

CLARA

Creating and Linking
Licensable Educational Resources



Cominlabs project
2022 - 2024



Patricia Serrano Alvarado
Nantes Université



GTNum #IA_E0 le 26 janvier 2023 à Nantes, France



Le projet CLARA vise à donner aux enseignants les moyens de faciliter la création de ressources éducatives (RE) sous licence basées sur celles qui existent déjà.

Le but est de proposer un ensemble pertinent d'RE cohérent avec une esquisse de cours et des licences compatibles.

Motivation pour la connection de REs

Un moteur de recherche Web renvoie un ensemble de liens
qui mènent à de pages Web.

Pour obtenir des informations pertinentes, vous devez lire
et naviguer parmi ces liens.

Ex. « cours sur les bases de données »

- ❖ Qu'est-ce que lie ces pages Web ?
- ❖ Parlent-elles des mêmes concepts ?
- ❖ A-t-on le droit de réutiliser et de « mixer » ces contenus ?

Google search renseigne la licence uniquement lors de la recherche d'images.



cours sur bases de données



Tous

Vidéos

Actualités

Images

Livres

Plus

Outils

Environ 293 000 000 résultats (0,37 secondes)

<https://www.i3s.unice.fr/~nlt/cours/licence> PDF

Cours SGBD 1 Concepts et langages des Bases de Données ...

IUT de Nice - **Cours SGBD1**. 4. I Notions intuitives. • **Base de données** ensemble structuré de données apparentées qui modélisent un univers réel.

224 pages

<http://sql.bdpedia.fr>

Cours de bases de données - Modèles et langages ...

Cours de bases de données - Modèles et langages • Schéma relationnel (6 points) • SQL (7 points) • Un peu de logique (3 points) • Algèbre (2 points) • Transactions ...

<https://laurent-audibert.developpez.com/Cours-BD/p...>

Bases de Données et langage SQL - Developpez.com

Plus précisément, on appelle **base de données** un ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l' ...

Chapitre 1 - Introduction aux... · 1-2. Système de gestion de...

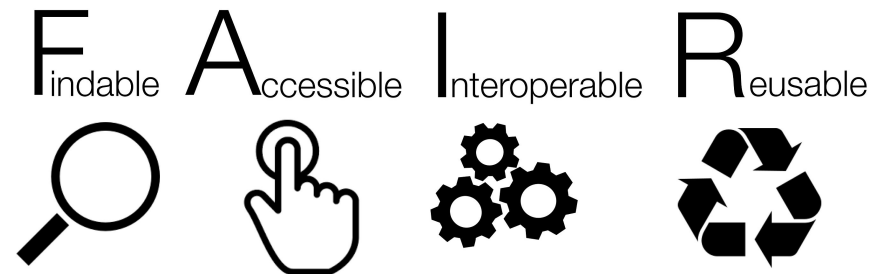
<https://gilles-hunault.leria-info.univ-angers.fr/co...> PDF

Introduction aux bases de données - GILLES HUNAUT (giluno)

Comment connecter les RE

- ❖ **Web Sémantique.** Extension du Web.
 - ❖ Fournit un réseau de ressources qui peuvent être traitées par des machines pour aider leurs utilisateurs à créer de nouvelles connaissances.
 - ❖ Vise à aider l'émergence de nouvelles connaissances en s'appuyant sur le web existant.
- ❖ **Web des données** (Linked Data, LD). Méthode de publication de données structurées sur le Web
 - ❖ Permet de lier et structurer l'information sur internet pour accéder simplement à la connaissance déjà existante.
 - ❖ Les données sont décrites en RDF à l'aide d'ontologies.

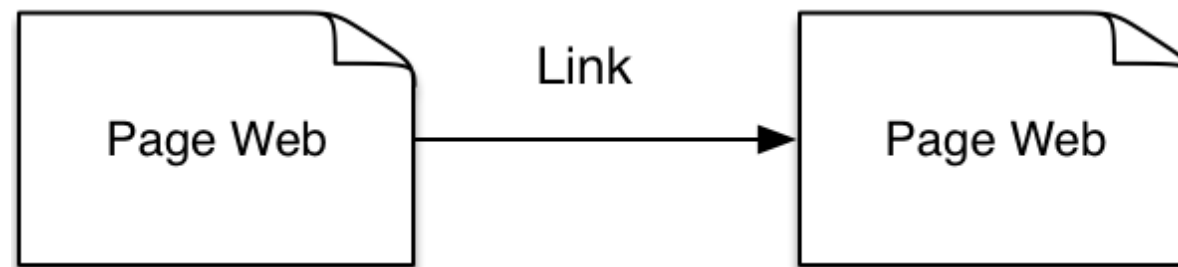
Les principes FAIR



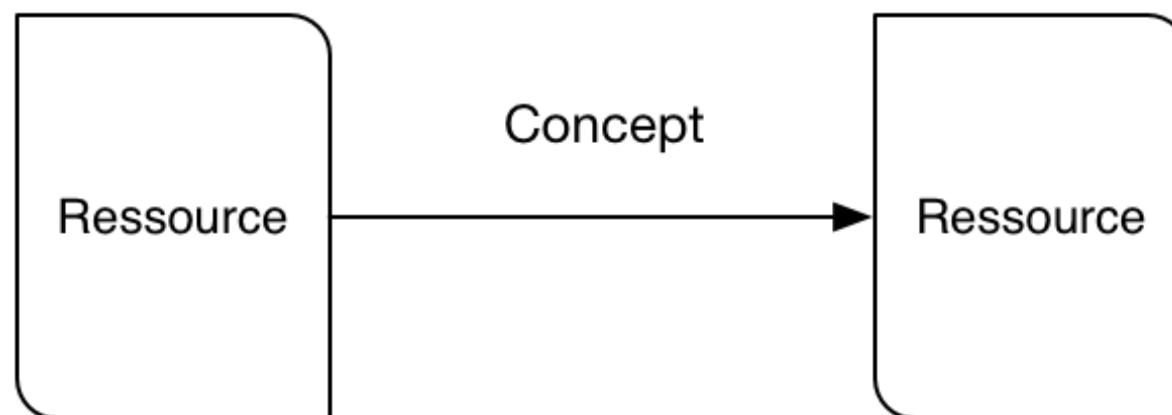
- ❖ **Faciles à (re)trouver.** Les données doivent être faciles à trouver par tous, humains et machines
- ❖ **Accessibles.** Les données FAIR ne sont pas obligatoirement des données ouvertes. Elles doivent être récupérables par leur identifiant en utilisant un protocole standard de communication et les (méta)données sont disponibles à des conditions connues, grâce à des licences claires.
- ❖ **Interopérables.** L'interopérabilité implique l'utilisation des métadonnées contextuelles précises, et de contenu et de formats respectant les grands standards internationaux et utilisant un vocabulaire respectant les principes FAIR.
- ❖ **Réutilisables.** La réutilisation (libre, conditionnelle ou payante) doit être facilité par l'utilisation de standards communs. Les (méta)données doivent être richement décrites, par une pluralité d'attributs précis et pertinents, incluant des détails sur leur provenance. Elles sont libérées avec une licence d'utilisation claire et accessible.

