

INTERROGER LE WEB DE DONNÉES

UDOS - 30 JUIN 2017

Pierre-Antoine Champin (UCBL/LIRIS)

<http://champin.net/2017/udos>

WEB DE DONNÉES

QU'EST-CE QUE LE WEB?

Un espace documentaire

- décentralisé (HTTP)
- interconnecté (URL)
- interopérable (HTML)

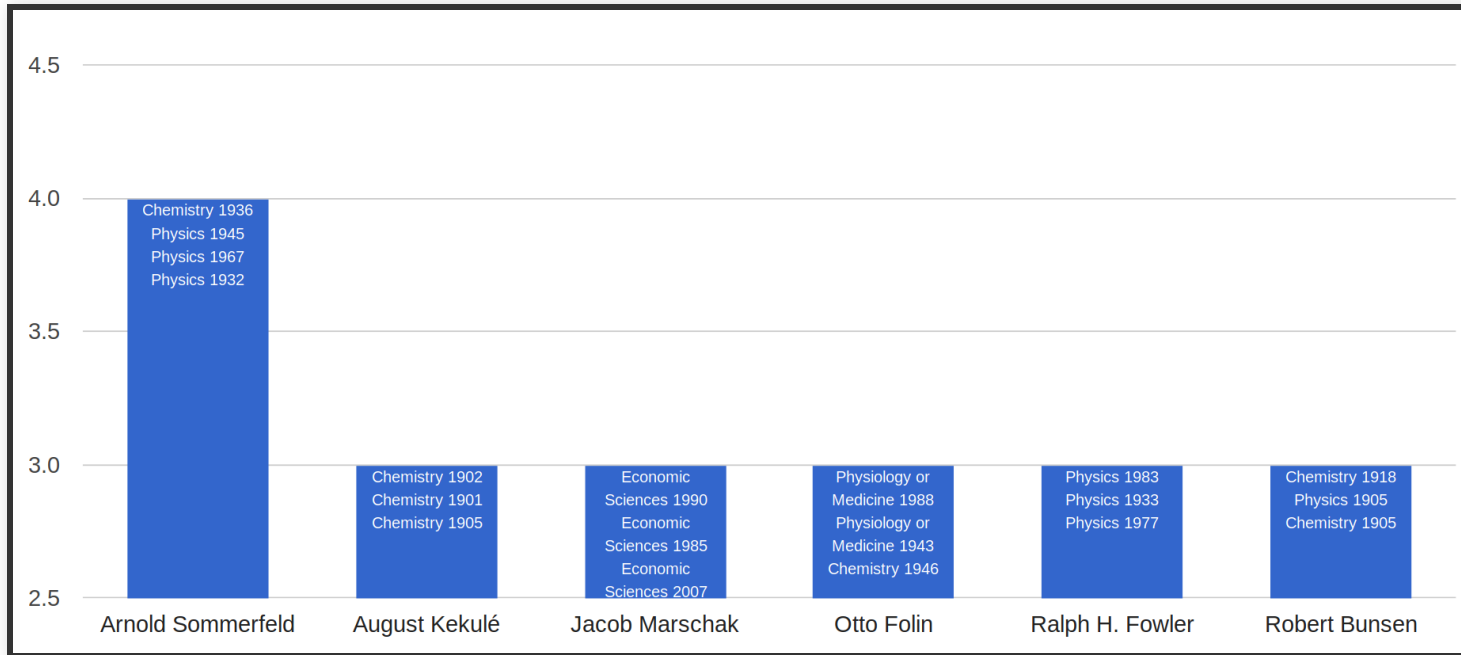
QU'EST-CE QUE LE WEB?

Un espace documentaire

- décentralisé (HTTP 2014, HTTP 2.0)
- interconnecté (URL, URI, IRI)
- interopérable (HTML5, CSS, JS)

LIMITES

Combien de scientifiques n'ont jamais reçu de prix Nobel personnellement, mais ont "produit" au moins trois prix Nobel différents par étudiant interposé ?

