Infrastructure décentralisée et sémantique pour l'apprentissage tout au long de la vie

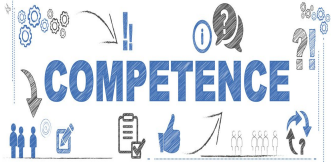
Hala Skaf-Molli, Patricia Serrano-Alvarado
Sara El hassad, Emmanuel Desmontils, Pascal Molli

GDD - LS2N
AWD 2019 - 22 Janvier 2019

Contexte

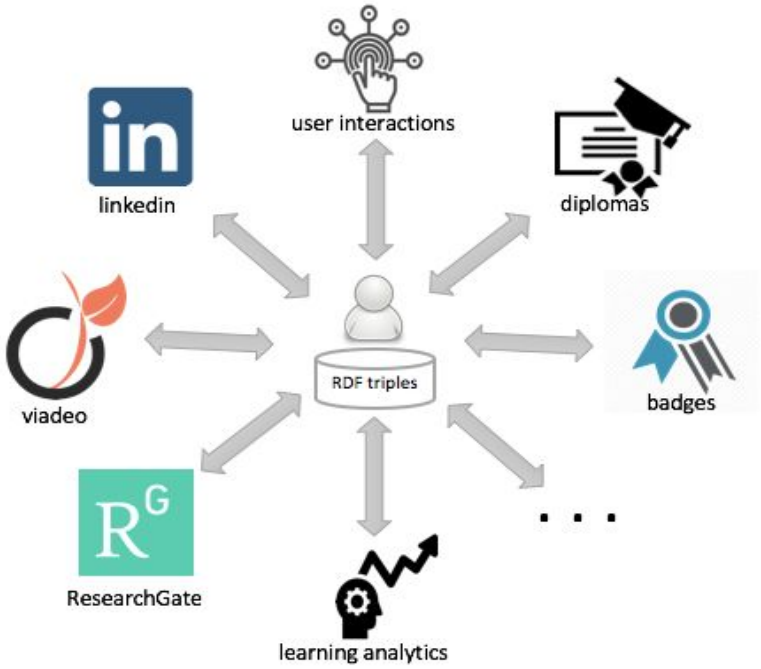
- L’apprentissage tout au long de la vie
  - de l’école jusqu’à l’université, et pendant toute la carrière
  - Données personnelles d'apprentissage conservées pendant longtemps
    - sous le contrôle de l'apprenant
  - Interaction avec différentes communautés

Données pour l'infrastructure



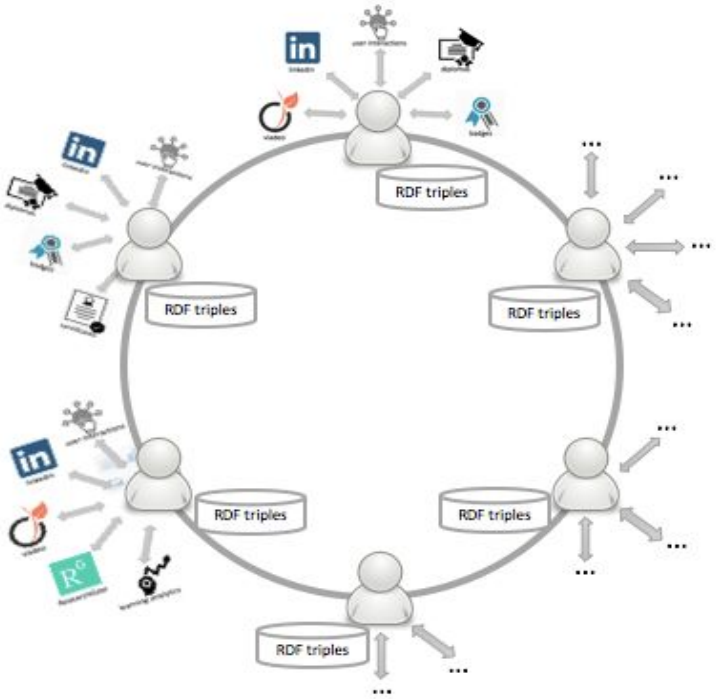
Première étape de la construction

Espace privé qui contient des données personnelles d'apprentissage



Deuxième étape de la construction

Collaborer avec les autres en confiance



Objective

Construire une **infrastructure de confiance** avec une **intégration flexible** des données avec un **contrôle d'usage personnalisé**.

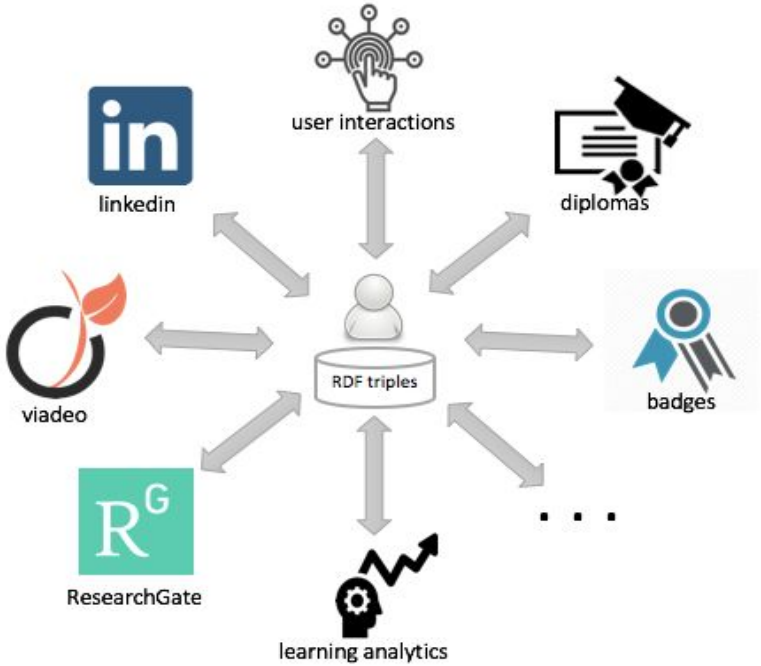
Etat de l'art

- Infrastructures client-serveur [1], [2], [3]
  ou Contrôle d'usage personnalisé limité [2]
  - Les données sont confiées à un tiers
- Infrastructure pair-à-pair [4], [5]
  - pas de contrôle d'usage

[1] https://www.linkedin.com/
[2] https://mahara.org/
[3] https://fr.viadeo.com/fr/
[4] Nejdl, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmér, M., Risch, T.: Edutella: a p2p networking infrastructure based on rdf. In: WWW (2012)
[5] Simon, B., Miklos, Z., Nejdl, W., Sintek, M., Salvachua, J.: Elena: A mediation infrastructure for educational services. In: WWW (2003)

Approche

Utiliser le Web sémantique pour construire l’**espace privé** et pour le **contrôle d’usage**



Pourquoi web sémantique ?