Web vs web des données

Web traditionnel



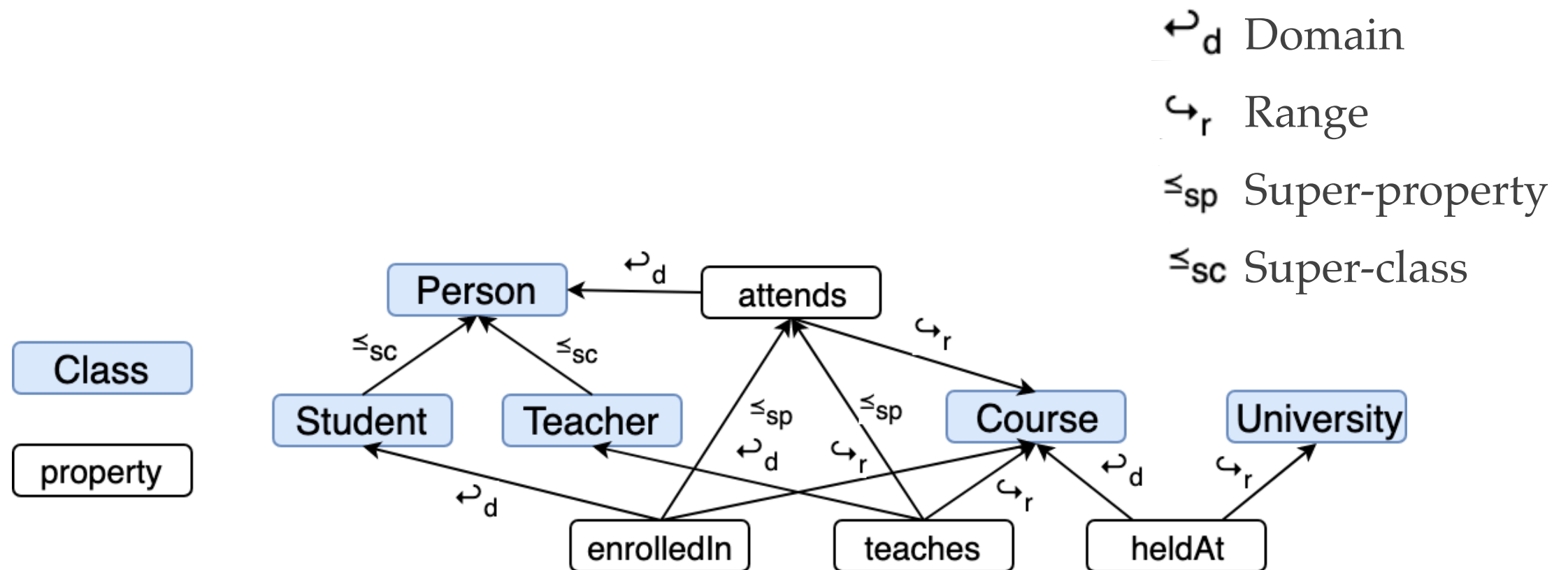
Web des données



Vocabulaire et ontologie*

- ❖ Il n'y a pas de séparation claire entre ce qu'on appelle les «vocabulaires» et les «ontologies».
- ❖ La tendance est d'utiliser le mot « ontologie » pour des ensembles de termes plus complexes et peut-être assez formels
- ❖ «Vocabulaire» est utilisé lorsqu'un formalisme aussi strict n'est pas nécessairement utilisé ou seulement dans un sens très vague.

Ex. d'ontologie

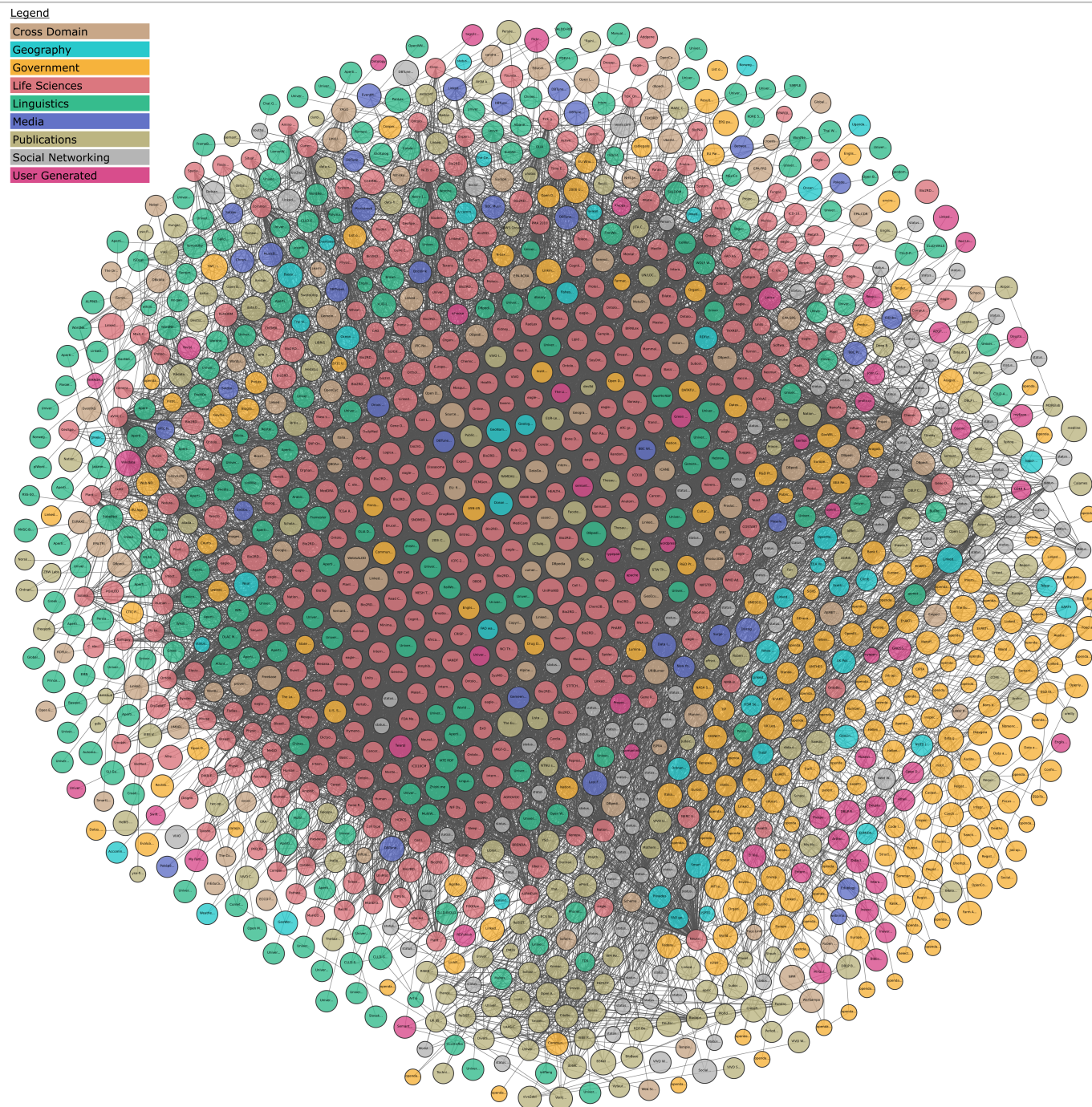


Graphe de connaissances

Knowledge graph (KG)

- ❖ Collection de descriptions interconnectées de concepts, d'entités, de relations et d'événements où :
 - ❖ Les descriptions ont une sémantique formelle compréhensible par les personnes et les machines ;
 - ❖ Les descriptions contribuent les unes aux autres, formant un réseau
 - ❖ Diverses données sont connectées et décrites par des métadonnées sémantiques selon le modèle de connaissances.
- ❖ KG ouverts : DBpedia, Wikidata, Freebase, YAGO, etc.
- ❖ Recherche de KG : LODAtlas*

Le web de données ouvert (LOD*)



1,255 datasets

Graphe du 2022-11-03

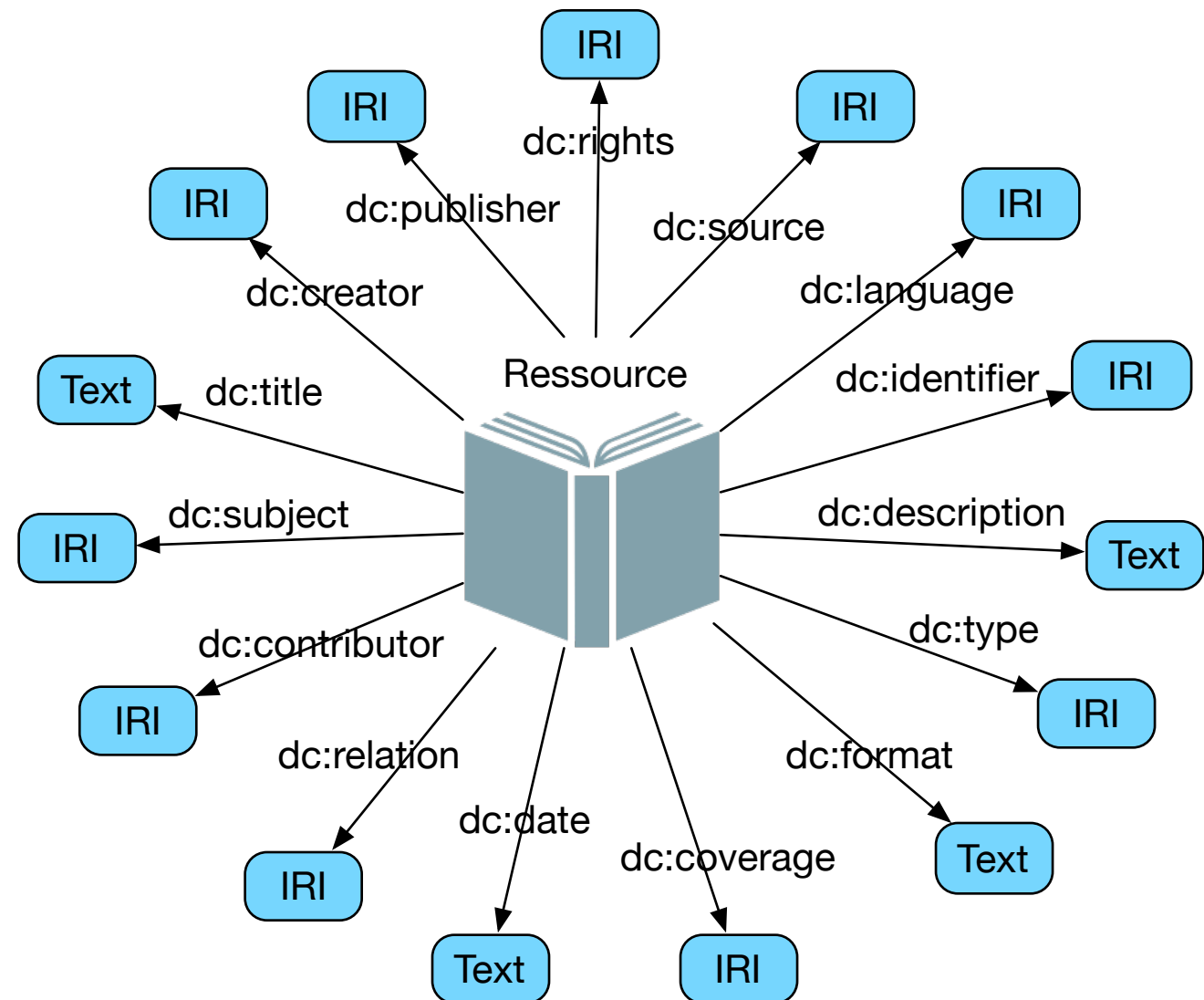
The Linked Open Data Cloud from lod-cloud.net