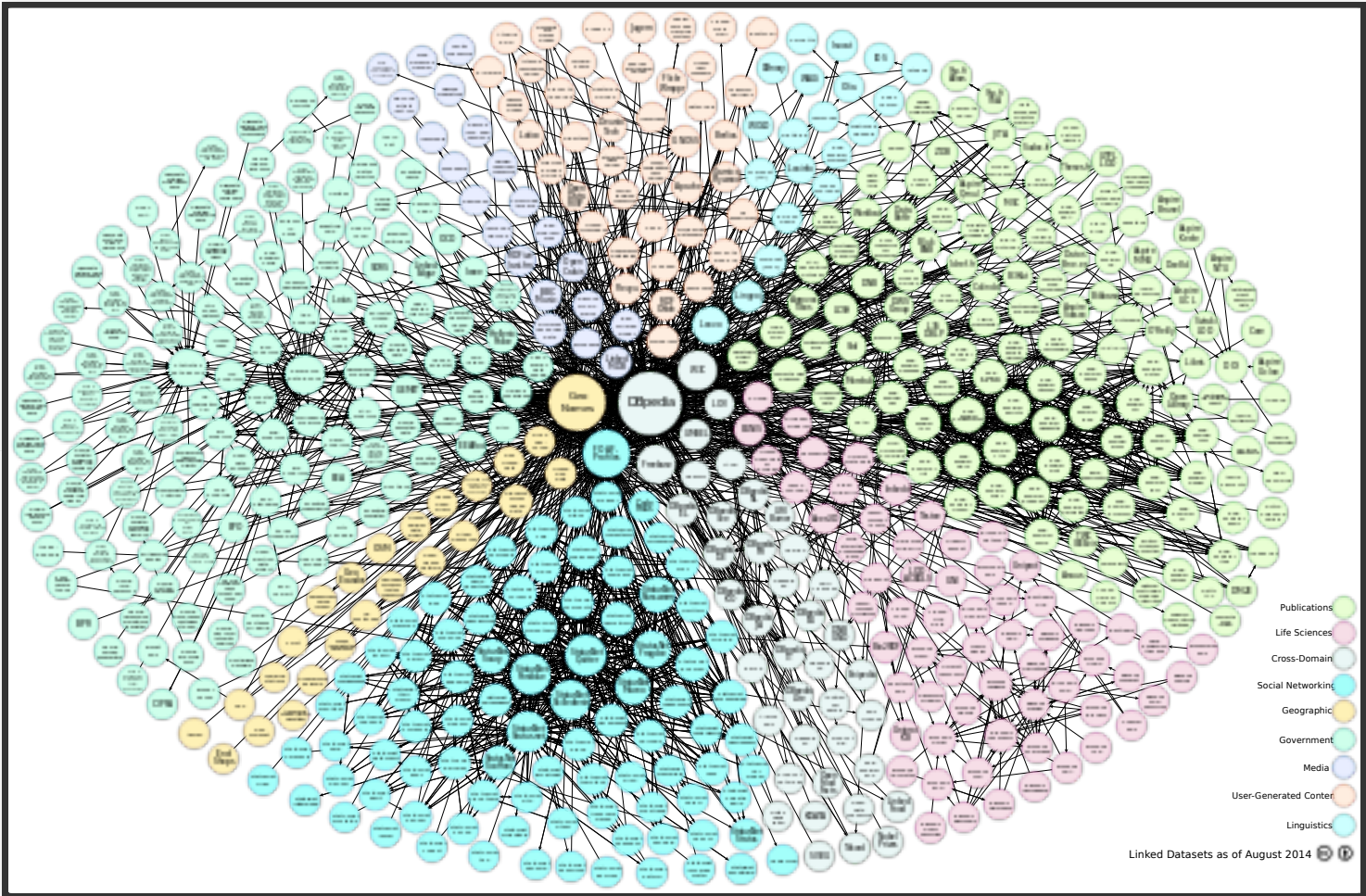


<<http://yasgui.org/short/B1EEbOQN->>

VERS UN WEB DE DONNÉES

Un espace de **données**

- décentralisé (HTTP)
- interconnecté (URL)
- interopérable (?)



VERS UN WEB DE DONNÉES

Un espace de **données**

- décentralisé (**HTTP**)
- interconnecté (**URL**)
- interopérable (**RDF**)