- **Les ontologies** permettent une intégration fluide des données d'apprentissage
  - CCSO[6]
- **SPARQL** permet d'interroger différentes sources de données

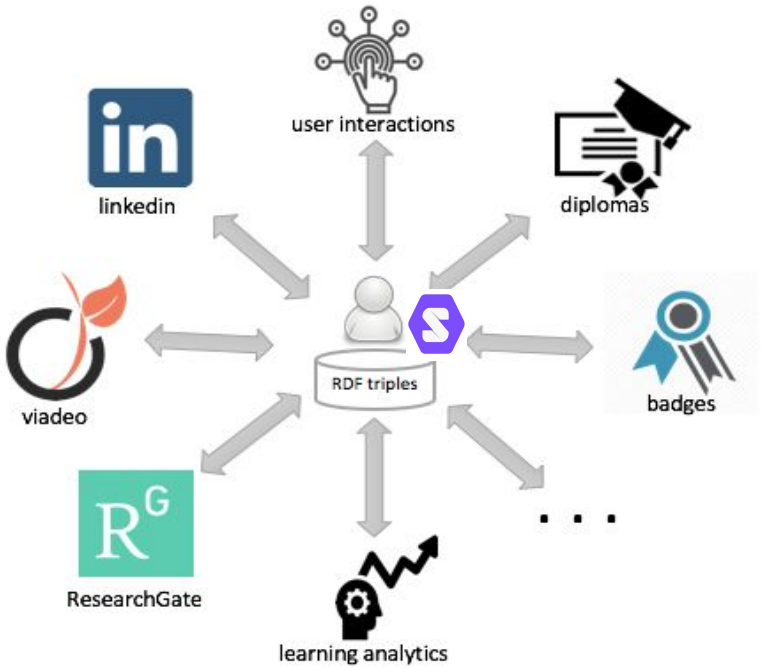
Learner ontologies

[6] Katis, E., H. Kondylakis, G. Agathangelos, and K. Vassilakis. Developing an ontology for curriculum & syllabus. In ESWC(2018)

Comment construire l’espace d’apprentissage privé?

Solid pour l'espace privé

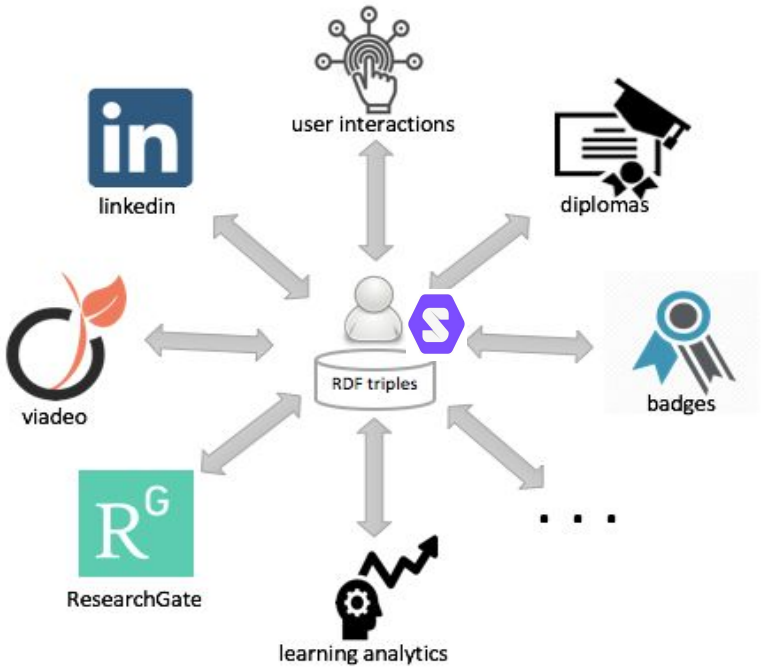
- Données transformées en RDF et stockées dans Solid [7]
  - Données sont accessibles online
  - Données gérées indépendamment de une application
  - Requêtes sur mes données



[7] Mansour, E., A. V. Sambra, S. Hawke, M. Zereba, S. Capadisli, A. Ghanem, A. Aboulnaga, et T. Berners-Lee (2016). A Demonstration of the Solid Platform for Social Web Applications. In WWW . Demo paper.

Exemple

```
select distinct ?d
where {
   :Me  edu:aDiplome  ?d .
   ?d  dc:title “Master” .
}
```



Pas possible de stocker toutes les données dans l'espace privé

- Exemple : le programme détaillé d'une formation stocké sur le site de l'université.
- Problème:
  - pas possible de répondre à certaines requêtes
  - Exemple :

```
select distinct ?c
where {
 :Me  edu:aDiplome   ?d .
 ?d  dc:title  "Master ALMA Nantes" .
 ?d  edu:programme  ?p
 }
```

- Besoin de faire l'intégration virtuelle de données

Intégration virtuelle

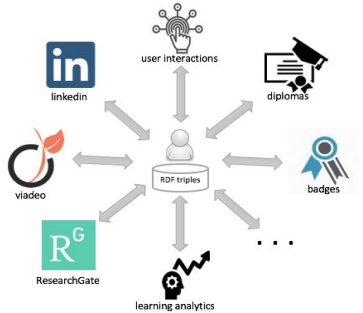
- Les données restent dans les sources et elles sont accessibles à la demande au moment de la requête
  - semLAV : Médiateur Local-As-View pour les requêtes SPARQL [9]
    - APIs accessibles pour interroger les données non-RDF [8].

[8] Moreau, B., P. Serrano-Alvarado, E. Desmontils, et D. Thoumas (2017). Queryingnon-RDF Datasets using Triple Patterns. In ISWC. Demo paper
[9] G. Montoya, L. D. Ibánez, H. Skaf-Molli, P. Molli, and M.-E. Vidal. SemLAV: Local-As-View Mediation for SPARQL. Transactions on Large-Scale Data- and Knowledge-Centered Systems, LNCS, Vol. 8420, p 33–58, 2014.

semLAV

médiateur

```
select distinct ?c
where {
  :Me  edu:aDiplome   ?d .
  ?d  dc:title “Master ALMA Nantes” .
  ?d  etu:programme  ?p
  }
```



-Diplômes
-title

UNIVERSITÉ DE NANTES

-Détails programme

Comment collaborer avec les autres en confiance?

Partage de données de confiance

- Chaque apprenant associe des politiques d'usage différente à ses données

```
<mypod/linkedin/sparql> a odrl:Asset ;
rdfs:comment "mes données collectés de LinkedIn" ;
odrl:hasPolicy <licence1> .
<mypod/badges/sparql> a odrl:Asset ;
rdfs:comment "mes badges numériques" ;
odrl:hasPolicy <licence2>
```

```
licence1 a odrl:Policy ;
rdfs:comment "licence1" ;
odrl:Permission
        [ odrl:action
         cc:read ; ] ;
odrl:prohibition
        [ odrl:action
        cc:archive ] .
```

```
licence2 a odrl:Policy ;
rdfs:comment "licence2" ;
odrl:Permission
        [ odrl:action ,
         cc:read ; ] ;
odrl:prohibition
        [ odrl:action
        cc:distribute ] .
```