Vocabulaires pour décrire des ressources éducatives

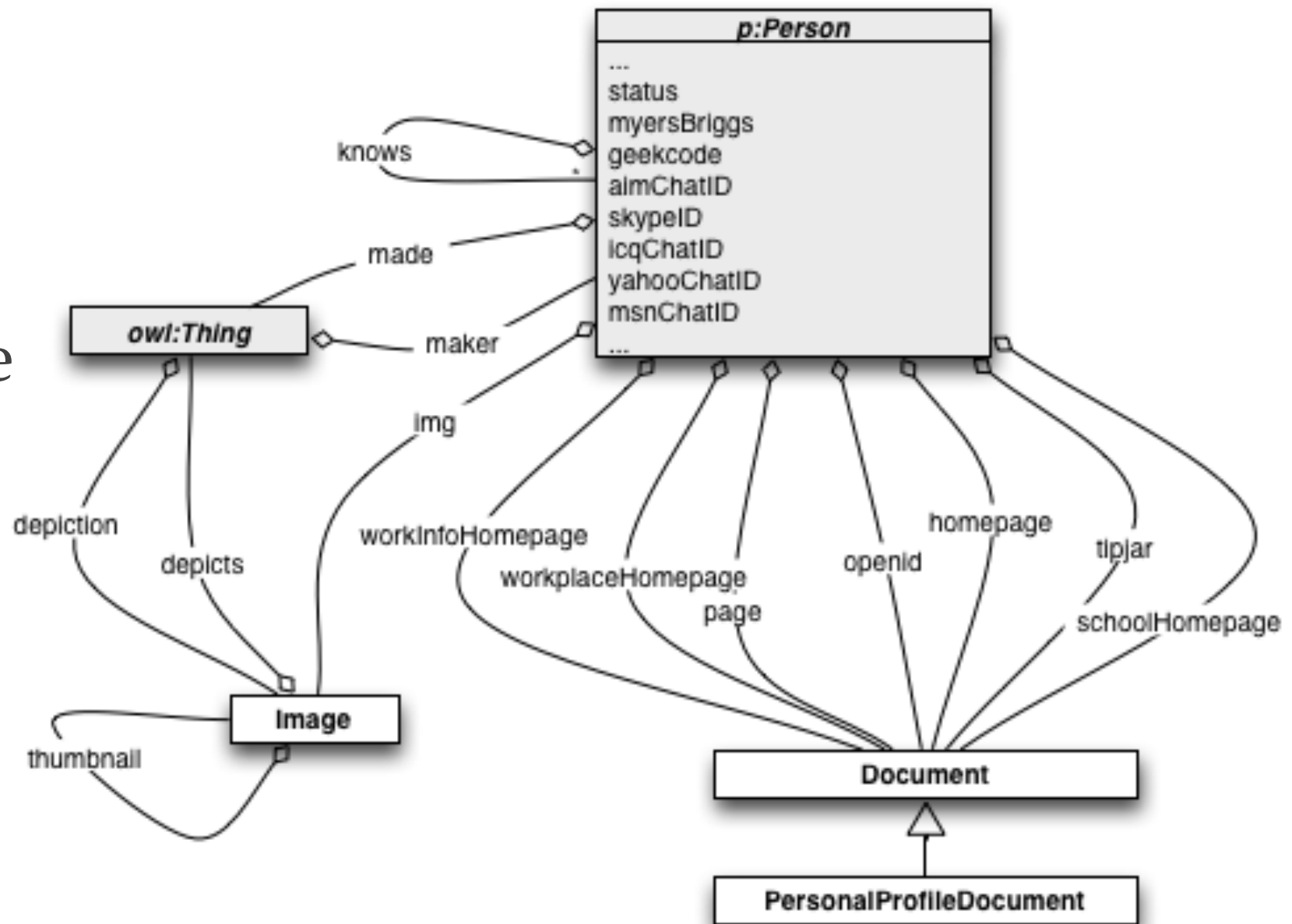
Dublin Core*

- ❖ Vocabulaire issu d'un consensus international et multidisciplinaire pour **décrire des documents** de manière simple et standardisée
- ❖ Socle commun d'éléments descriptifs suffisamment structuré pour permettre une **interopérabilité minimale** entre des systèmes conçus indépendamment les uns des autres.



Friend of a friend (FOAF)*

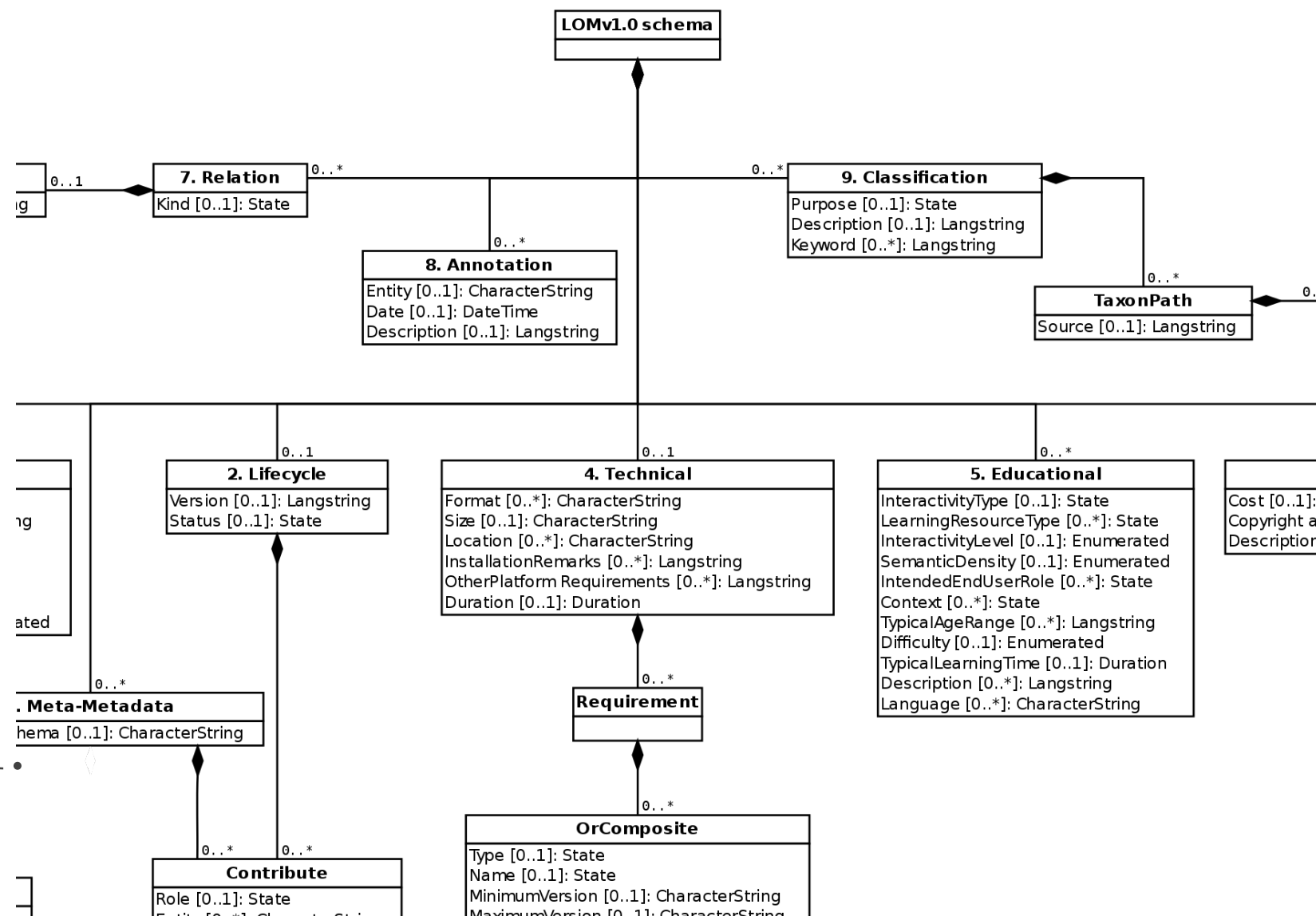
- ❖ Vocabulaire permettant de décrire des personnes et les relations qu'elles entretiennent entre elles.