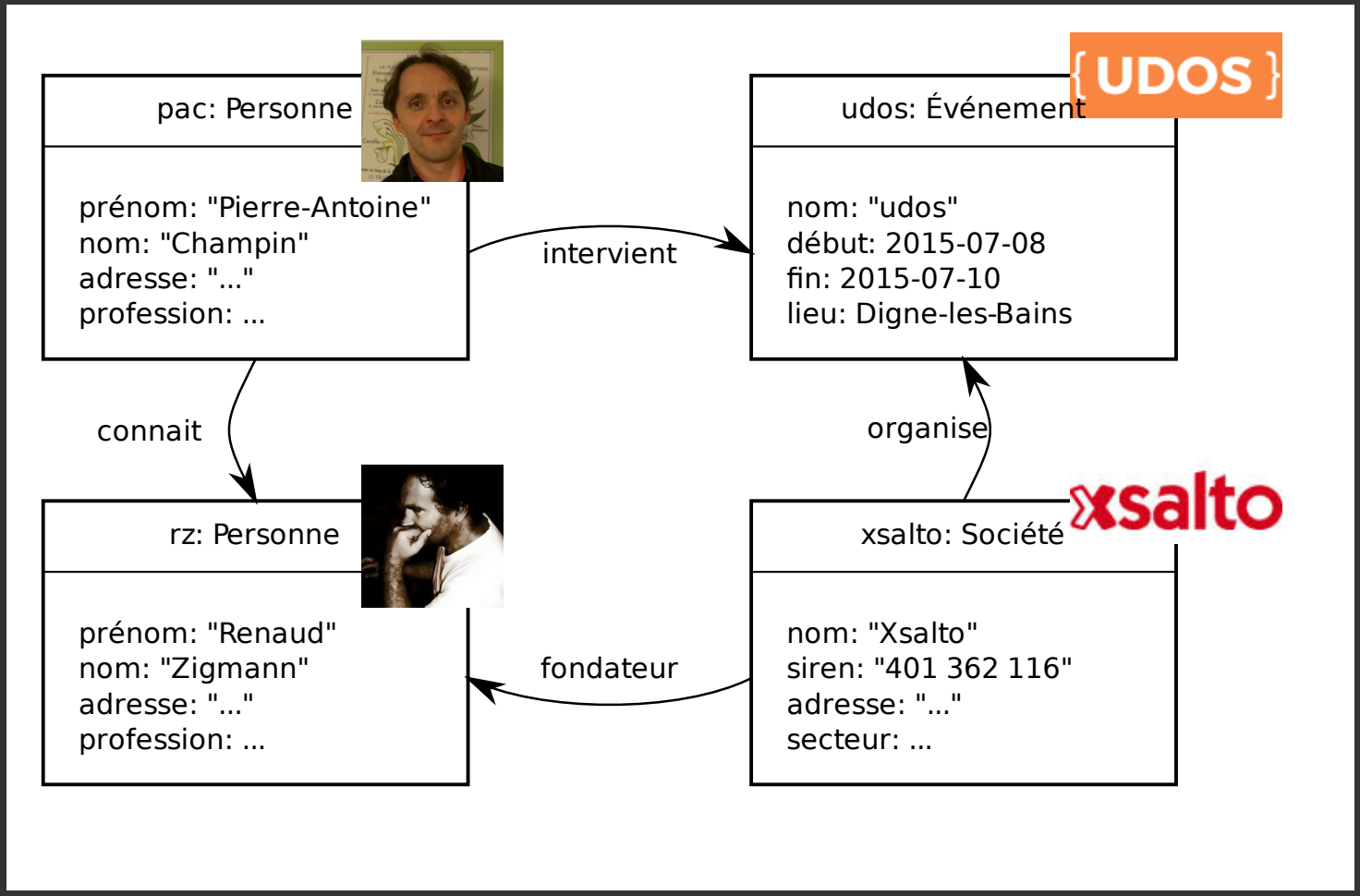
LE BOURGEOIS GENTILHOMME.

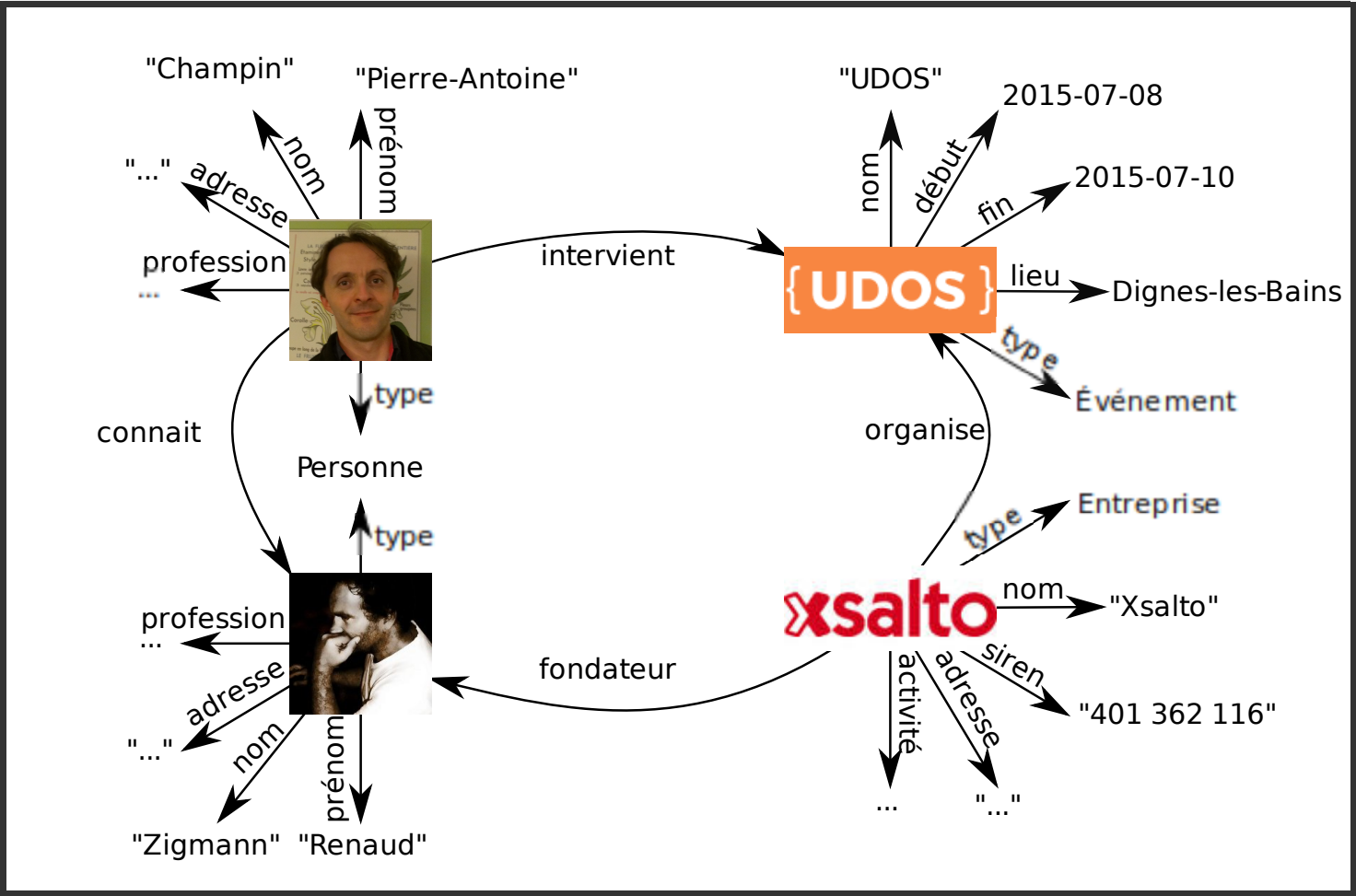
M. JOURDAIN.