Partage de données de confiance

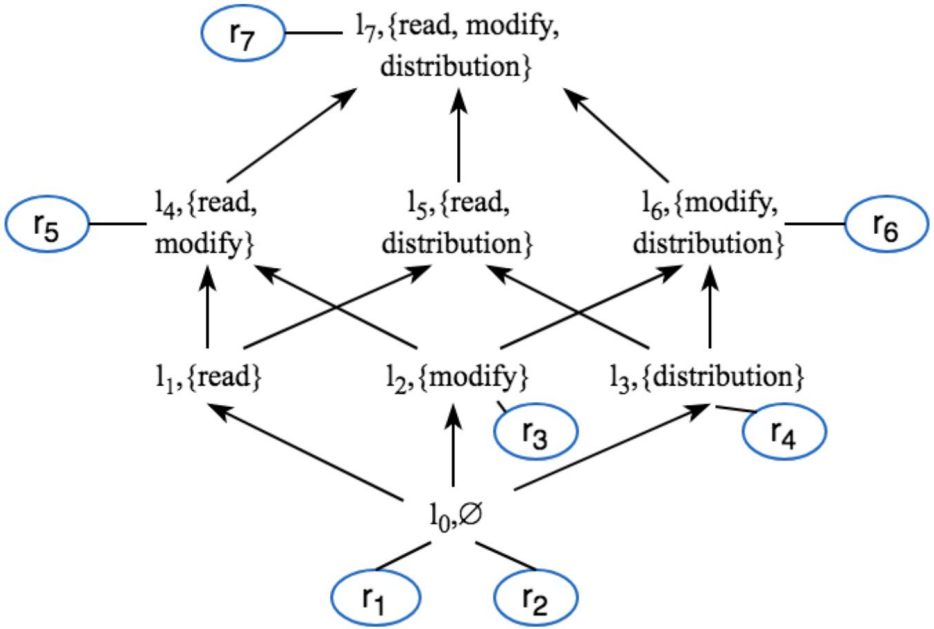


figure 3 : Représentation d'une classification des ressources basée sur les licences [10]

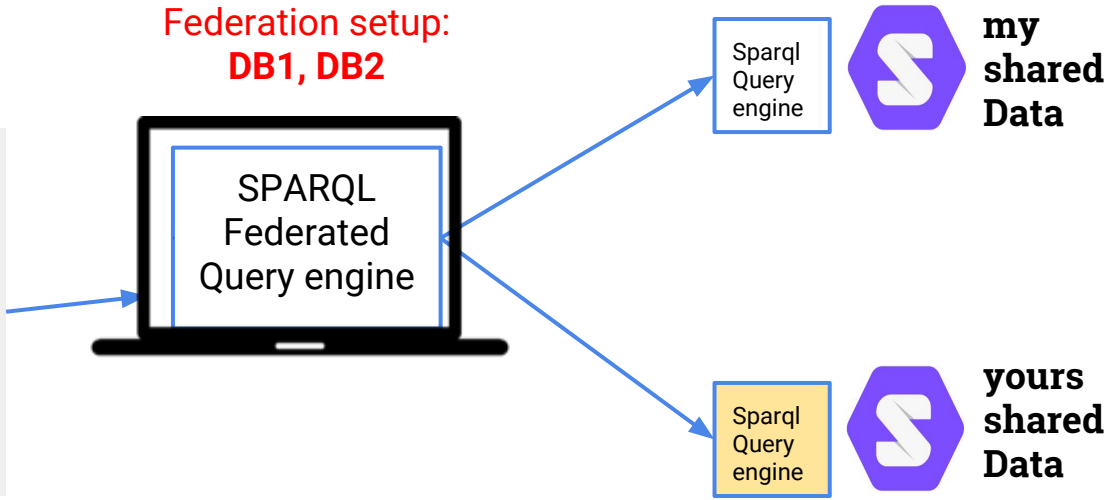
[10] Moreau, B., P. Serrano-Alvarado, et E. Desmontils (2018). CaLi : A Lattice-Based Model for License Classifications. In BDA.

Comment interroger les données protégées sous licences ?

Requête SPARQL

- **Trouver les MOOC suivis par les apprenants qui ont suivi le MOOC Web sémantique et que je n'ai pas suivis?**

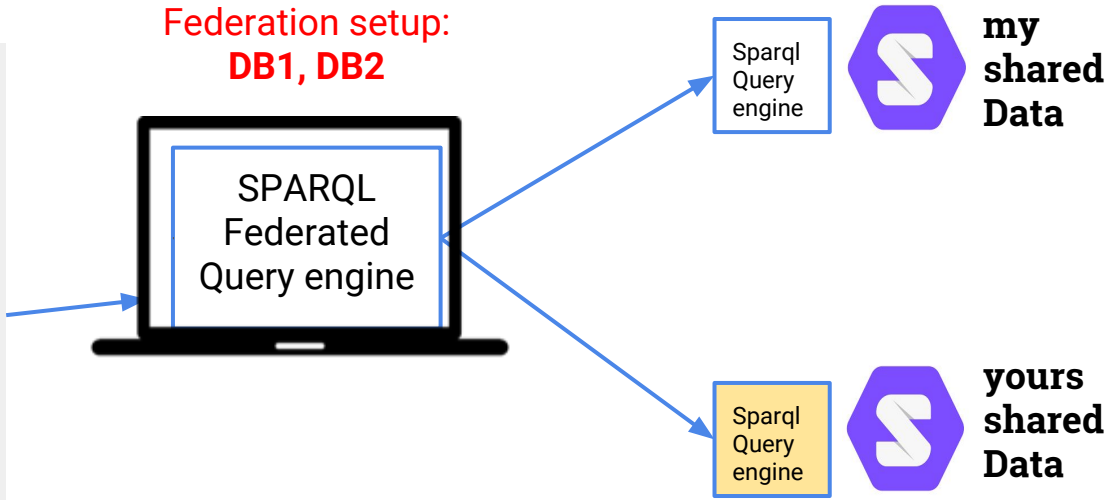
```
select distinct ?c
where {
  :Me edu:connait ?p .
  ?p edu:suit "Web Semantic" .
  ?p edu:suit ?c .
  filter not exists { :ME edu:suit ?c}
  }
```



Requête SPARQL avec licence

- **Trouver les MOOC suivis par les apprenants qui ont suivi le MOOC Web sémantique et que je n'ai pas suivis?**

```
select distinct ?c
FROM FEDERATION <file://myfed.ttl>
where {
GRAPH ?g {
  :Me  foaf:knows  ?p .
  ?p dbprop:follow "Web Semantic" .
  ?p dbprop:follow ?c .
  filter not exists { :ME dbprop:follow ?c}
}
 ?g odrl:hasPolicy ?l .
 ?l  odrs:compatibleWith "query licence"
  }
```



