Ontologies et données FAIR

Cassia Trojahn

MEthodes et ingénierie des Langues, des Ontologies et du DIscours (MELODI)
Institut de Recherche en Informatique de Toulouse (IRIT)
cassia.trojahn@irit.fr

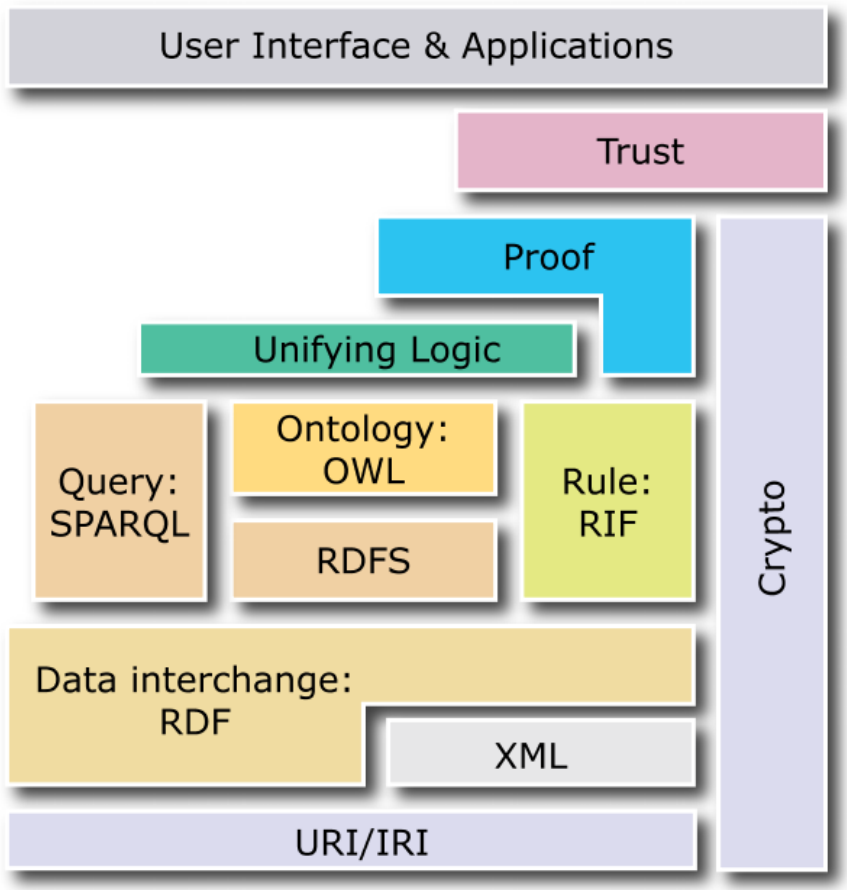
Des données aux connaissances

- Plusieurs sources de données (hétérogènes) sur le Web
- Besoin de prendre des décisions à partir de ces données
- Besoin de les lier, de raisonner sur ces données
- Besoin de produire de la connaissance

La vision du Web sémantique

"Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data"

