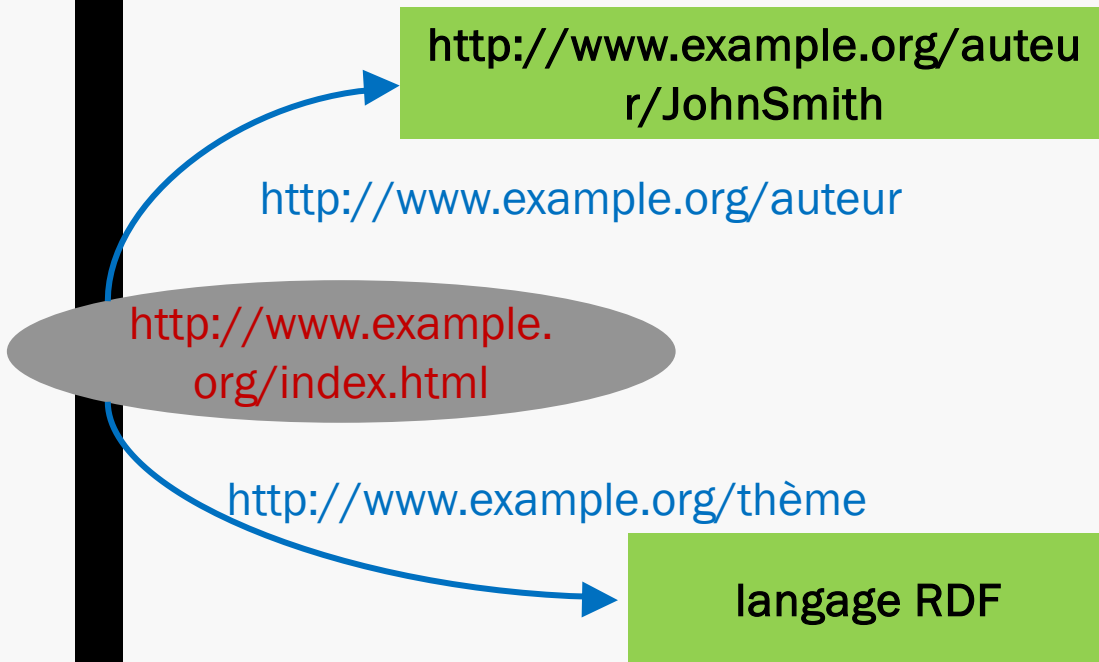
  <rdf:Description rdf:about="http://www.example.org/index.html">
    <ex:auteur rdf:parseType="Collection">
      <rdf:Description rdf:about="http://www.example.org/auteur/JohnSmith"/>
      <rdf:Description rdf:about="http://www.example.org/auteur/GeraudPelap"/>
      <rdf:Description rdf:about="http://www.example.org/auteur/PaulGagnan"/>
    </ex:auteur>
    <ex:theme>langage RDF</ex:theme>
  </rdf:Description>

</rdf:RDF>
```

RDF: N-Triples

- Une recommandation W3C depuis Février 2014
- Format texte brut qui sérialise un graphe RDF à raison d'un triplet par ligne du fichier texte et encodé uniquement en US-ASCII.
- La plus basique est la syntaxe N-triples, qui écrit chaque triplet sous la forme
 <URI du sujet> < URI du prédicat> < URI de l'objet ou littéral>
- Langage très simple et rapide à lire et écrire par programme.

RDF: N-Triples



```
< http://www.example.org/index.html >  
<http://www.example.org/auteur >  
<http://www.example.org/auteur/JohnSmith>.
```