Contrôle d'usage des données personnelles

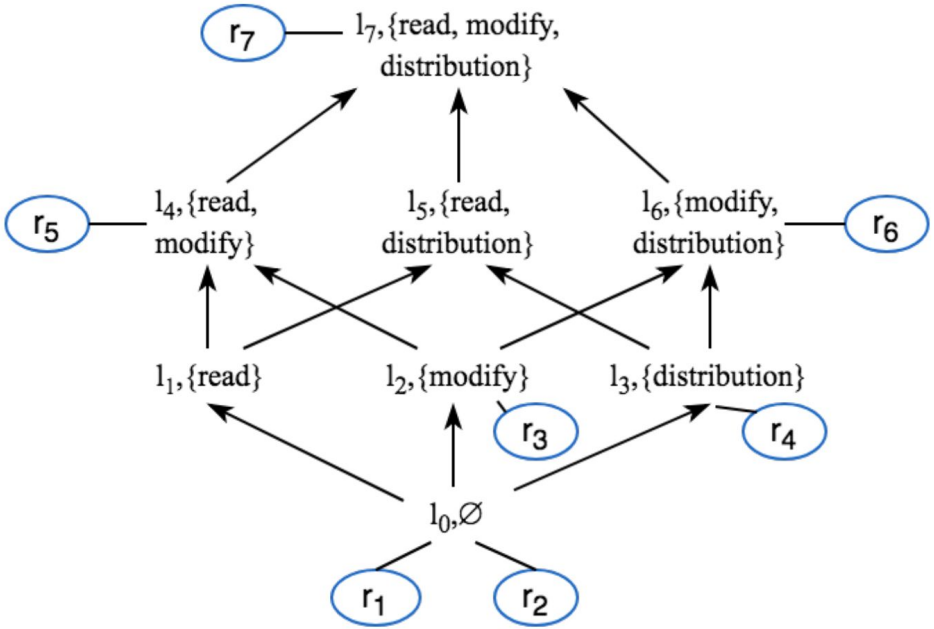


figure 3 : Représentation d'une classification des ressources basée sur les licences [10]

[10] Moreau, B., P. Serrano-Alvarado, et E. Desmontils (2018). CaLi : A Lattice-Based Model for License Classifications. In BDA.

Contrôle d'usage des données personnelles

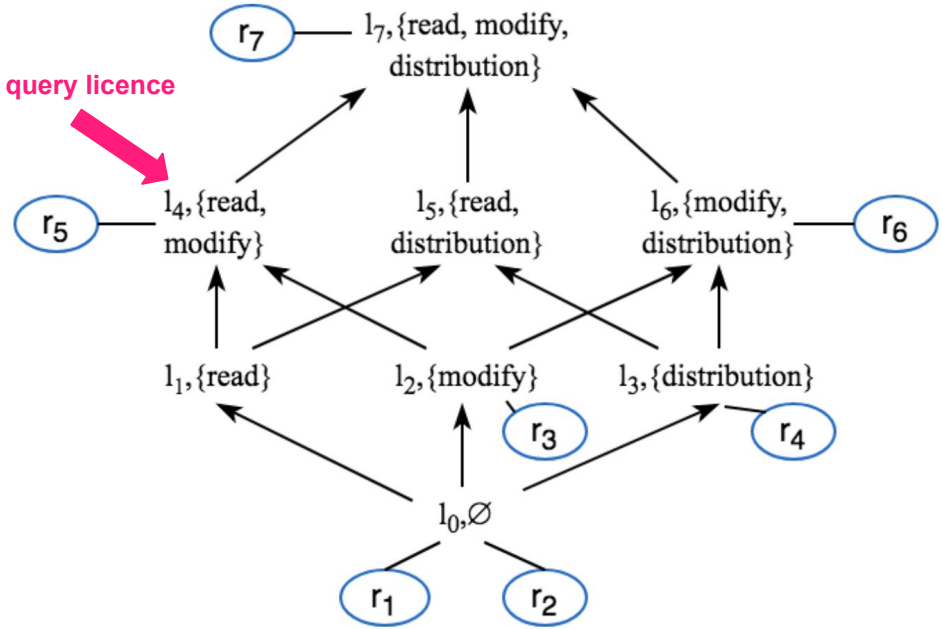


figure 3 : Représentation d'une classification des ressources basée sur les licences [10]

[10] Moreau, B., P. Serrano-Alvarado, et E. Desmontils (2018). CaLi : A Lattice-Based Model for License Classifications. In BDA.

Contrôle d'usage des données personnelles

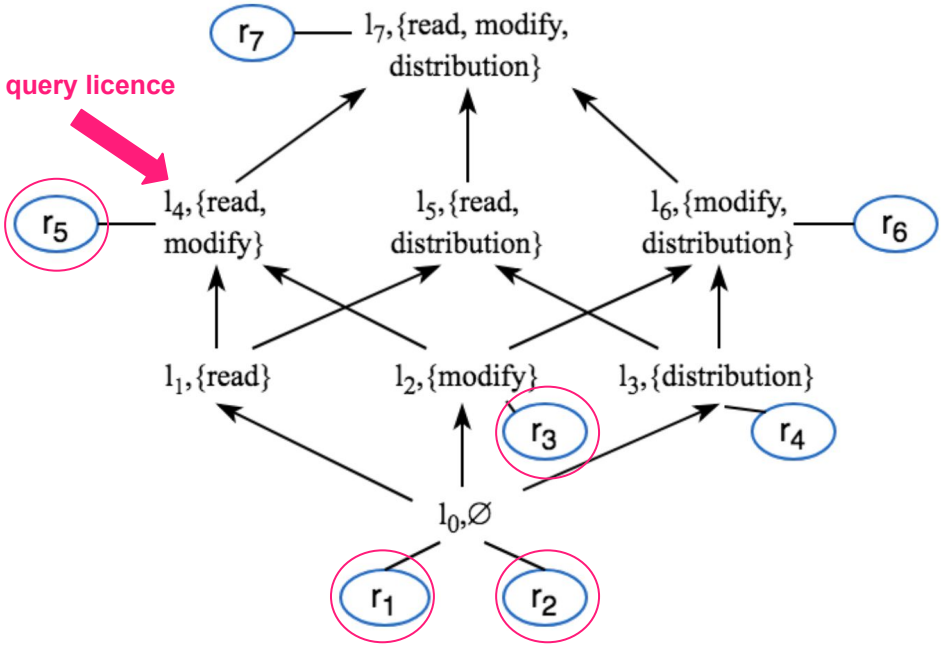


figure 3 : Représentation d'une classification des ressources basée sur les licences [10]

[10] Moreau, B., P. Serrano-Alvarado, et E. Desmontils (2018). CaLi : A Lattice-Based Model for License Classifications. In BDA.

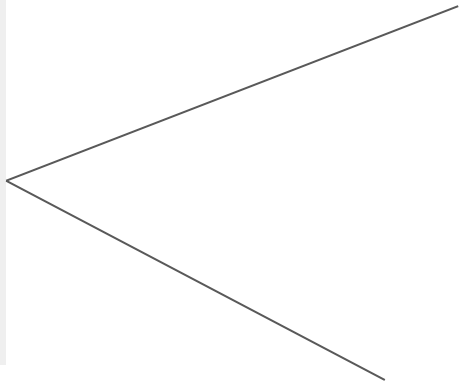
Conclusion

- Infrastructure décentralisation basée sur le web sémantique pour l’apprentissage tout au long de la vie.
  - Deux services pour l’intégration des données sémantiques.
  - Partage de données personnelles de confiance
- Expériences préliminaires avec des étudiants du deuxième cycle en sciences de l’éducation dans le cadre du projet SEDELA.

Merci

semLAV

```
select distinct ?c
where {
  :Me  foaf:knows  ?p .
  ?p dbprop:follow “Web Semantic” .
  ?p dbprop:follow ?c .
  filter not exists { :ME dbprop:follow ?c}
  }
```



-**Compétences**
-Expériences Professionnelles
-**Formations**

-Diplômes
-**Détails cours**

UNIVERSITÉ DE NANTES

Old presentation

Introduction

- L’apprentissage tout au long de la vie commence lors de la formation initiale, de l’**école** jusqu’à l’**université**, et se poursuit pendant toute la carrière, avec les différents **emplois** que nous pouvons occuper.
- Il n’existe actuellement aucune infrastructure fournissant un environnement de confiance pour l’apprentissage à long terme.
- Pourquoi le web sémantique?