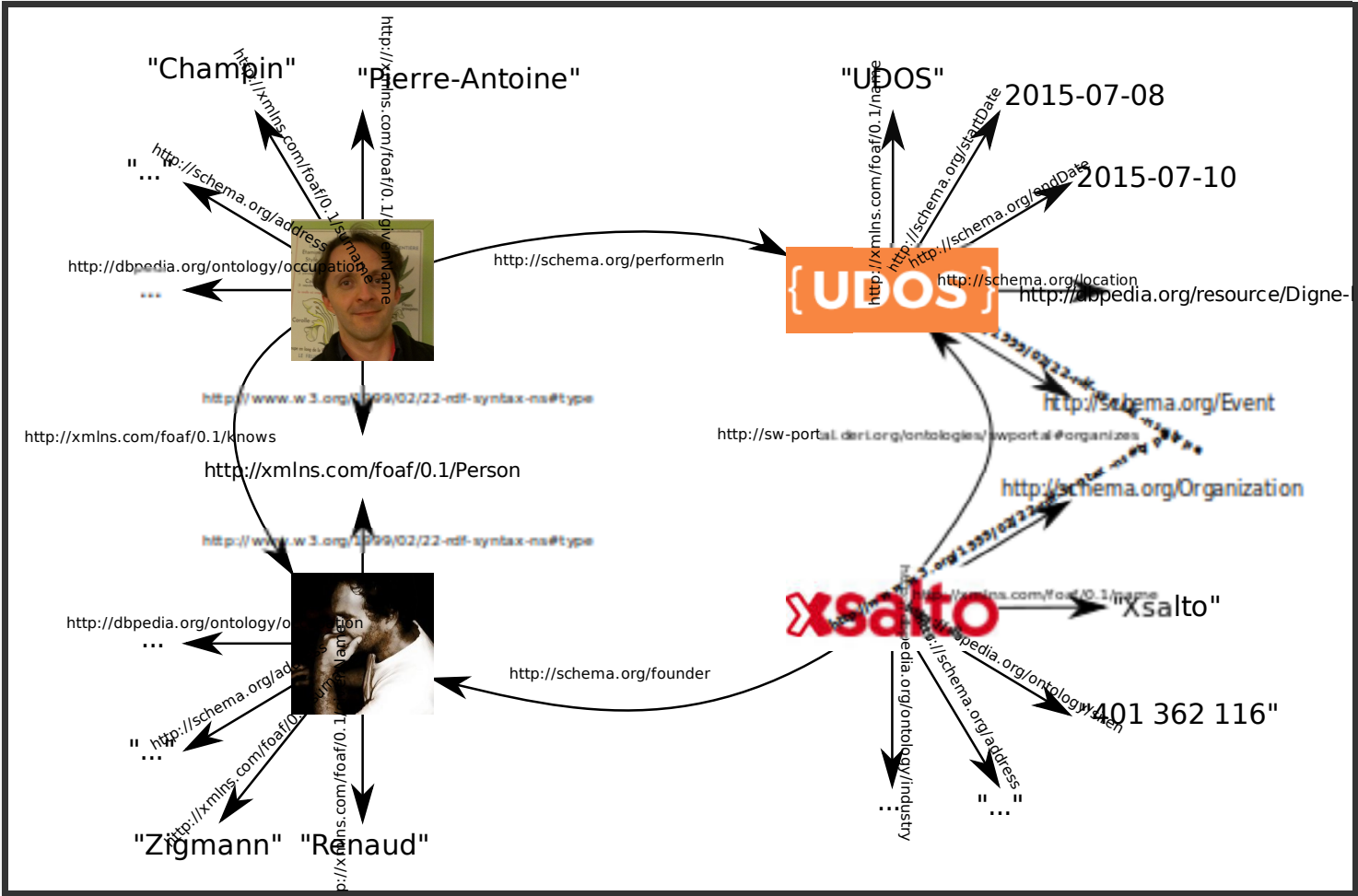
Suivez-moi, que j'aie un peu montrer mon habit par la ville.