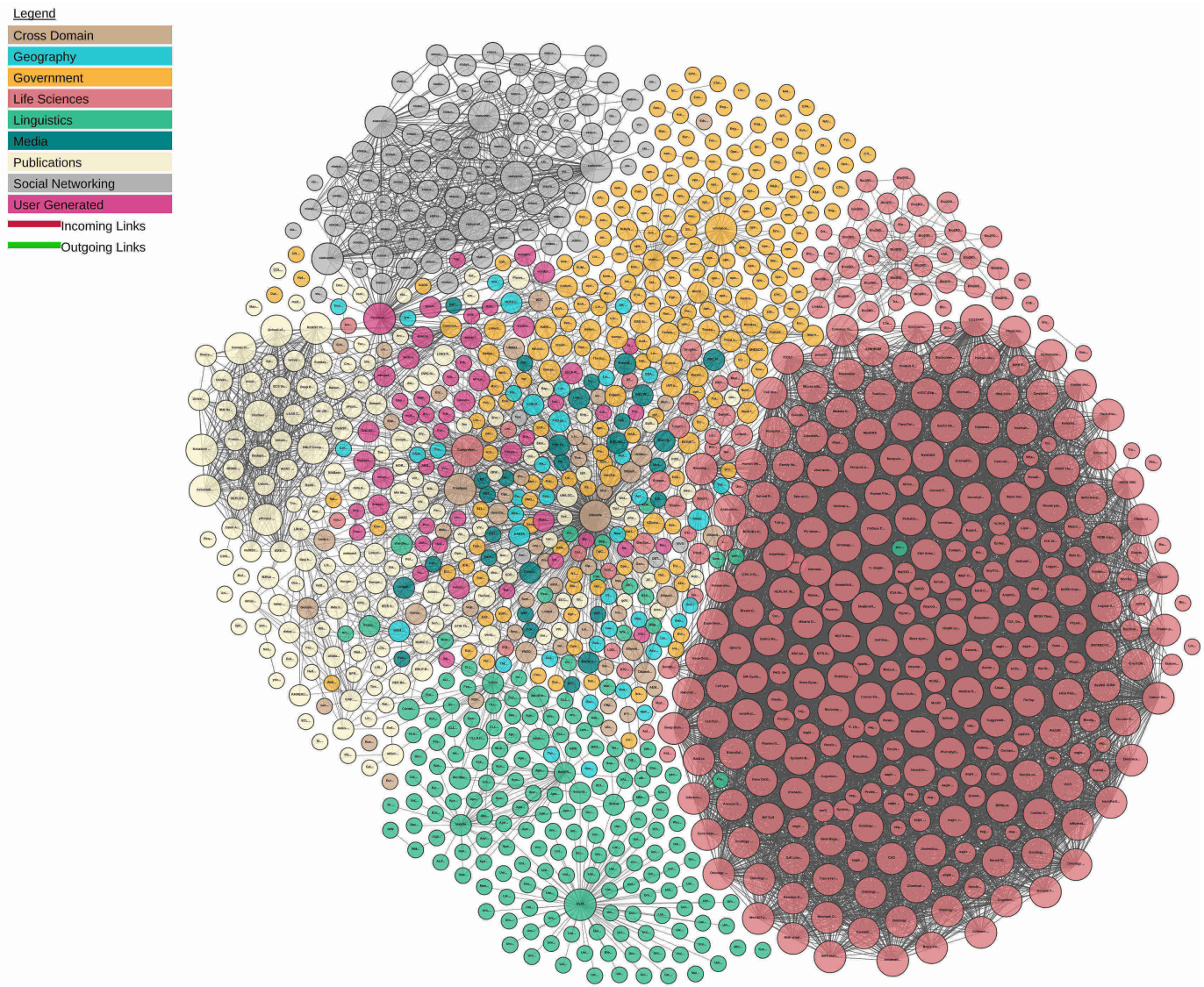
W3C



La vision du Web de données liées

“’Semantic Web’ refers to W3C’s vision of the Web of linked data”