Learner ontologies

Infrastructure décentralisée et sémantique

- Intégration des données d'apprentissage personnelles des apprenants dans un espace privé

- Collaboration personnalisée avec les autres apprenants et partage des données personnelles de confiance

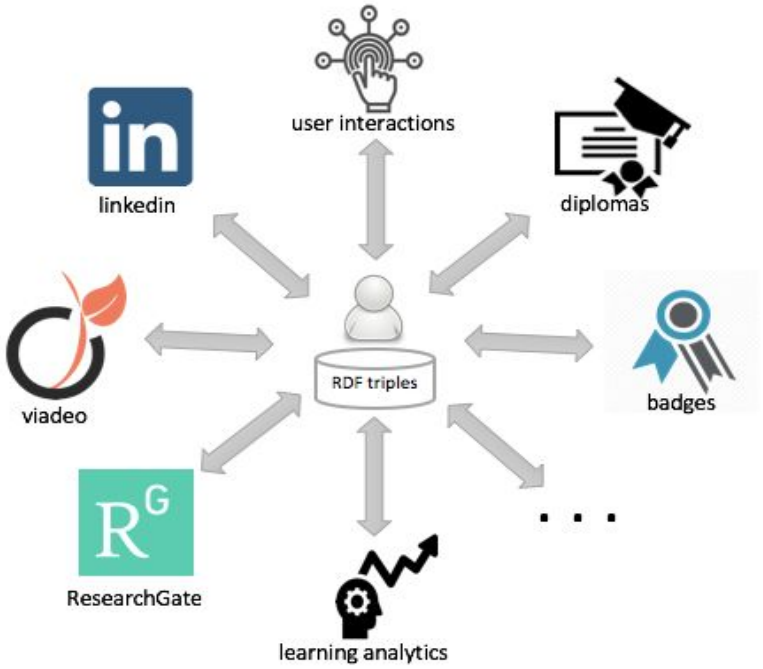


figure 1 : Informations personnelles

Infrastructure décentralisée et sémantique

- Intégration des données d'apprentissage personnelles des apprenants dans un espace privé

- Collaboration personnalisée avec les autres apprenants et partage des données personnelles de confiance

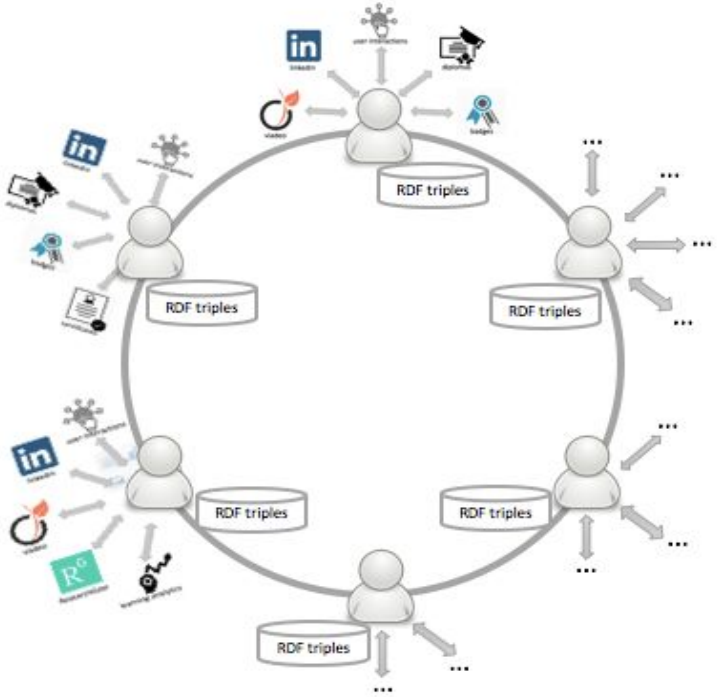
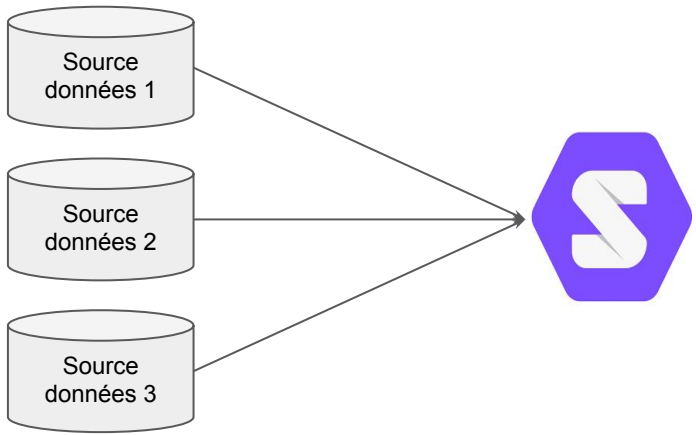


figure 2 : Réseau social d'apprentissage

Comment collecter mes données personnelles d’apprentissage ?

Intégration des données personnelles

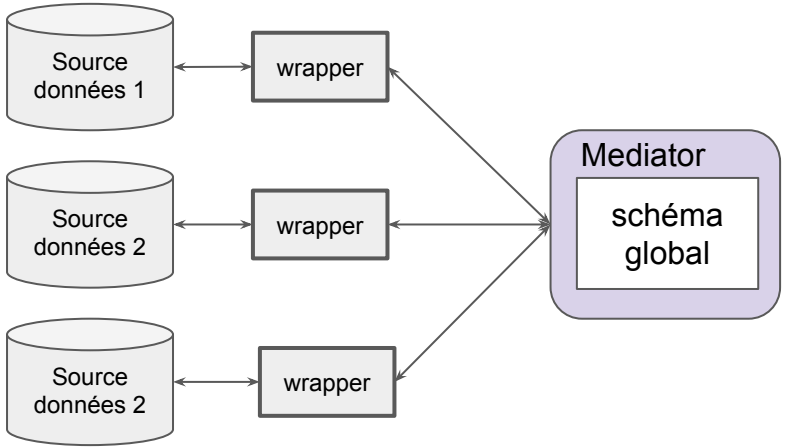
- Data warehousing
  - Transformer toutes les données sous la même forme et les stocker localement (Solid [1]).
  - Solid : plateforme décentralisée pour les applications web social. Les données sont accessibles online et elles sont gérées indépendamment des applications.
  - limite :
    - problème de la fraîcheur des données non-stables
    - gestion des vues matérialisées
    - mémoire de stockage limité



[1] Mansour, E., A. V. Sambra, S. Hawke, M. Zereba, S. Capadisli, A. Ghanem, A. Aboulnaga, et T. Berners-Lee (2016). A Demonstration of the Solid Platform for Social Web Applications. In WWW . Demo paper.