sc. II. cc. 1.

RDF







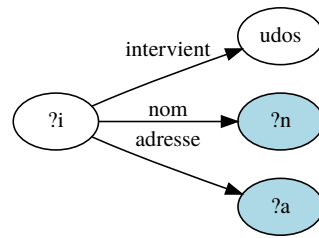
VOCABULAIRE / ONTOLOGIE

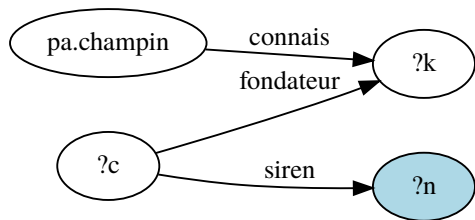
- Ensemble d'IRIs décrivant
 - les catégories du domaine (classes)
 - les propriétés (attributs, relations)
 - les objets du domaine

SPARQL

PRINCIPE

Décrire les relations entre les données que l'on cherche sous forme d'un *sous-graphe*.





PRÉFIXES

PREFIX s: <http://schema.org/>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

...

s:Event → <http://schema.org/Event>

s:startDate → <http://schema.org/startDate>

s:endDate → <http://schema.org/endDate>

owl:sameAs → <http://www.w3.org/2002/07/
/owl#sameAs>

...

TERMES

<code>s:Event</code>	IRI abrégé
<code><http://example.org/></code>	IRI complet
<code>"hello world"</code>	Litéral chaîne
<code>42</code> <code>3.14</code> <code>1e-10</code>	Litéraux numérique
<code>?x</code>	Variable

TRIPLET

Chaque arc du graphe est représenté par un triplet :

`<http://champin.net/#pa> rdf:type s:Person .`



GRAPHE

Le graphe requête est l'union des arcs.

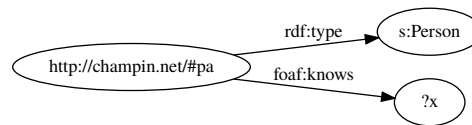
```
<http://champin.net/#pa> rdf:type s:Person .  
<http://champin.net/#pa> foaf:knows ?x .  
?x foaf:surname "Zigmann" .
```



FACTORISATION (1)

Même sujet :

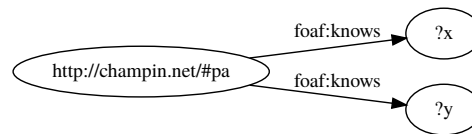
```
<http://champin.net/#pa>  
  rdf:type s:Person ;  
  foaf:knows ?x .
```



FACTORISATION (2)

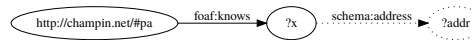
Même sujet et même prédicat :

`<http://champin.net/#pa> foaf:knows ?x, ?y .`



SOUS-GRAPHE OPTIONNEL

```
<http://champin.net/#pa> foaf:knows ?x.  
OPTIONAL { ?x schema:address ?addr }
```



par opposition à

```
<http://champin.net/#pa> foaf:knows ?x.  
?x schema:address ?addr .
```

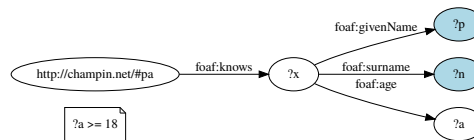


FILTRES

```
<http://champin.net/#pa> foaf:knows ?x.  
?x s:birthDate ?bd.  
FILTER ( STR(?bd) > "1999-06" )
```

SÉLECTION

```
SELECT ?p ?n {  
  <http://champion.net/#pa> foaf:knows ?x .  
  ?x foaf:givenName ?p ;  
    foaf:surname ?n ;  
    foaf:age ?a .  
  
  FILTER (?a >= 18)  
}
```