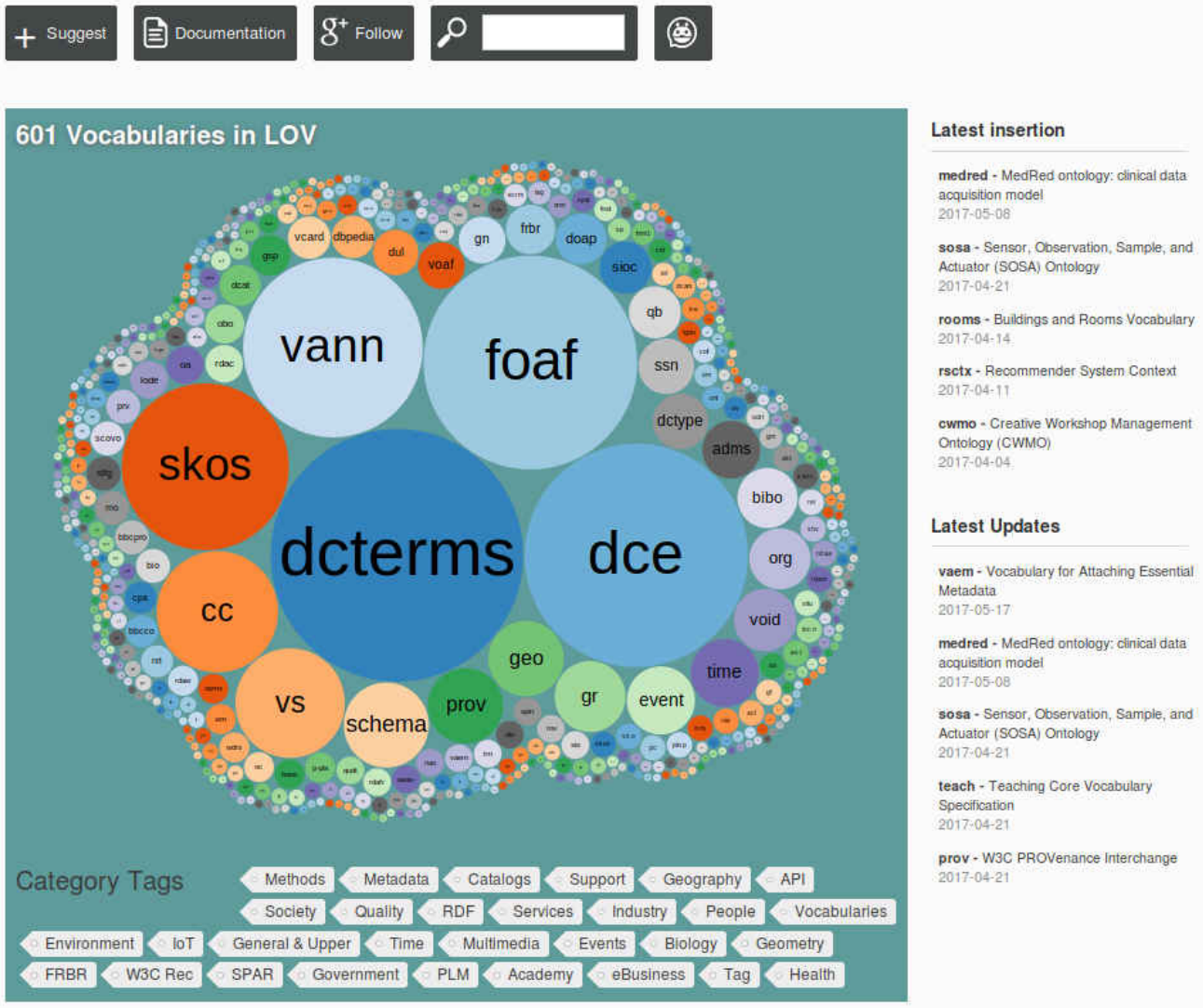
W3C



http://lod-cloud.net (2017-02-20)

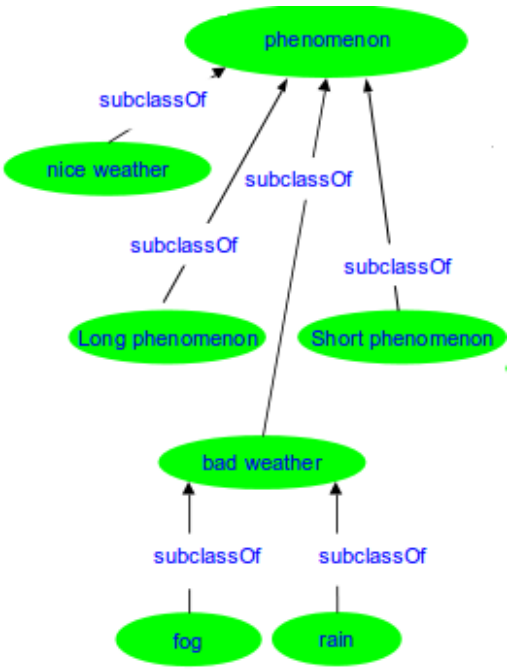
Comment décrire les données ?

Linked Open Vocabularies (LOV)



https://lov.okfn.org/dataset/lov/

Des ontologies plus expressives pour supporter le raisonnement



TBox :