Intégration des données personnelles

- Intégration virtuelle
  - les données restent dans les sources et elles sont accessibles à la demande
  - APIs accessibles pour interroger les données non-RDF [2].
  - Local-As-View (LAV) médiateurs
    - **Requêtes** utilisateur sont posées à une couche intermédiaire (**schéma global**)
    - **Sources de données** locales : des vues sur le schéma global
    - Requête réécrite avec les sources de données



[2] Moreau, B., P. Serrano-Alvarado, E. Desmontils, et D. Thoumas (2017). Queryingnon-RDF Datasets using Triple Patterns. In ISWC. Demo paper

Local-As-View (LAV)

La requête est définie en utilisant le schéma global :

myDiploma(N,L,Y):- name(M,N), hasDiploma(M,D), label(D,L), Year(D,Y)

Sources sont des vues sur le schéma global :

LinkedIn(P,N,L,C) :- name(P,N), hasDiploma(P,R), label(R,L), hasCompetence(P,C)
UniversityNantes(P,W,D,Y) :- name(P,W), hasDiploma(P,D), Year(D,Y)

Requête réécrite en utilisant les mappings de LAV :

myDiploma(N,L,Y) = r1(N,L,Y) Union r2(N,L,Y) where:

- r1(N,L,Y) :- LinkedIn(P,N,L,C), University(P,W,D,Y),
- r2(N,L,Y) :- University(P,N,D,Y), LinkedIn(P,N',L,C)

Approches de réécriture LAV

Très grand nombre de réécritures : $O((M \times V)^N)$

M: taille maximale des vues

V: nombre des vues

N: nombre des prédicats de la requête

Q(P,C,L,W,N):- member(P,C), label(C,L), knows(P,W), name(W, N)

NB of views	NB of rewriting
5	60
10	960
20	15,360
40	245,760
160	6.2914e+07

Peut-on vraiment exécuter ce grand nombre de réécritures?

R. Pottinger and A. Y. Halevy. **MiniCon: A scalable algorithm for answering queries using views**. VLDB J. 2001
Y. Arvelo et al. **Compilation of Query-Rewriting Problems into Tractable Fragments of Propositional Logic.** AAAI. 2006
G. Konstantinidis and J. L. Ambite. **Scalable query rewriting: a graph-based approach**. SIGMOD Conference. 2011

SemLAV [3]

- Médiateur avec une matérialisation intelligente des vues pertinentes
    1. Première étape de l'algorithme Bucket [4]
    2. Les vues dans chaque Bucket sont classées
    3. Les k meilleures vues sont matérialisées
    4. requête initial exécutée avec différentes stratégies [5]

[3] G. Montoya, L. D. Ibánez, H. Skaf-Molli, P. Molli, and M.-E. Vidal. SemLAV: Local-As-View Mediation for SPARQL. Transactions on Large-Scale Data- and Knowledge-Centered Systems, LNCS, Vol. 8420, p 33–58, 2014.
[4] A. Levy et al. Querying Heterogeneous Information Source Using Source Description. VLDB 1996.
[5] P.Folz, G. Montoya, H.Skaf-Molli, P. Molli, and M. Vidal. Parallel data loading during querying deep web and linked open data with SPARQL. ISWC 2015

Comment contrôler mes données personnelles ?
Comment collaborer avec les autres en confiance ?

Partage de données de confiance

- Chaque apprenant associe une politique d'usage différente à ses données

```
<mypod/linkedin/sparql> a odrl:Asset ;
rdfs:comment "mes données collectés de LinkedIn" ;
odrl:hasPolicy <licence1> .
<mypod/badges/sparql> a odrl:Asset ;
rdfs:comment "mes badges numériques" ;
odrl:hasPolicy <licence2>
```

```
licence1 a odrl:Policy ;
rdfs:comment "licence1" ;
odrl:Permission
        [ odrl:action
         cc:read ; ] ;
odrl:prohibition
        [ odrl:action
        cc:archive ] .
```

```
licence2 a odrl:Policy ;
rdfs:comment "licence2" ;
odrl:Permission
        [ odrl:action ,
         cc:read ; ] ;
odrl:prohibition
        [ odrl:action
        cc:distribute ] .
```

Partage de données de confiance

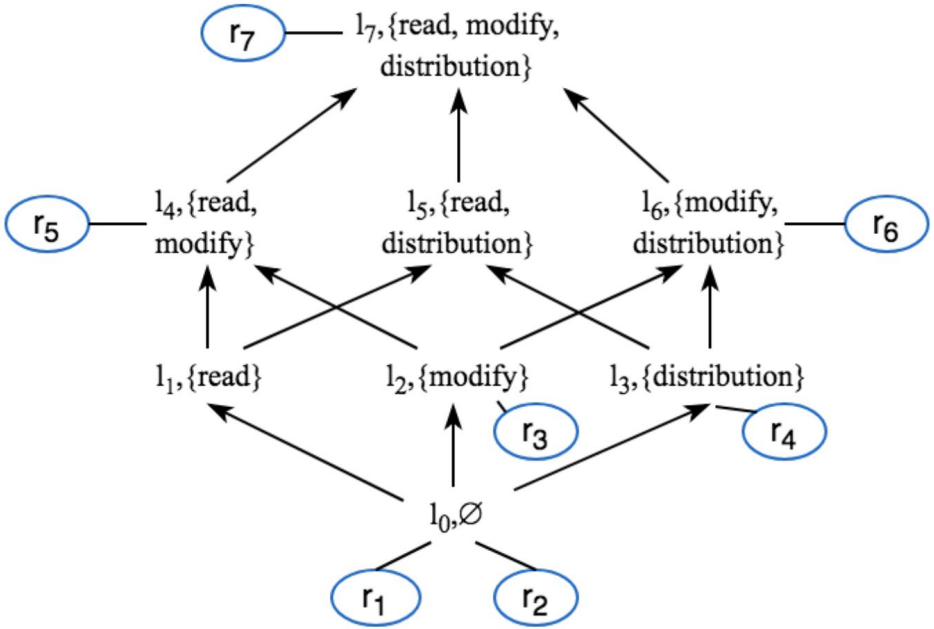


figure 3 : Représentation d'une classification des ressources basée sur les licences [6]

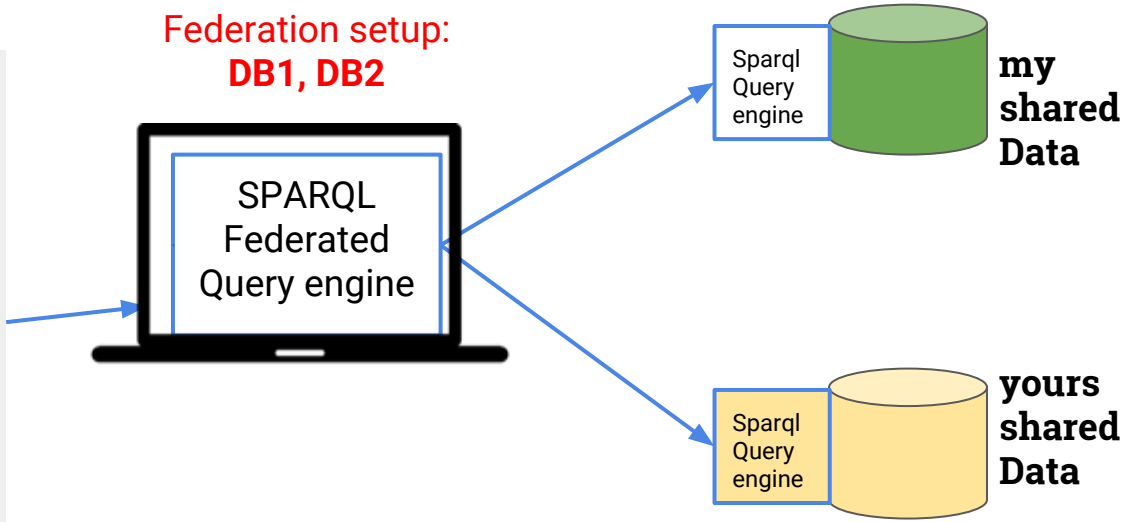
[6] Moreau, B., P. Serrano-Alvarado, et E. Desmontils (2018). CaLi : A Lattice-Based Model for License Classifications. In BDA.