À VOUS DE JOUER

<http://champin.net/2017/udos/tuto>

POUR ALLER PLUS LOIN

TROUVER DES SOURCES DE DONNÉES

- Datahub
- SPARQL endpoint status
- *follow your nose* with VOID

YASGUI, FONCTIONS AVANCÉES

- représentations graphiques
- tableau croisé

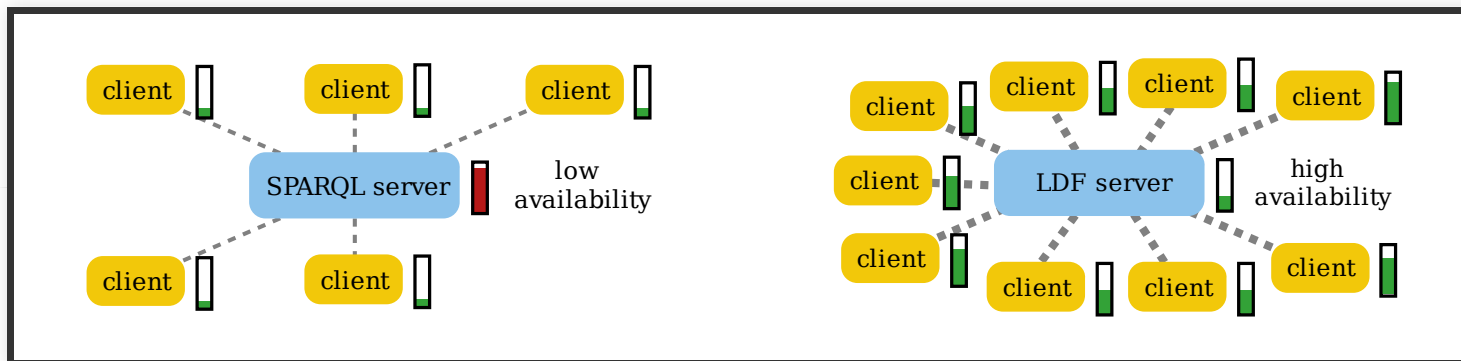
REQUÊTES FÉDÉRÉES

```
# (... déclarations de préfixe ...)  
SELECT ?s {  
  [] a nobel:NobelPrize ;  
    nobel:category cat:Physics ;  
    nobel:year 1965 ;  
    nobel:laureate ?l .  
  ?l foaf:name ?name ; owl:sameAs ?iri .  
  
  SERVICE <http://dbpedia.org/sparql> {  
    ?iri dbo:doctoralStudent ?s .  
  }  
}
```

<<http://yasgui.org/short/r1S4PLfE->>

LINKED DATA FRAGMENTS

<http://linkeddatafragments.org/>



RDB 2 RDF

- Intégration de données
- Publication (sélective) sur le Web de données
- [D2R](#) et autres implémentations