```
< http://www.example.org/index.html >  
<http://www.example.org/theme > "langage RDF ".
```


RDF: Turtle : Terse RDF Triple Language

- Turtle (Terse RDF Triple Language) est, depuis février 2014, un standard approuvé par le W3C.
- Syntaxe très utilisée et a inspiré d'autres langages comme SPARQL
- Est un format textuel compact pour sérialiser un graphe RDF.
- Moins verbeux que N-Triples.
- La syntaxe Turtle reprend celle des N-triples en y ajoutant des facilités syntaxiques pour rendre le code plus lisible .

RDF: Turtle

The diagram illustrates the expansion of a URI. A central URI, `http://www.example.org/index.html`, is shown in a grey oval. Two blue arrows originate from this central URI. One arrow points to a green box containing the URI `http://www.example.org/auteur/JohnSmith`. The other arrow points to a green box containing the text `langage RDF`.

`http://www.example.org/auteur/JohnSmith`

`http://www.example.org/auteur`

`http://www.example.org/index.html`

`http://www.example.org/thème`

`langage RDF`

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
```

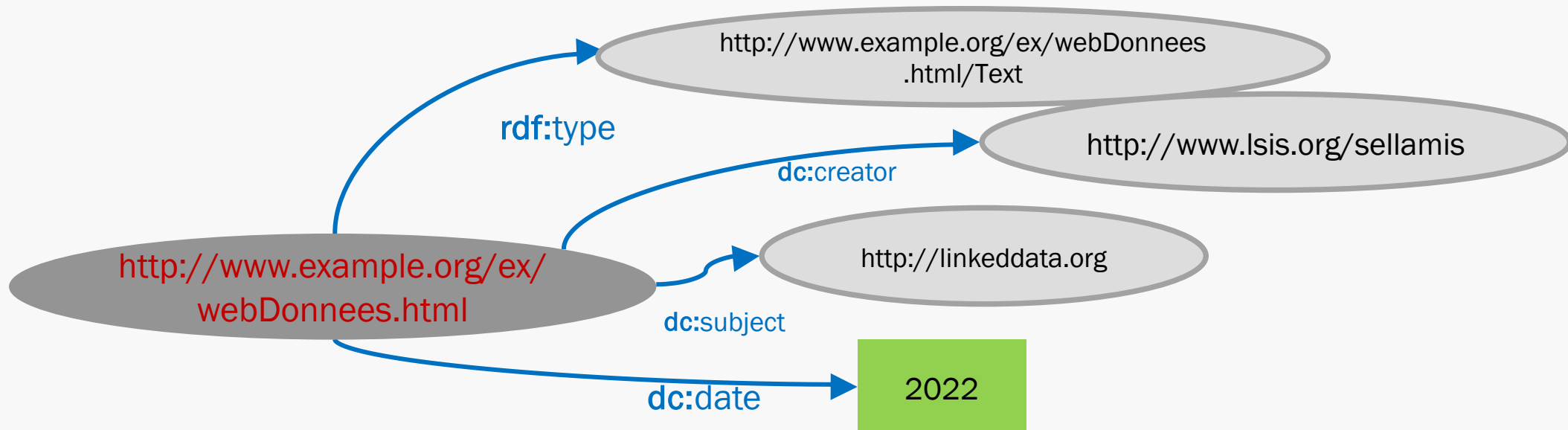
```
@prefix ex: <http://example.org/stuff/1.0/>.
```

```
<http://www.example.org/index.html >  
ex:auteur <http://www.example.org/auteur/JohnSmith>;  
ex:thème "langage RDF".
```

RDF: Turtle

■ Exercice:

Créer un document Turtle correspondant à l'exercice précédent



RDF: Turtle

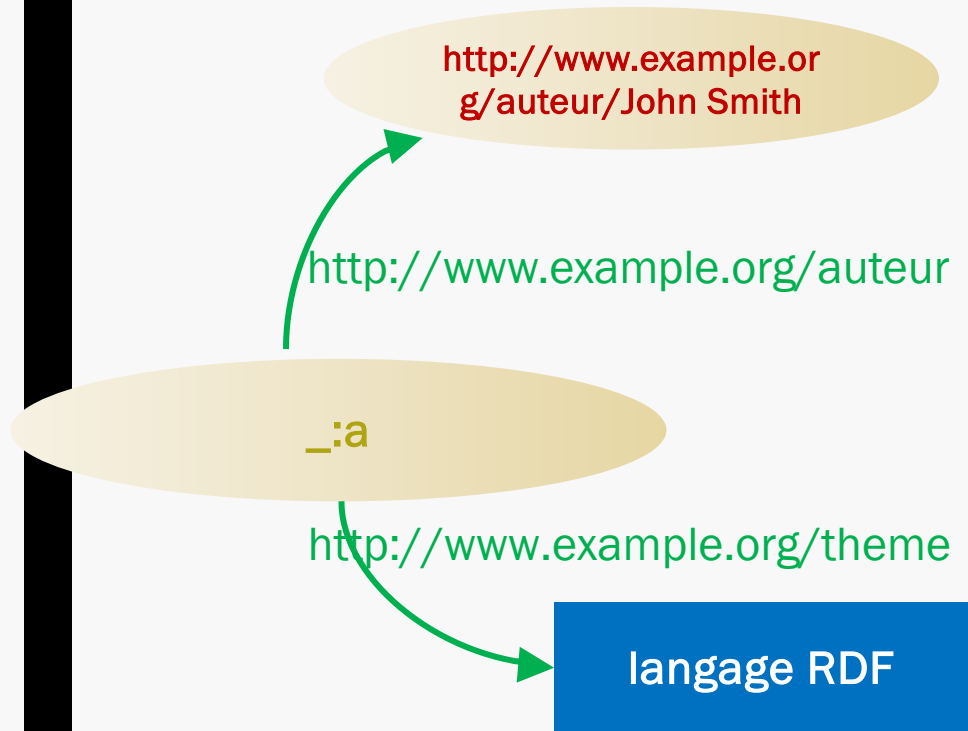
■ Exercice:

Créer un document Turtle correspondant à l'exercice précédent

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
  
<http://www.example.org/ex/webDonnees.html>  
a <http://www.example.org/ex/webDonnees.html/Text> ;  
  dc:creator <http://www.lsis.org/sellamis> ;  
  dc:subject <http://linkeddata.org> ;  
  dc:date "2022"^^xsd:year .
```

RDF: Turtle

■ Les nœuds vides



```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
```

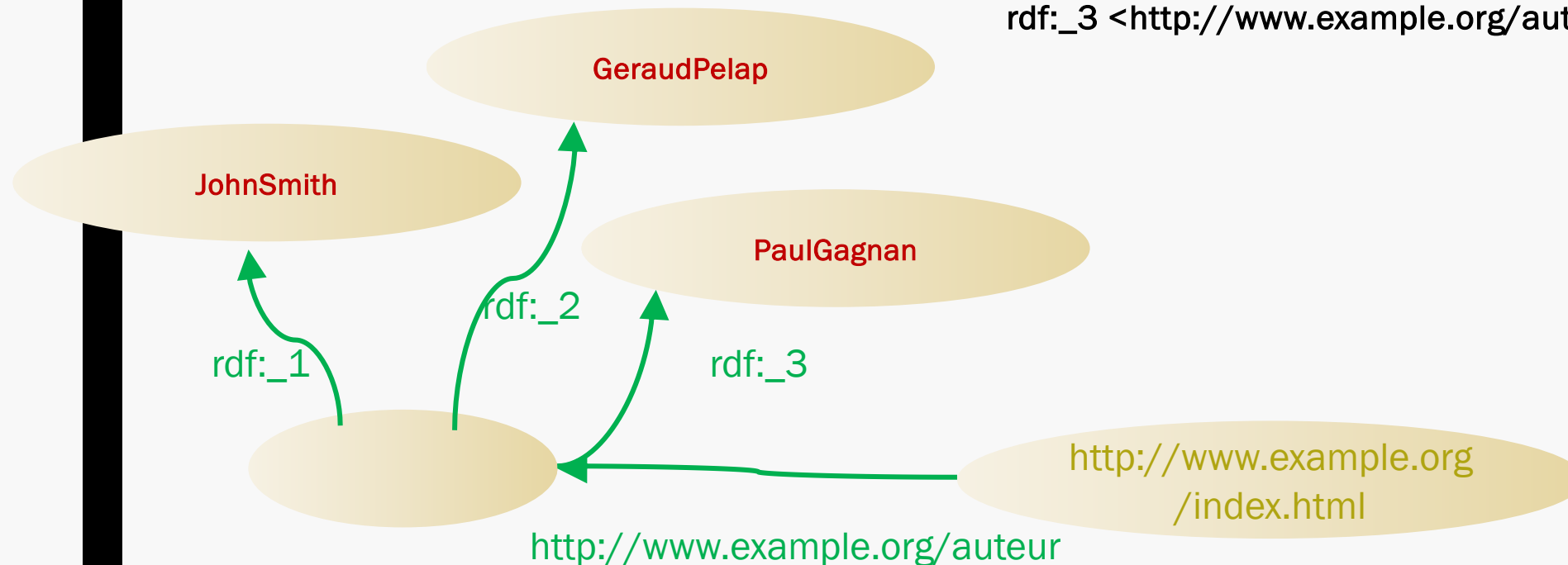
```
@prefix ex: <http://example.org/stuff/1.0/>.
```