(1) $Phenomenon \sqsubseteq hasDuration$

(2) $ShortPhenomenon \equiv Phenomenon \sqcap \exists hasDuration.\leq 15$

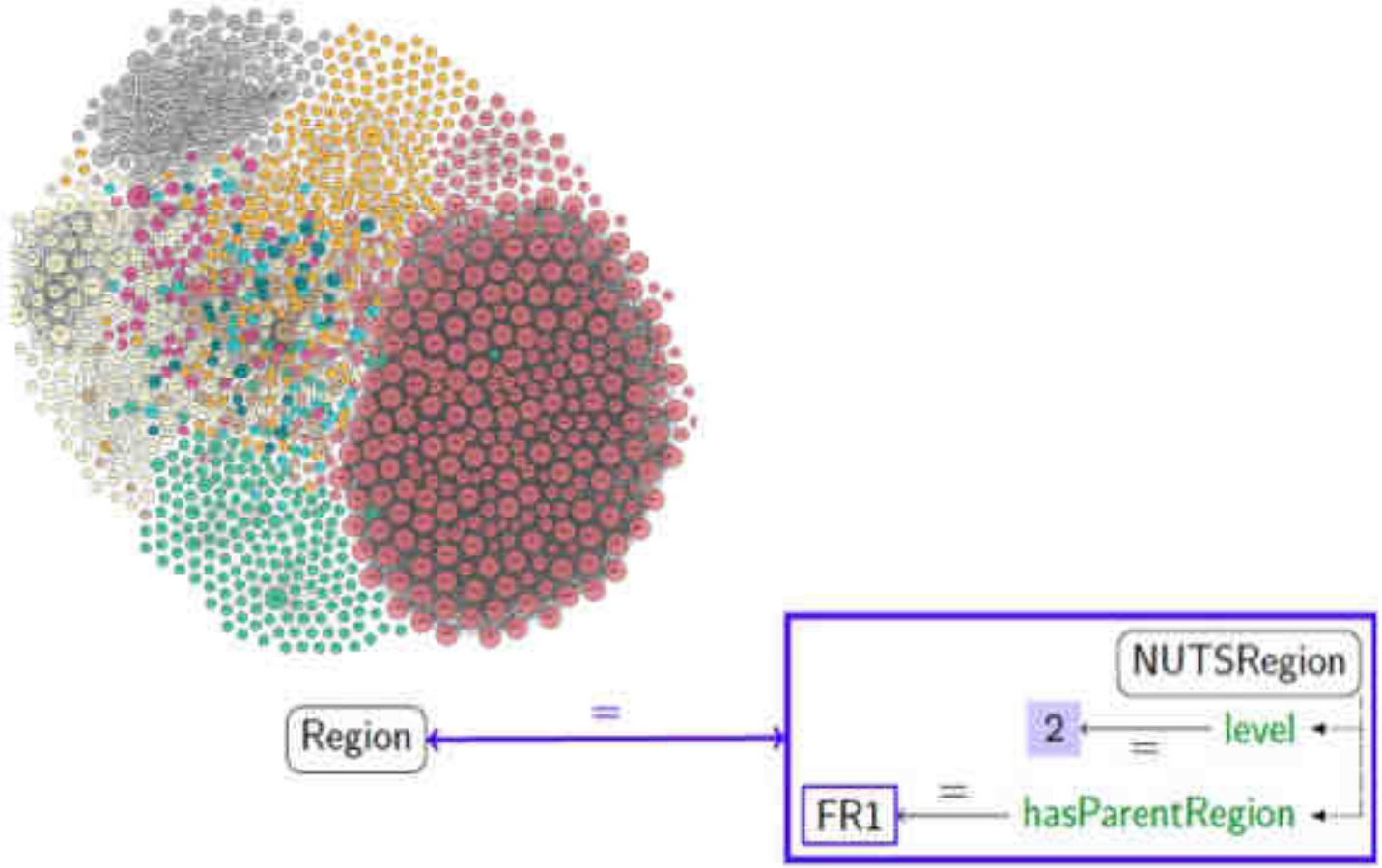
ABox :

```
What we assert :
     P1 hasDuration 15 min

What we infer :
     P1 is a Phenomenon
     P1 is a ShortPhenomenon
```