CHOISIR UN VOCABULAIRE

<http://lov.okfn.org/>

MERCI

PUBLIER DES DONNÉES LIÉES

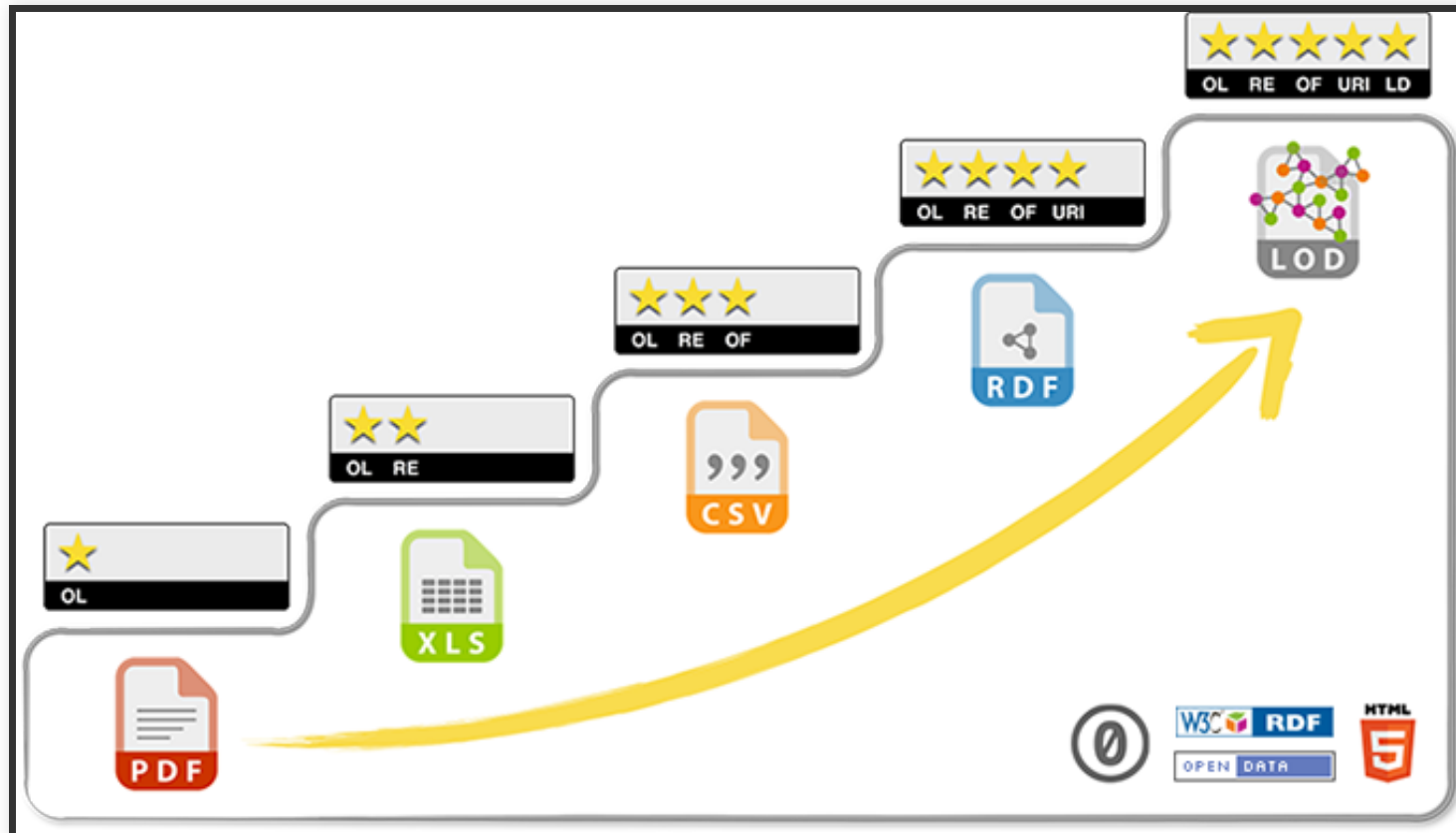
UDOS - 30 JUIN 2017

Pierre-Antoine Champin (UCBL/LIRIS)

<http://champin.net/2017/udos>

DE RDF AU WEB DE DONNÉES

5 ÉTOILES



<http://5stardata.info/fr/>

LES QUATRE PRINCIPES

- Utiliser des *IRIs* (URLs) pour nommer les objets d'intérêt
- Y donner accès à une *description* de ces ressources
- ... dans un format de données *standard*
- ... et *pointant* vers les IRIs d'autres objets d'intérêt

<http://www.w3.org/DesignIssues/LinkedData.html>

FOLLOW YOUR NOSE

<http://dbpedia.org/resource/Digne-les-Bains>

Négociation de contenu

→ Le site Web *est* l'API

DE QUOI AI-JE BESOIN ?

Un triple store (e.g. [Virtuoso](#))

...ou pas

VOCABULAIRES

VOCABULAIRE / ONTOLOGIE

- Ensemble d'IRIs décrivant
 - les types d'objet d'un domaine (classes)
 - les propriétés (attributs, relations)
 - les objets notables du domaine
- Décrits à l'aide d'un *méta-vocabulaire*
 - [RDFS](#) (sémantique simple)
 - [OWL](#) (sémantique plus complexe)
 - [SHACL](#) (contraintes structurelles)

OÙ TROUVER DES VOCABULAIRES ?

- [Linked Open Vocabulary](#)
- [Schema.org](#)

J'AI DÉJÀ UN SITE WEB

RDFA : RDF IN ATTRIBUTES

RDFa est une utilisation d'attributs (existants ou supplémentaires) de (X)HTML pour y inclure du RDF (à la manière des micro-formats) :

- facilite la migration de contenus HTML vers RDF
- facilite la maintenance en cohérence de la version HTML et des données RDF (*DRY : Don't Repeat Yourself*)

Syntaxe: <http://www.w3.org/TR/rdfa-primer/>

Valideur: <http://rdfa.info/tools>

Distiller: <http://www.w3.org/2012/pyRdfa/>

DÉMONSTRATION

<http://rdfa.info/play>

`<p>`

Pierre-Antoine Champin est l'un des intervenants de `UDOS 2017`, les 29 et 30 juin.

`</p>`

RDFA 101

vocab	préfixe par défaut
prefix	permet de spécifier d'autres préfixes (<i>alla</i> SPARQL)
resource	IRI de la ressource décrite
typeof	type(s) de la ressource d'écrite
property	spécifie la relation entre la ressource décrite et le "contenu" de la balise

J'AI DÉJÀ UNE API

JSON-LD

- JSON-LD (JSON Linked Data) permet d'interpréter une structure JSON comme du RDF,
- grâce à un *contexte* (implicite ou explicite).
- Objectif : faciliter l'adoption de RDF (syntaxe abstraite) auprès des développeurs d'applications web.