```
[ ex:auteur  
  <http://www.example.org/auteur/JohnSmith>;  
  ex:thème "langage RDF"].
```

RDF: Turtle

■ Les conteneurs

```
<http://www.example.org/index.html> ex:auteur  
[ a rdf:Bag ;  
  rdf:_1 <http://www.example.org/auteur/JohnSmith> ;  
  rdf:_2 <http://www.example.org/auteur/GeraudPelap>  
  ;  
  rdf:_3 <http://www.example.org/auteur/PaulGagnan> ].
```

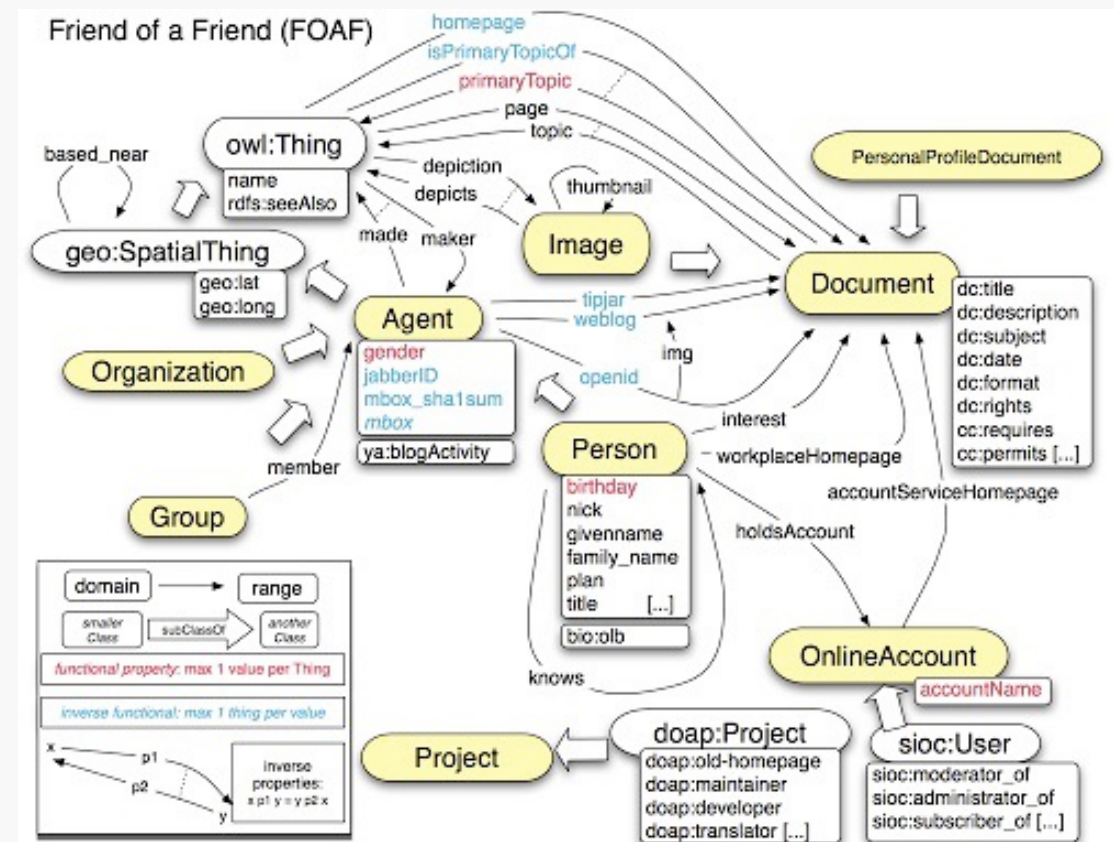


Vocabulaires RDF

- RDFS (RDF Schema): pour définir son propre vocabulaire
- Autres vocabulaires:
 - *FOAF (Friend of a friend) pour des relations sociales,*
 - *schema.org pour des moteurs de recherche, ...*
 - *dbpedia pour annoter wikipedia(<http://fr.dbpedia.org/>):*
 - Projet phare du web des données développé par l'université de Leipzig, l'université libre de Berlin et l'OpenLink software sous licence libre.
 - Produit depuis 2007 une source de données RDF extraites de Wikipedia.
 - Données librement accessibles sur le Web
 - Un accès aux données via un service de téléchargement de l'ensemble de la source et un service de requêtage SPARQL permettant une extraction des données.

RDF: foaf(friend of a friend)

- Un vocabulaire RDF qui permet de décrire de manière structurée une personne et ses relations.
- RDF peut contenir diverses informations:
 - *surnom* ;
 - *adresse* ;
 - *numéro de téléphone* ;
 - *profession*.



En résumé

- RDF permet que les informations sur les ressources soient manipulées par des applications