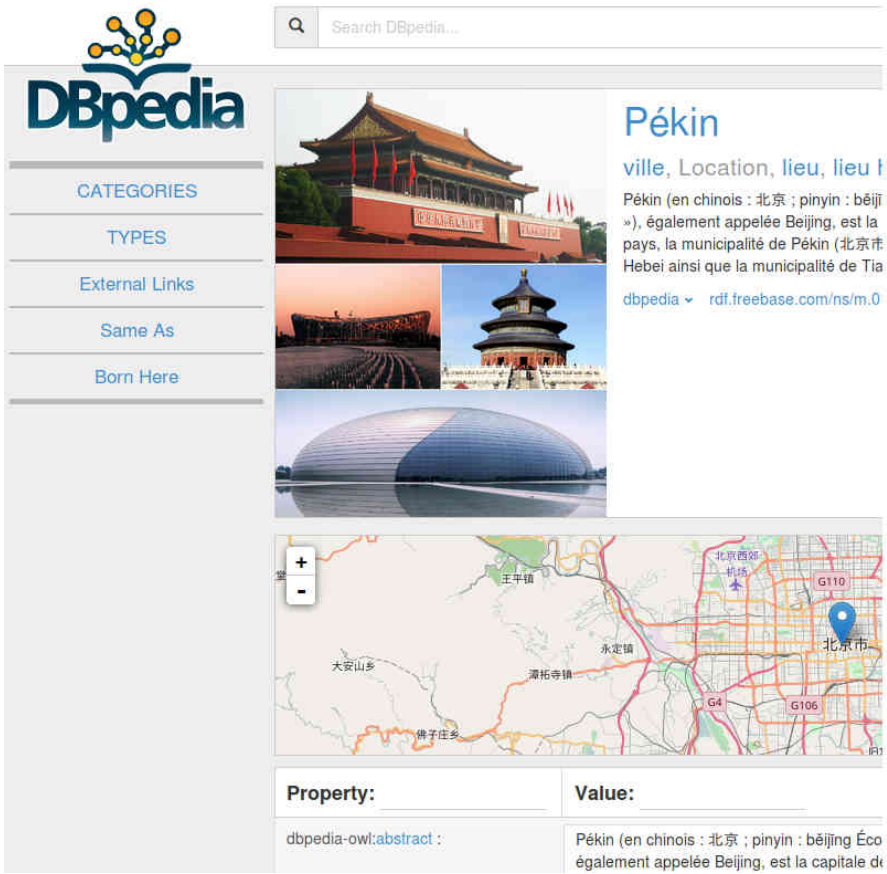
Comment lier les données ?

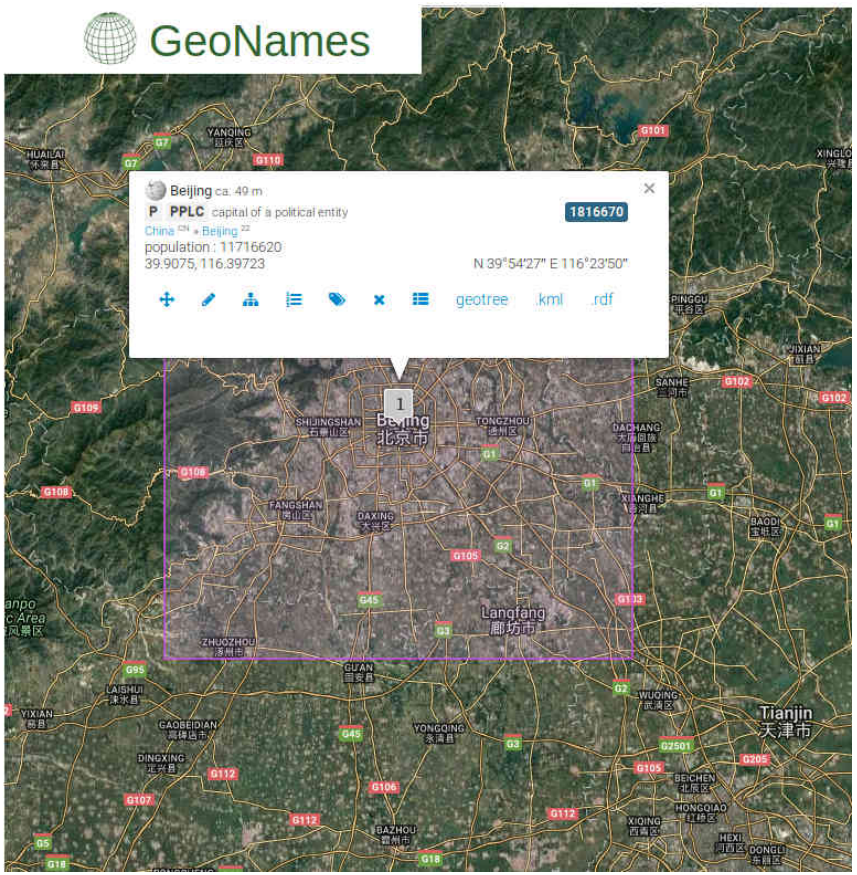
L'alignement pour résoudre l'hétérogénéité : comment aligner les vocabulaires ?