Learning Object Metadata (LOM)*

- ❖ Permet de décrire de ressources d'enseignement et d'apprentissage.

- ❖ Standard IEEE 1484.12.1-2002 (LOM)1.

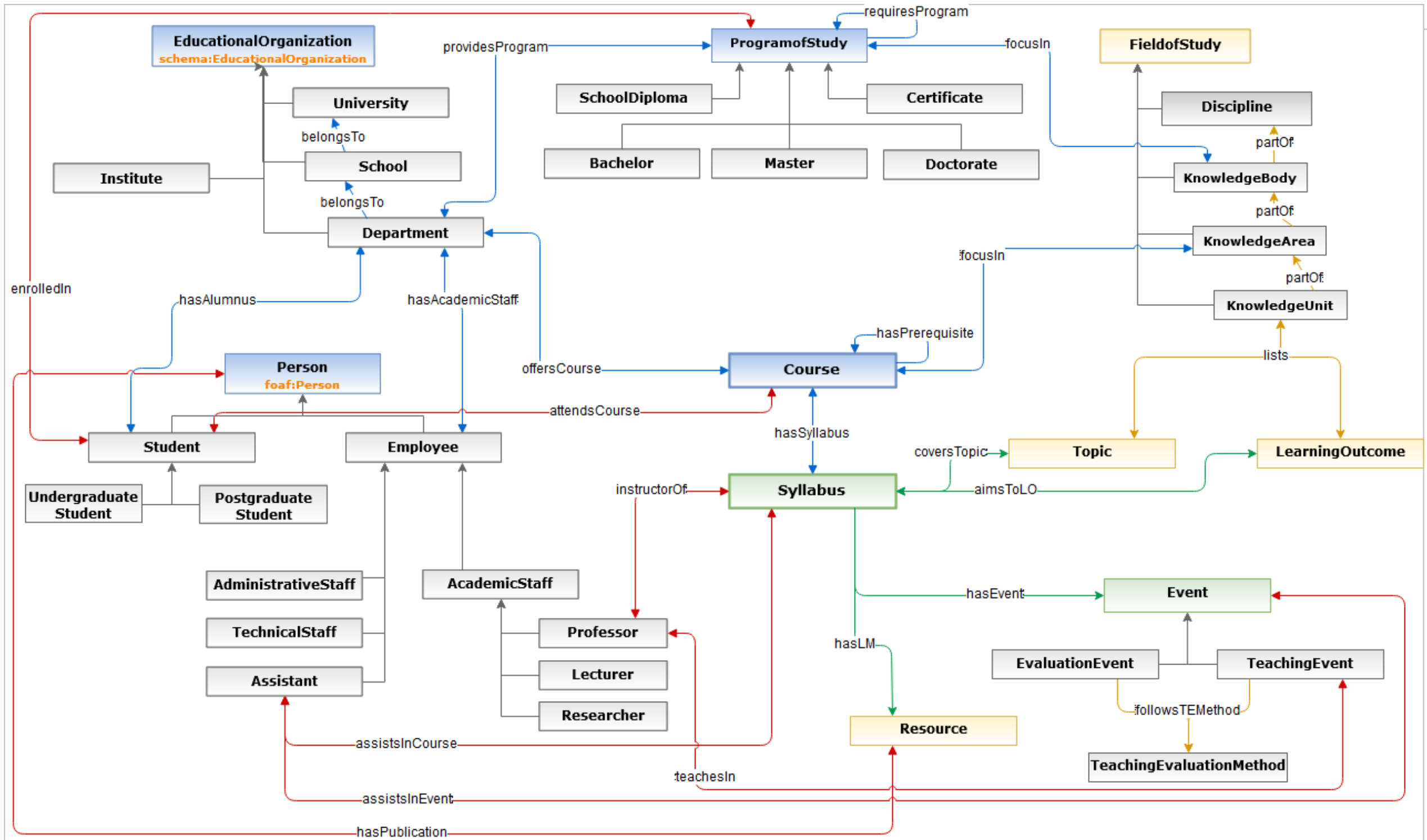


Teaching Core Vocabulary (TEACH)*

Classes: | [Course](#) | [Assignment](#) | [Material](#) | [Teacher](#) | [StudentGroup](#) | [Student](#) | [Module](#) | [StudyProgram](#) | [Lecture](#) | [Building](#) | [Room](#) |

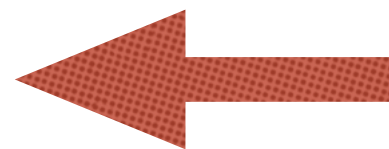
Properties: | [courseDescription](#) | [courseTitle](#) | [hasDescription](#) | [hasTitle](#) | [bookingNumber](#) | [teacher](#) | [teacherOf](#) | [studentGroup](#) | [weeklyHours](#) | [grading](#) | [academicTerm](#) | [room](#) | [building](#) | [reportTemplate](#) | [reviewTemplate](#) | [opponentOf](#) | [hasAssignment](#) | [hasAssignmentMaterial](#) | [hasCourseMaterial](#) | [reading](#) | [nextReading](#) | [deadline](#) | [deadlineDraftReport](#) | [deadlineFinalReport](#) | [deadlineReviewReport](#) | [arrangedAt](#) | [notArrangedAt](#) | [module](#) | [studyProgram](#) | [hasTeachingMethod](#)

Curriculum Course Syllabus Ontology (CCSO)*



Les vocabulaires peuvent être utilisés ensemble

- ❖ Autres vocabularies
 - ❖ AIISO - Academic Institution Internal Structure Ontology,
 - ❖ MLO - Metadata for Learning Opportunities,
 - ❖ XCRI-CAP - XCRI Course Advertising Profile,
 - ❖ BIBO - The Bibliographic Ontology,
 - ❖ W3C Ontology for Media Resources,
 - ❖ Open Provenance Model Vocabulary,
 - ❖ etc.
- ❖ Moteur de recherche de vocabulaires: LOV*



Licences lisibles par machine dans le Web de données

Ex. CC-BY NC

Licence publique Creative Commons Attribution - Utilisation non commerciale 4.0 International

Lorsque Vous exercez les Droits accordés par la licence (définis ci-dessous), Vous acceptez d'être lié par les termes et conditions de la présente Licence publique Creative Commons Attribution - Utilisation non commerciale 4.0 International (la « Licence publique »). Dans la mesure où la présente Licence publique peut être interprétée comme un contrat, Vous bénéficiez des Droits accordés par la licence en contrepartie de Votre acceptation des présents termes et conditions, et le Donneur de licence Vous accorde ces droits en contrepartie des avantages que lui procure le fait de mettre à disposition l'Œuvre sous licence en vertu des présents termes et conditions.

Article 1 - Définitions.