Syntaxe: <http://www.w3.org/TR/json-ld-syntax/>

Valideur: <http://json-ld.org/playground/>

DÉMONSTRATION

<http://json-ld.org/playground/>

```
{ "type": "event",  
  "homepage": "http://udos.fr/",  
  "from": "2017-06-29",  
  "to": "2017-06-30",  
  "with": {  
    "id": "http://champin.net/#pa",  
    "type": "person",  
    "first_name": "Pierre-Antoine",  
    "last_name": "Champin"  
  }  
}
```

TRANSFORMATIONS DE DONNÉES AVEC JSON-LD

- expand
- compact
- flatten
- frame

J'AI SEULEMENT UNE BASE DE
DONNÉES

D2RQ

<http://d2rq.org/>

- mapping SQL → RDF

POUR ALLER PLUS LOIN

- intégration de plusieurs bases
- mapping SPARQL → SQL

ET SPARQL ALORS?

UTILISATION EN LOCAL

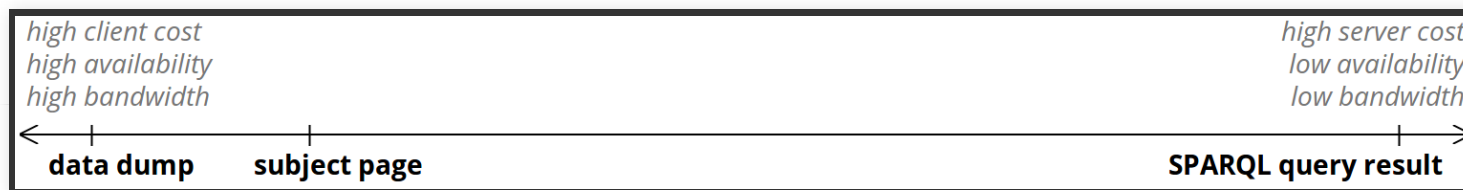
- sur les données agrégées en parcourant le Web

SPARQL PROTOCOL

- échange de requête et de résultats sur HTTP
- complémentaire de *follow your nose*
- découvrable *via* VOID

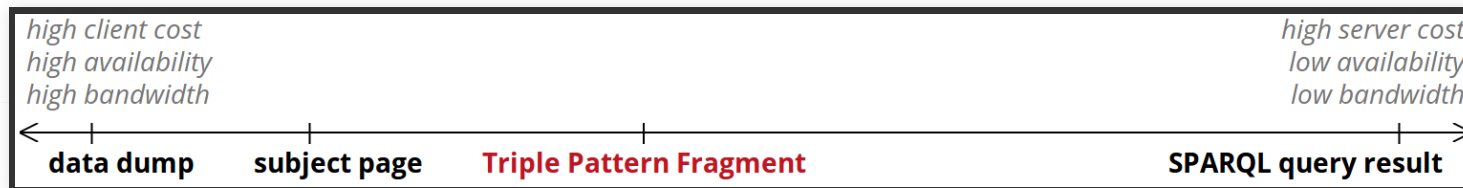
<http://data.nobelprize.org/page/laureate/6>

LINKED DATA FRAGMENT



<http://linkeddatafragments.org/>

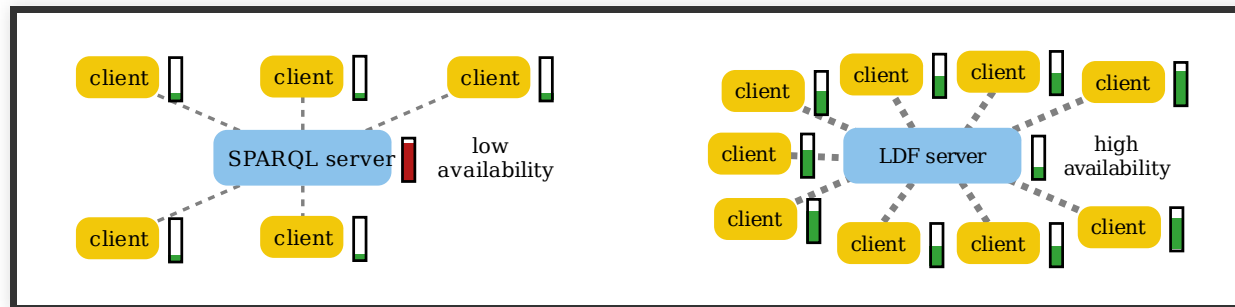
LINKED DATA FRAGMENT



<http://linkeddatafragments.org/>

TRIPLE PATTERN FRAGMENT

- Le serveur se content de répondre à des requêtes très simples ([exemple](#))
- Le client décompose sa requête complexe en de multiples requêtes simples. ([exemple](#))
- Fédération de requête transparente ([exemple](#))



<http://linkeddatafragments.org/>