[Euzenat and Shvaiko, 2013]

L'alignement pour résoudre l'hétérogénéité : comment aligner les instances ?

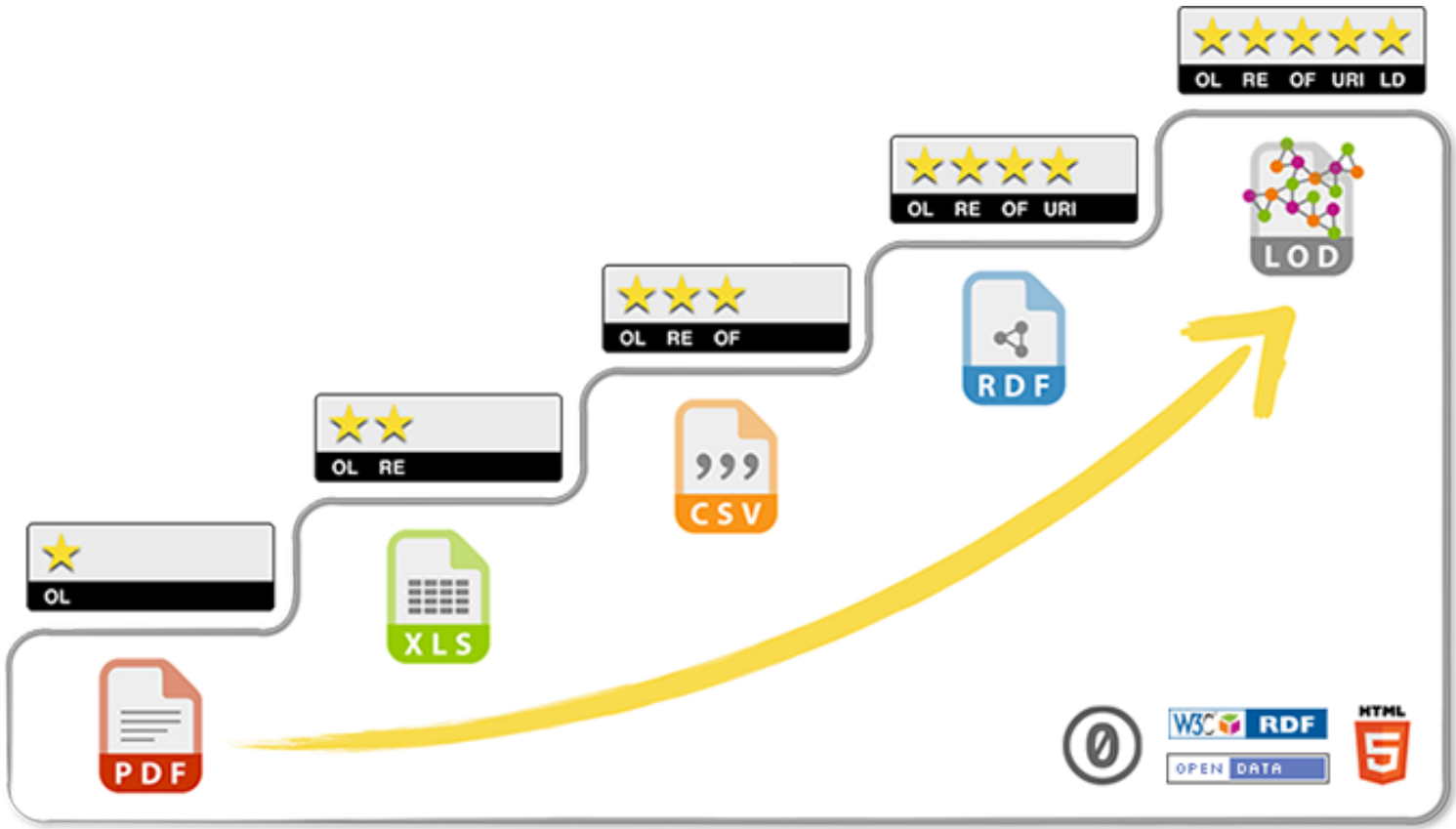




[Daskalaki et al., 2016]