- a. **Œuvre dérivée** signifie œuvre protégée par les Droit d'auteur et droits connexes, dérivée ou adaptée de l'Œuvre sous licence et dans laquelle l'Œuvre sous licence est traduite, retouchée, arrangée, transformée, ou modifiée de telle façon que l'autorisation du Donneur de licence est nécessaire, conformément aux dispositions des Droit d'auteur et droits connexes. Dans le cas de la présente Licence publique, lorsque l'Œuvre sous licence est une œuvre musicale, une représentation publique ou un enregistrement sonore, la synchronisation de l'Œuvre sous licence avec une image animée sera considérée comme une Œuvre dérivée aux fins de la présente Licence publique.
- b. **Licence d'Œuvre dérivée** signifie licence par laquelle Vous accordez Vos Droit d'auteur et droits connexes portant sur Vos contributions à l'Œuvre dérivée, selon les termes et conditions de la présente Licence publique.
- c. **Droit d'auteur et droits connexes** signifie droit d'auteur et/ou droits connexes incluant, notamment, la représentation, la radio et télédiffusion, l'enregistrement sonore et le Droit sui generis des producteurs de bases de données, quelle que soit la classification ou qualification juridique de ces droits. Dans le cadre de la présente

Ex. CC-BY NC lisible par machine

- ❖ Représentation de la licence CC BY-NC avec les technologies du Web sémantique
 - ❖ en RDF
 - ❖ avec usage d'ontologies bien connues (CC, ODRL)

```
:cc-by-nc a odrl:Policy ;  
rdfs:label "CC BY-NC";  
  
...  
odrl:permission  
  [odrl:action  
    cc:Distribution,  
    cc:Reproduction,  
    cc:DerivativeWorks];  
  
odrl:duty  
  [odrl:action  
    cc:Notice,  
    cc:Attribution];  
  
odrl:prohibition  
  [odrl:action  
    cc:CommercialUse].
```

Ex. CC-BY NC lisible par machine

Une licence est un ensemble
d'*actions* réparties sur des *status*.



: 12 actions

<https://creativecommons.org/ns>

ODRL Information Model 2.2 : 72 actions

W3C Recommendation 15 February 2018

<https://www.w3.org/TR/odrl-vocab/#actionConcepts>

```
:cc-by-nc a odrl:Policy ;  
rdfs:label "CC BY-NC";  
...  
odrl:permission  
  [odrl:action  
    cc:Distribution,  
    cc:Reproduction,  
    cc:DerivativeWorks];  
odrl:duty  
  [odrl:action  
    cc:Notice,  
    cc:Attribution];  
odrl:prohibition  
  [odrl:action  
    cc:CommercialUse].
```


Compatibilité de quelques licences populaires

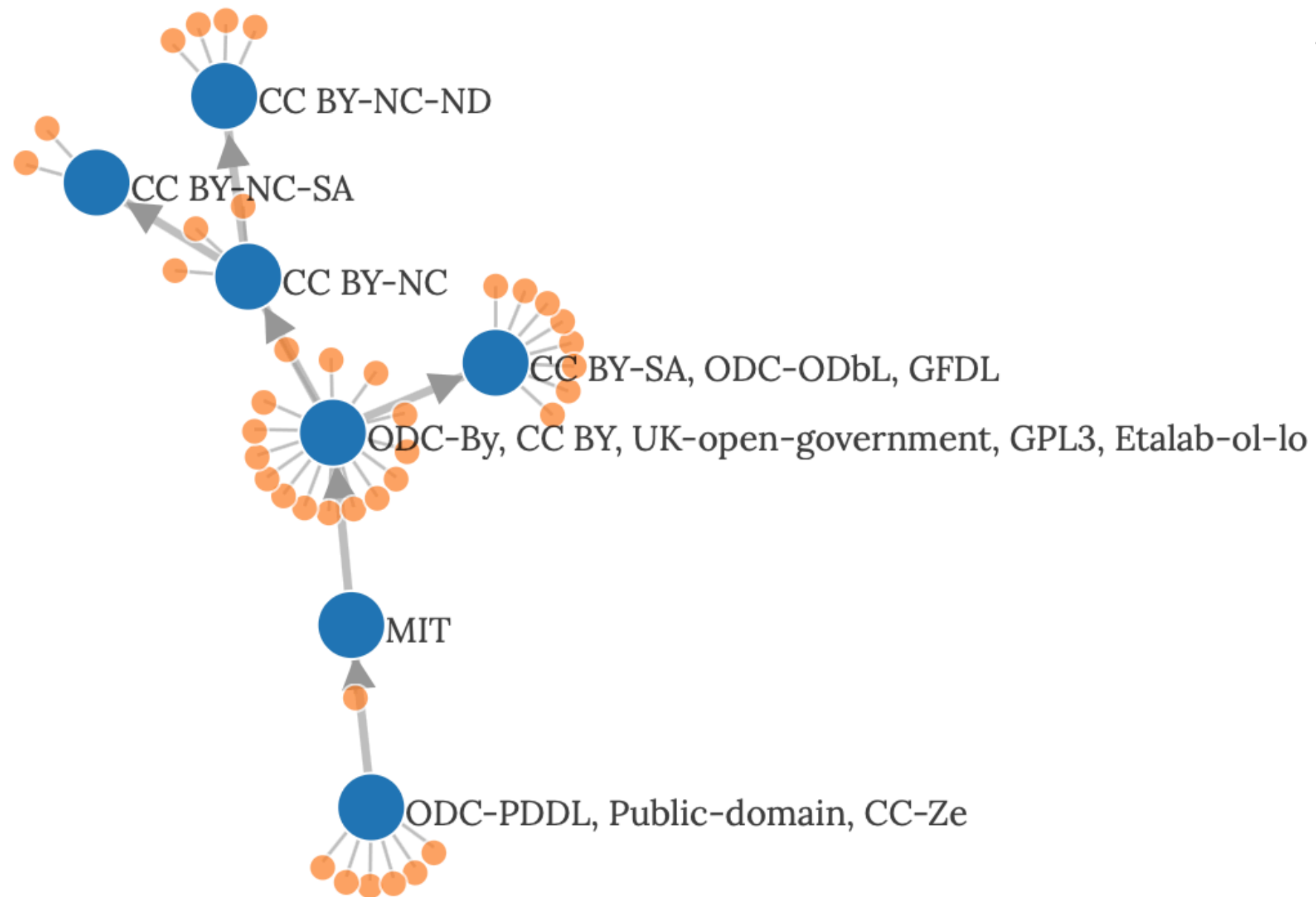


Image prise de : <http://cali.priloo.univ-nantes.fr/rep/graph>

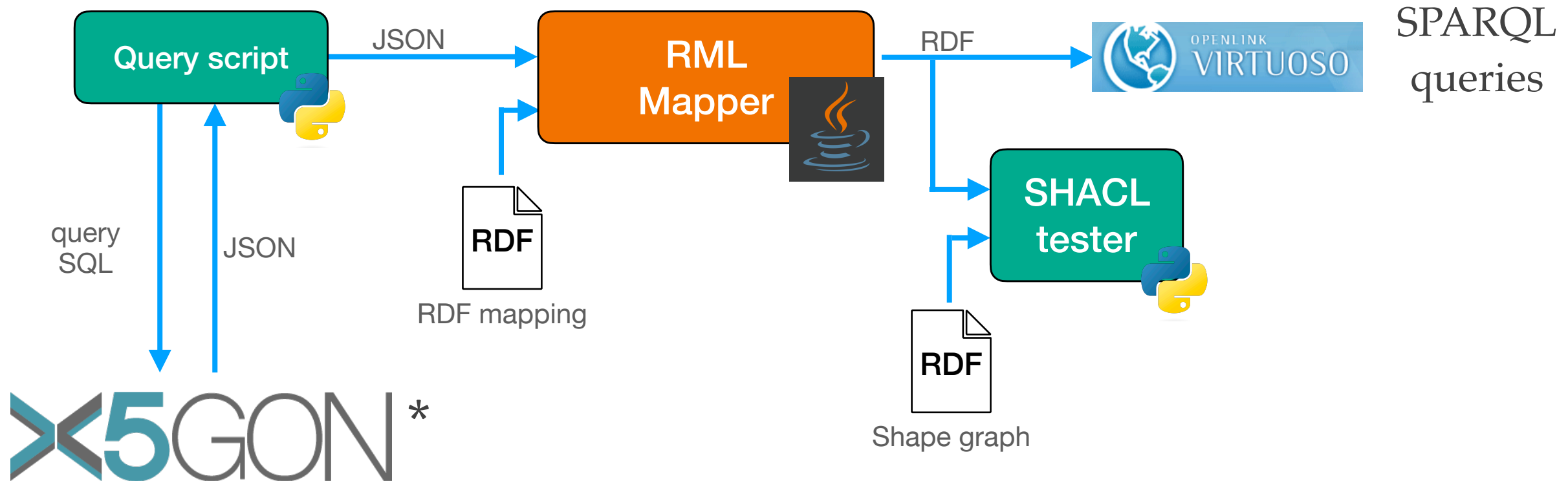
B. Moreau, P. Serrano-Alvarado, M. Perrin, E. Desmontils. **Modelling the Compatibility of Licenses.**
In *ESWC2019*, <https://hal.archives-ouvertes.fr/hal-02069076>

Construction et interrogation d'un graphe de connaissances de ressources éducatives sous licence

Travaux en cours

- ❖ Construction d'un corpus d'RE liées avec une représentation sémantique riche (métadonnées disponibles, licences, similitudes, dépendances, etc.).
- ❖ Conception un moteur de recherche et de navigation qui facilite la recherche de ressources pédagogiques pertinentes compatibles avec les licences.