Comment interroger les données protégées sous licences ?

Requête SPARQL avec licence

- *new-cours(C):- knows(Me,P), follow(P, “Web Semantic”), follow(P,C), name(P,N), ,notfollow(Me,C), In(C, G), hasPolicy(G,L), isCompliant(L,”query licence”)*

```
select distinct ?c
FROM FEDERATION <file://myfed.ttl>
where {
GRAPH ?g {
  :Me  foaf:knows  ?p .
  ?p dbprop:follow “Web Semantic” .
  ?p dbprop:follow ?c .
  filter not exists { :ME dbprop:follow ?c}
}
 ?g odrl:hasPolicy ?l .
 ?l  odrs:compatibleWith “query licence”
  }
```



Contrôle d'usage des données personnelles

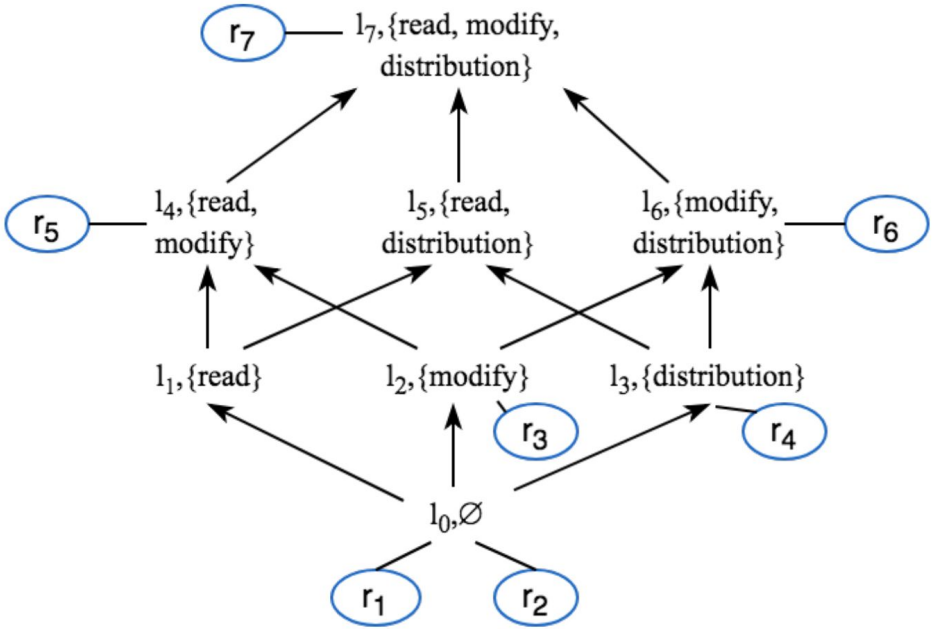


figure 3 : Représentation d'une classification des ressources basée sur les licences [6]

[6] Moreau, B., P. Serrano-Alvarado, et E. Desmontils (2018). CaLi : A Lattice-Based Model for License Classifications. In BDA.

Contrôle d'usage des données personnelles

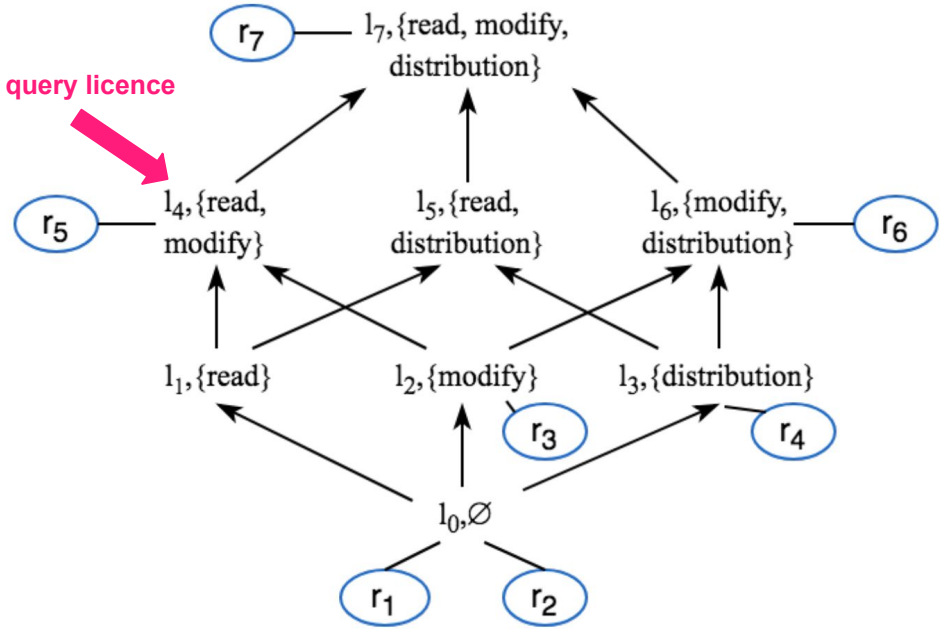


figure 3 : Représentation d'une classification des ressources basée sur les licences [6]

[6] Moreau, B., P. Serrano-Alvarado, et E. Desmontils (2018). CaLi : A Lattice-Based Model for License Classifications. In BDA.

Contrôle d'usage des données personnelles

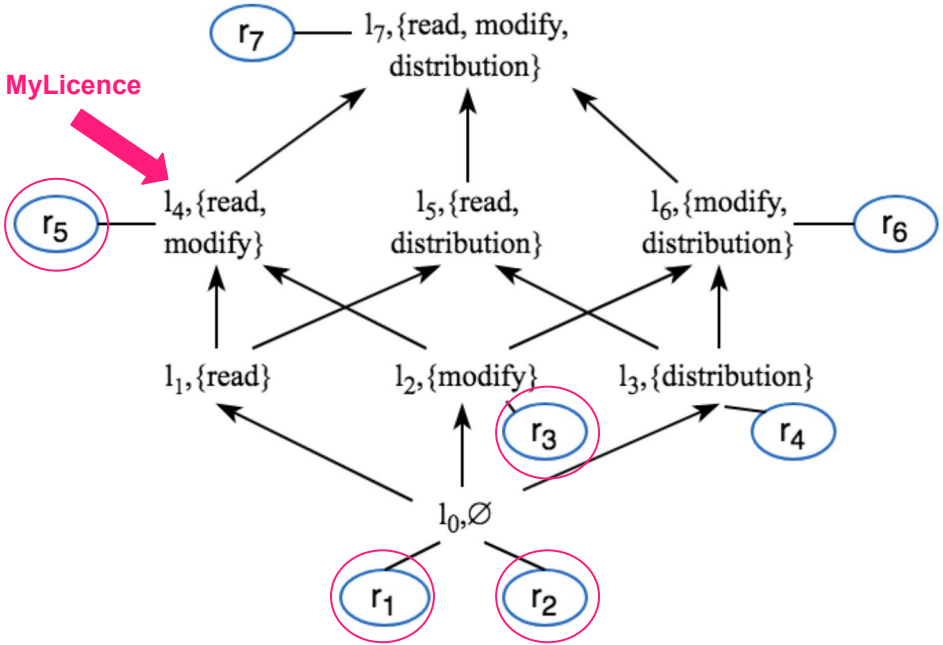


figure 3 : Représentation d'une classification des ressources basée sur les licences [6]

[6] Moreau, B., P. Serrano-Alvarado, et E. Desmontils (2018). CaLi : A Lattice-Based Model for License Classifications. In BDA.