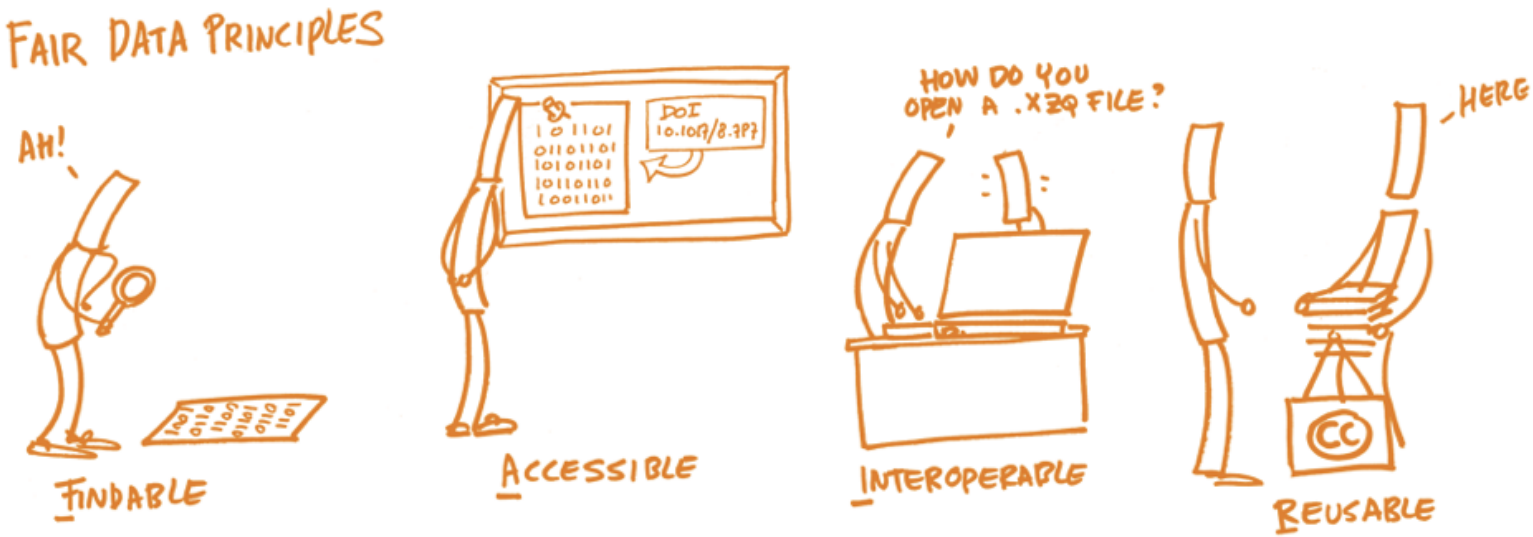
Des données ouvertes aux données ouvertes et liées

Données 5 étoiles (by Tim Berners-Lee)