CLARA pipeline



*X5GON European project <https://wp3.x5gon.org/>

Recommandations du W3C pour le web sémantique

- ❖ SHACL <https://www.w3.org/TR/2017/REC-shacl-20170720/>
- ❖ SPARQL <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>
- ❖ RDF <https://www.w3.org/TR/REC-rdf-syntax/>
- ❖ RML qui se base sur R2RML <https://www.w3.org/TR/r2rml/>
- ❖ RDF-star (pas encore pret) https://w3c.github.io/rdf-star/cg-spec/editors_draft.html

Collecte actuelle de RE

61 570 854

triplets RDF

11 947

auteurs

45 524

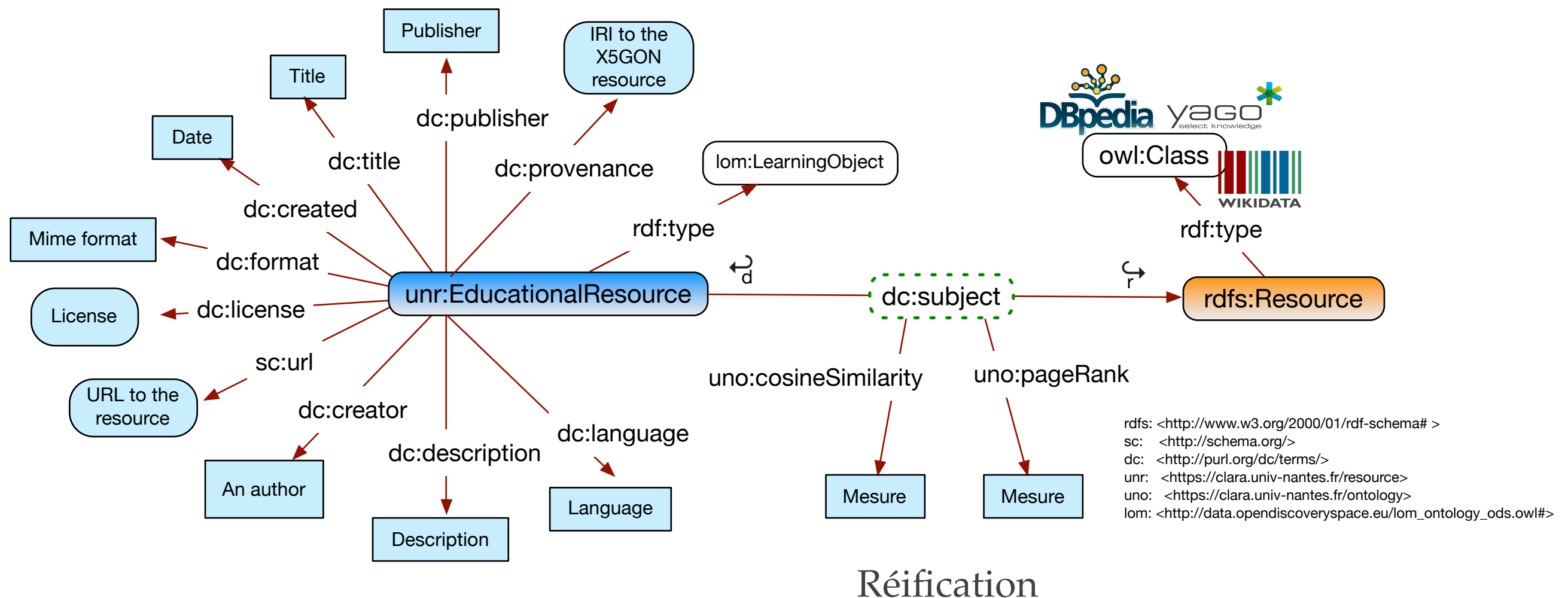
RE issues du projet X5GON

135 069

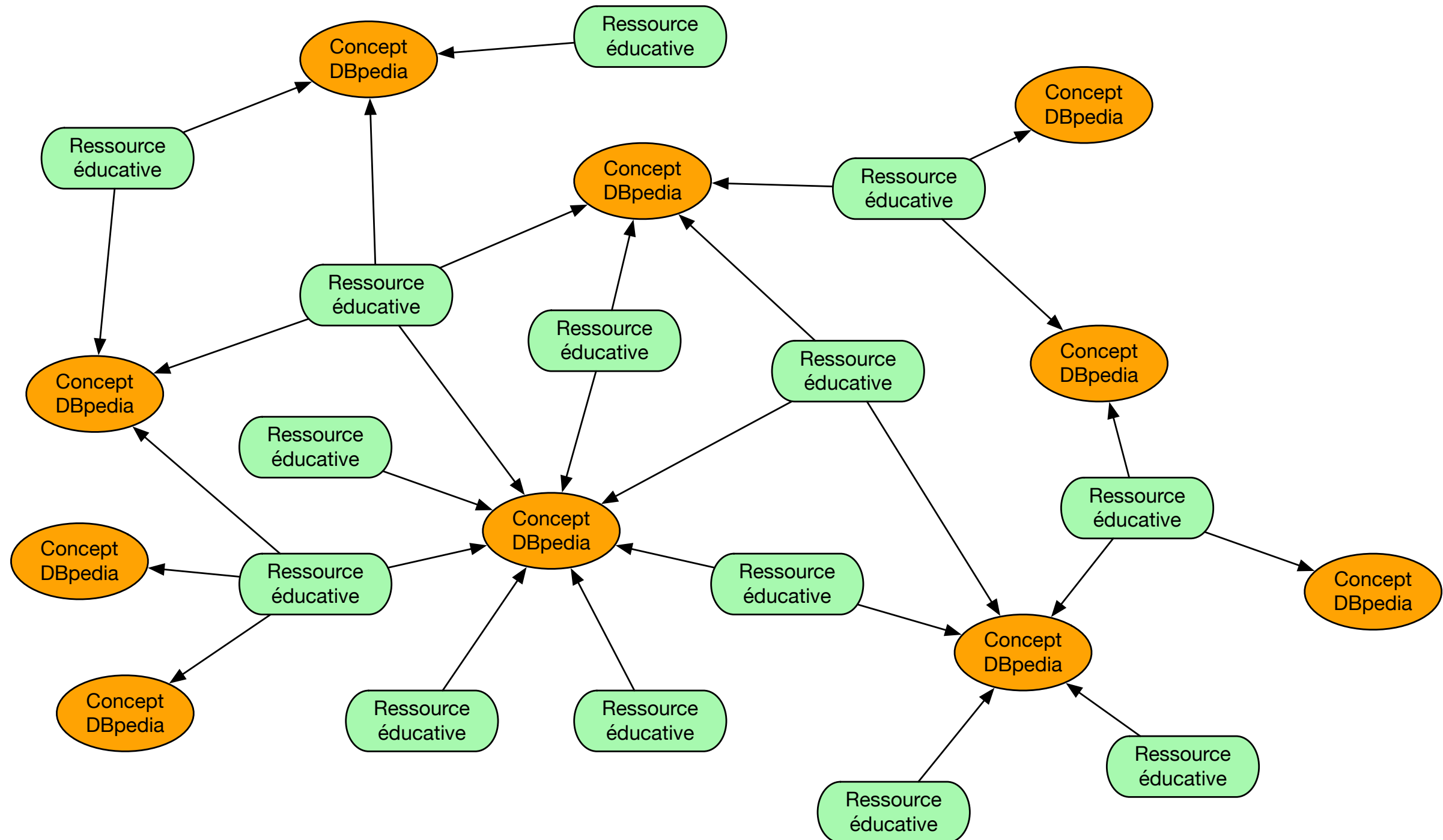
concepts
(sujets identifiés
dans les RE)*