- RDF est un modèle flexible et extensible
- Utilisation de URI pour désigner les ressources décrites par le graphe RDF

Références

- Sites Web:
 - *Spécifications FOAF*: <http://xmlns.com/foaf/spec/>
 - *Recommandation RDF*: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>
 - *Tutoriel*: <http://www.linkdatatools.com/semantic-web-basics>
 - *Support de cours de Michel Gagnon*: <https://moodle.polymtl.ca/course/view.php?name=INF8410>
 - <https://www.w3.org/2001/12/semweb-fin/w3csw>
- Supports de cours
 - *Support de cours de Pierre Antoine Champin, Université de Lyon*
 - *Support de cours de Fabien Duchateau, Université de Lyon*
 - *Cours Web Sémantique Bernard Espinasse. Polytech Marseille*
 - *Cours de Bilasco inspiré de RDF Primer W3C*
 - *Cours de Odile Papini*
- Livres
 - *Fabien Gandon, Catherine Faron-Zucker, Olivier Corby. Le Web sémantique. Comment lier les données et les schémas sur le Web?*
- MOOC: MOOC Web sémantique et Web de données: <https://www.fun-mooc.fr/courses/course-v1:inria+41002+session03/about>
- Données RDF: Linked Open data cloud : <https://lod-cloud.net/>
 - <https://www.w3.org/wiki/DataSetRDFDumps>
 - <http://www.rdfhdt.org/datasets/>
 - DBPedia : <https://old.datahub.io/dataset/dbpedia>
 - LODLIVE (<http://fr.lodlive.it/>) : Le projet LodLive fournit un exemple de l'usage qui peut être fait des standards du web sémantique (RDF, SPARQL) pour parcourir librement les ressources RDF.