Conclusion

- Les technologies du Web sémantique et la décentralisation constituent le pilier de notre infrastructure d'apprentissage tout au long de la vie.
  - Deux services pour l’intégration des données sémantiques.
  - Partage de données personnelles de confiance
- Expériences préliminaires avec des étudiants du deuxième cycle en sciences de l’éducation dans le cadre du projet SEDELA.

Merci

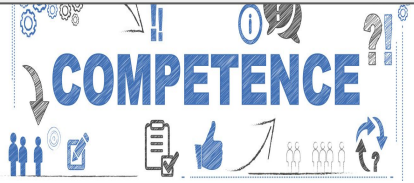
Motivation

UNIVERSITÉ DE NANTES

doyoubuzz

UNIVERSITE MONTPELLIER

Trouver les cinq MOOC les plus suivis par les apprenants qui ont suivi le MOOC Web sémantique ?



UNIVERSITÉ DE RENNES 1

Intégration des données

- Intégration virtuelle
  - données restent dans les sources et elles sont accessibles à la demande
  - APIs accessibles pour interroger les données non-RDF [2].
  - semLAV : Médiateur Local-As-View pour les requêtes SPARQL

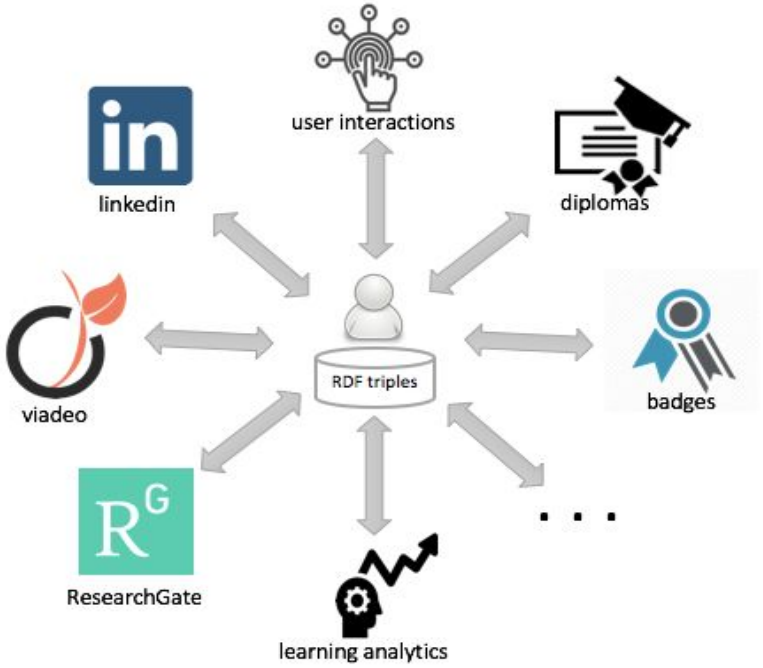
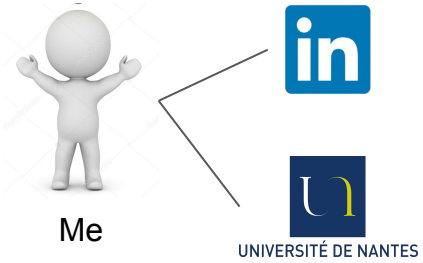


figure 1 : Informations personnelles

Motivation

- Communauté d'apprenants



Trouver les cinq MOOC les plus suivis par les apprenants qui ont suivi le MOOC Web sémantique ?

Motivation

Trouver les cinq MOOC les plus suivis par les apprenants qui ont suivi le MOOC Web sémantique ?

Réseau social d’apprentissage décentralisé

- Infrastructure sémantique et décentralisée
  - Contrôle complet des données personnelles
  - Collaboration personnalisée
  - Partage des données personnelles en confiance

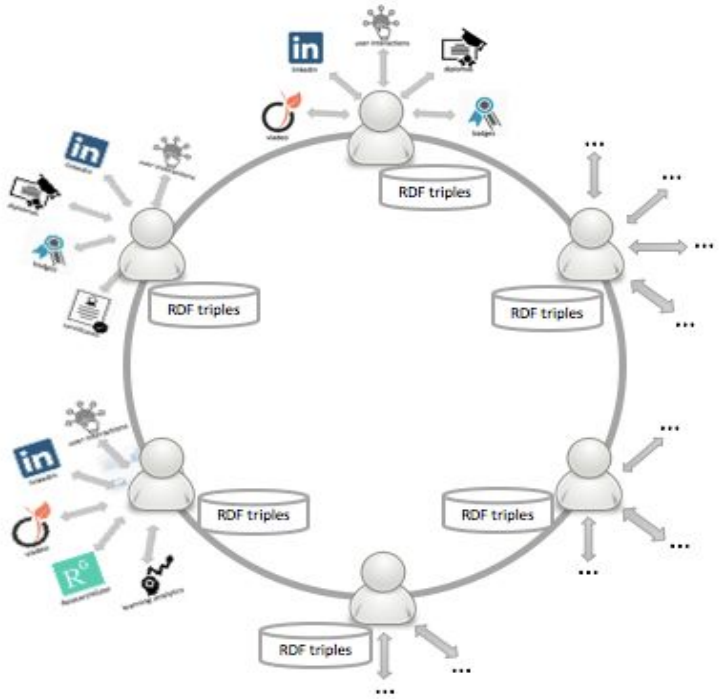
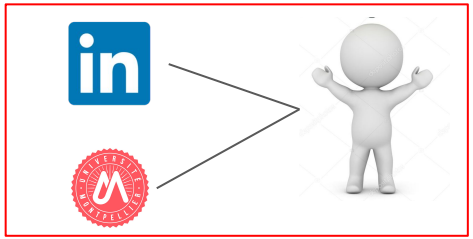


figure 2 : Réseau social d’apprentissage

Objectives



Espace privé qui contient des données personnelles d'apprentissage

Objectives

Trouver les cinq MOOC les plus suivis par les apprenants qui ont suivi le MOOC Web sémantique ?

Collaborer avec les autres en confiance

Pourquoi web sémantique ?

- **Les ontologies** permettent une intégration fluide des données d'apprentissage
  - CCSO[6], ontologie du curriculum et du programme[7]
- **SPARQL** permet d'interroger différentes sources de données

[6] Chung, H. et J. Kim. An ontological approach for semantic modeling of curriculum and syllabus in higher education. International Journal of Information and Education Technology (2016).