*<https://wikifier.org>

CLARA ontology

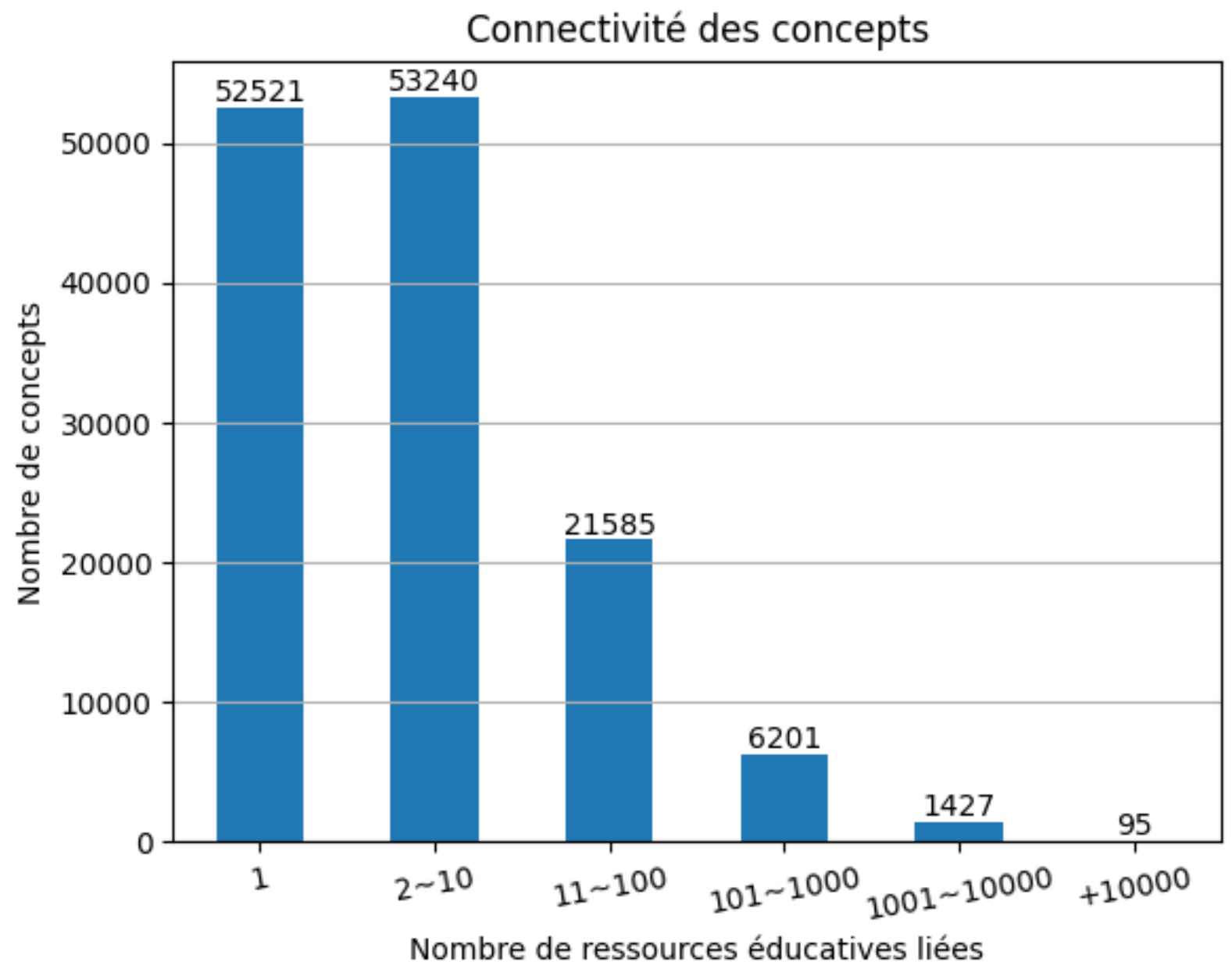


Les RE et leurs concepts



Connectivité des RE

- 52 521 concepts n'ont qu'une RE liée : connectivité nulle.
- 53 240 concepts ont entre 2 et 10 RE liées : connectivité pauvre ou bonne.
- 95 concepts sont hyper-représentés dans les RE : forte connectivité.



CLARA search engine (v0)

Reset filters and topics

Path graph



Neural network



Add a new topic

FILTERS

language

format

☒ application/pdf

☒ audio/mpeg

☒ video/mp4

license

☐ by-nc-nd/2.5/ch

☐ by-nc-nd/3.0

☐ by-nc-nd/3.0/de

☐ by-nc-nd/3.0/us

☐ by-nc-sa/2.5/si

☐ by-nc-sa/3.0/de

☐ by-nc-sa/3.0/us

☐ by-nc-sa/4.0

Online Prediction on Large Diameter Graphs

[Check it!](#)

Guy Lever
English
video/mp4

We continue our study of online prediction of the labelling of a graph. We show a fundamental limitation of Laplacian-based algorithms: if the graph has a large diameter then the number of mistakes made by such algorithms may be proportional to the square root of the number of vertices, even when tackling simple problems. We overcome this drawback by means of an efficient algorithm which achieves a logarithmic mistake bound. It is based on the notion of a spine, a path graph which provides a linear embedding of the original graph. In practice, graphs may exhibit cluster structure; thus in the last part, we present a modified algorithm which achieves the "best of both worlds": it performs well locally in the presence of cluster structure, and globally on large diameter graphs.

Path graph

[by-nc-nd/3.0/](#)

Sequence learning with hidden units in spiking neural networks

[Check it!](#)

Johanni Brea
English
application/pdf

We consider a statistical framework in which recurrent networks of spiking neurons learn to generate spatio-temporal spike patterns. Given biologically realistic stochastic neuronal dynamics we derive a tractable learning rule for the synaptic weights towards hidden and visible neurons that leads to optimal recall of the training sequences. We show that learning synaptic weights towards hidden neurons significantly improves the storing capacity of the network. Furthermore, we derive an approximate online learning rule and show that our learning rule is consistent with Spike-Timing Dependent Plasticity in that if a presynaptic spike shortly precedes a postsynaptic spike, potentiation is induced and otherwise depression is elicited.

Neural network

[by-nc-nd/3.0/](#)

Opportunités en recherche 1/2

- ❖ Si la requête utilisateur Q ne donne pas assez de RE intéressantes, comment étendre la portée de Q pour garantir un nombre k de RE ?
 - ❖ Q peut être étendu à un grand nombre de requêtes
 - ❖ laquelle pourrait être la plus intéressante ou la plus proche de la requête de l'utilisateur ?
 - ❖ comment prendre en compte la réification ?
- ❖ Une fois qu'un nombre k de RE est trouvé, comment les classer du plus au moins intéressant vis-à-vis de Q ?

Relâchement des contraintes d'une requête

Quelques règles de relâchement

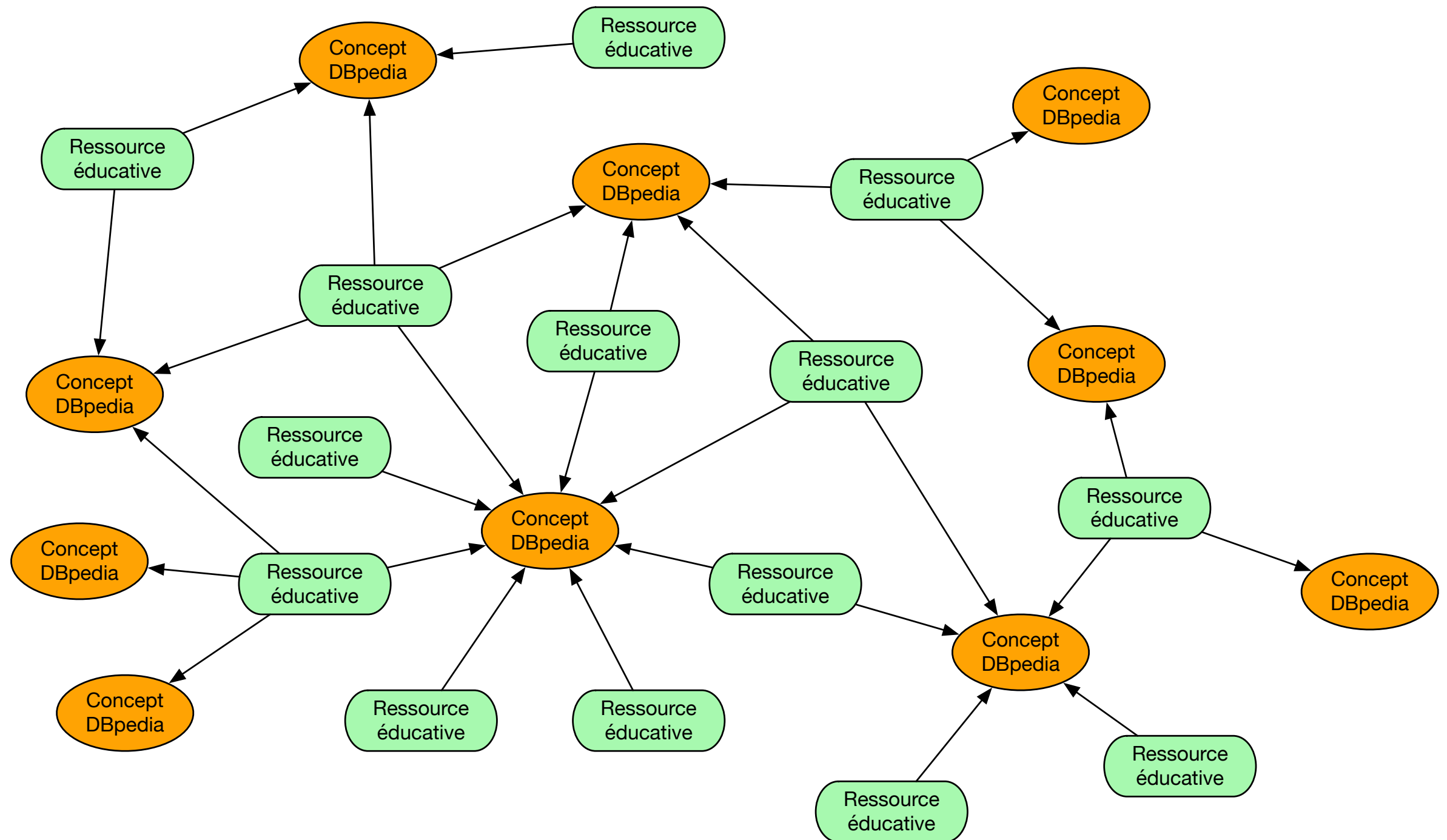
- Relâchement de classe avec la super-classe
- Relâchement de propriété avec la super propriété
- Relâchement simple d'une donnée avec une variable

<i>Q (original query)</i>	<i>Result no licensable</i>	<i>Q'3b4b with Simple relaxations</i> Sim=0.44 in F1	<i>Result licensable</i> <i>if D3 excluded</i>
<pre>SELECT * WHERE { ?student rdf:type ex:Student . #tp1@{D3} ?student ex:enrolledIn ?course . #tp2@{D3} ?course ex:heldAt ex:UniversityOfNantes . #tp3@{D1} ex:Ben ex:teaches ?course . #tp4@{D2} }</pre>		<pre>SELECT * WHERE { ?student rdf:type ex:Student . #tp1@{D2,D3} ?student ex:enrolledIn ?course . #tp2@{D2,D3} ?course ex:heldAt ?y . #tp3b@{D1} ?x ex:teaches ?course . #tp4'b@{D2} }</pre>	
<i>Q'4b with Simple relaxation</i> Sim=0.66 in F1	<i>Result licensable</i> <i>if D3 excluded</i>	<i>Q'4d with Simple and Property relaxations</i> Sim=0,33 in F2	<i>Result licensable</i> <i>if D2 excluded</i>
<pre>SELECT * WHERE { ?student rdf:type ex:Student . #tp1@{D2,D3} ?student ex:enrolledIn ?course . #tp2@{D2,D3} ?course ex:heldAt ex:UniversityOfNantes . #tp3@{D1} ?x ex:teaches ?course . #tp4'b@{D2} }</pre>		<pre>SELECT * WHERE { ?student rdf:type ex:Student . #tp1@{D2,D3} ?student ex:enrolledIn ?course . #tp2@{D2,D3} ?course ex:heldAt ex:UniversityOfNantes . #tp3@{D1} ?x ex:attends ?course . #tp4'd@{D2,D3} }</pre>	

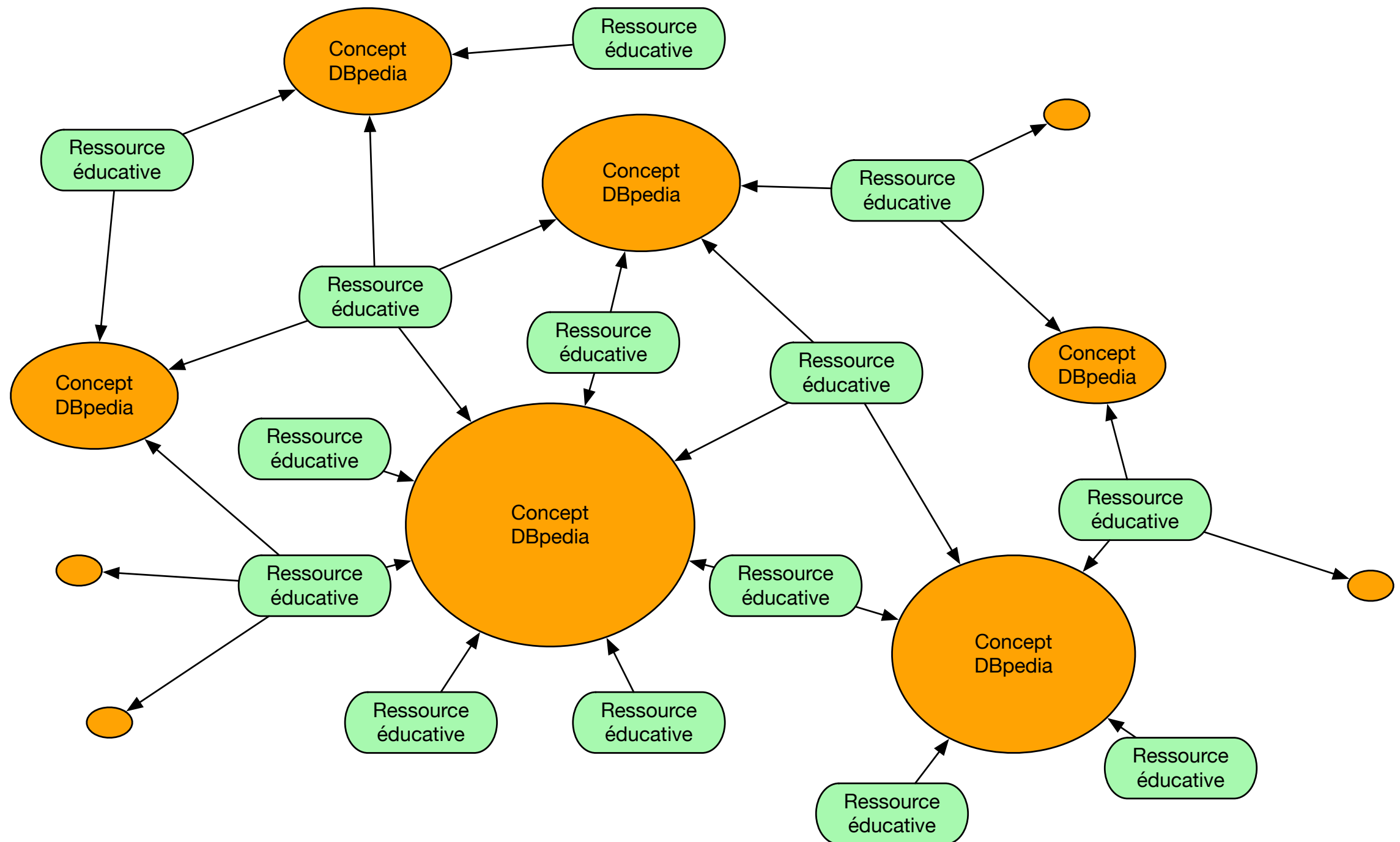
Opportunités en recherche 1/2

- ❖ Comment mesurer l'importance des concepts
 - ❖ Via différentes métriques : connectivité, pertinence
 - ❖ Avec des algorithmes de scoring, e.g., pagerank
- ❖ Comment identifier les domaines représentés
 - ❖ dans le graphe de connaissances
 - ❖ dans chaque RE

Importance des concepts dans le graphe



Importance des concepts dans le graphe



Conclusions

- ❖ CLARA est un projet sur 3 ans, nous sommes à la fin de la première année.
- ❖ A la fin de la seconde année on fera de testes utilisateur
 - ❖ Evaluer la qualité des RE résultat d'une recherche à partir de mots-clé ou d'un *sketch de cours*.
 - ❖ Evaluer l'expérience utilisateur