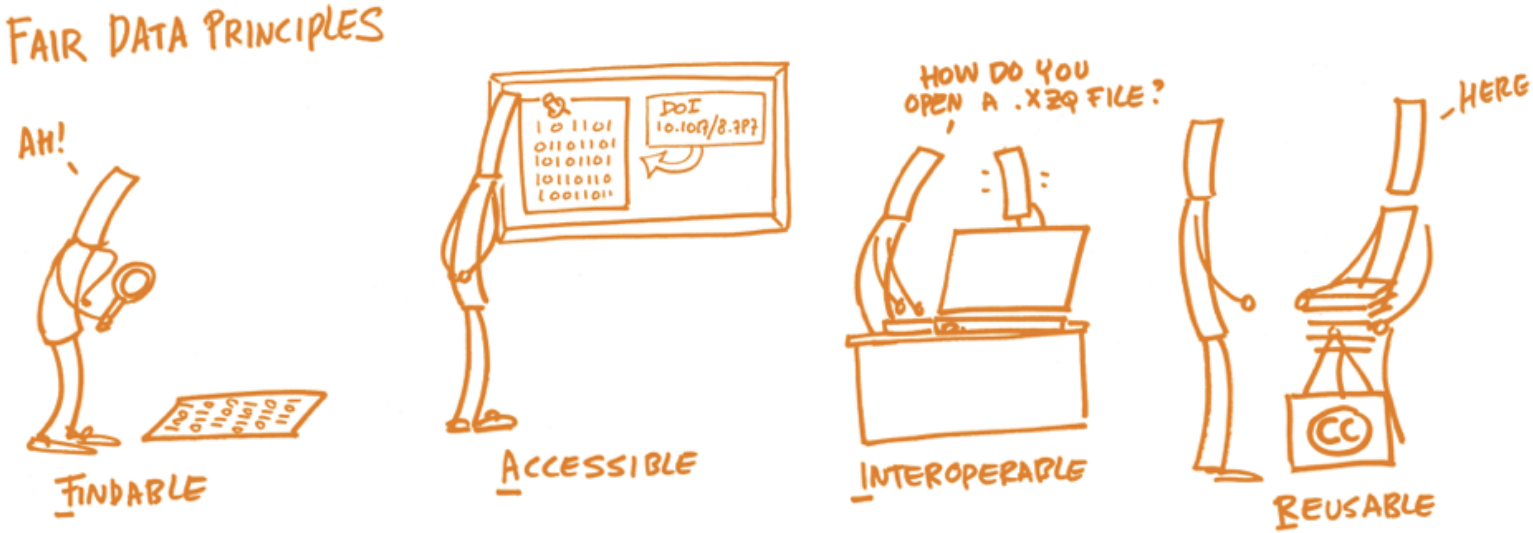


http://5stardata.info/fr/

Les données FAIR



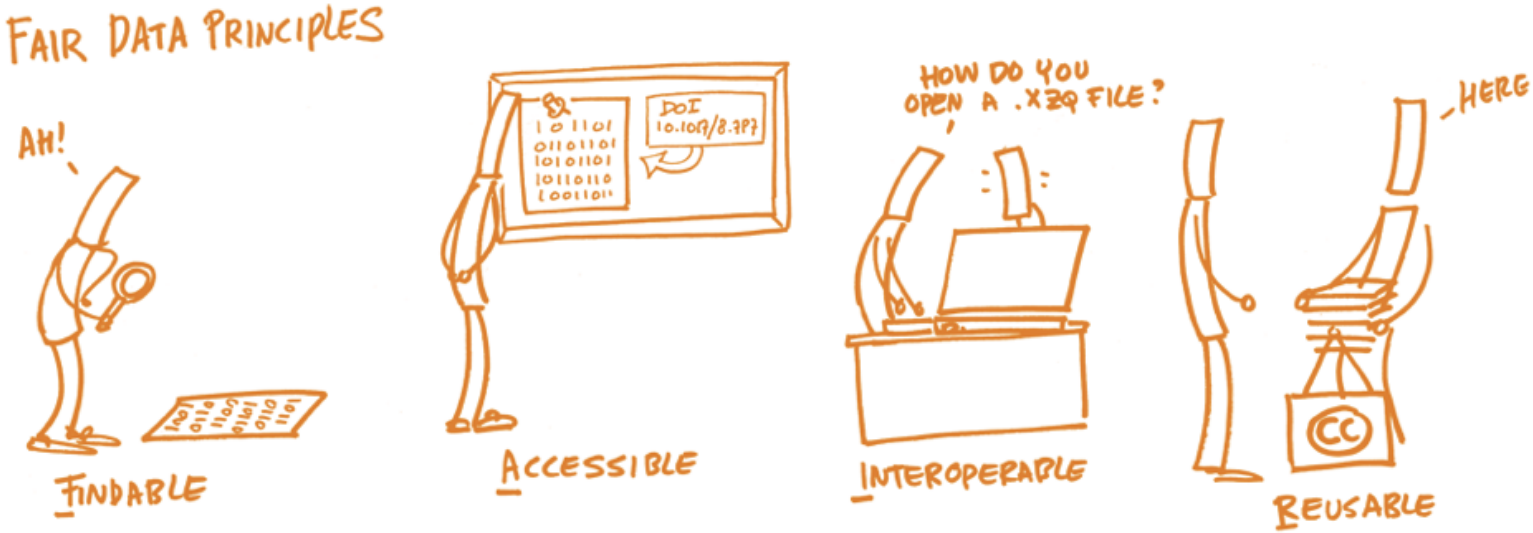
Source:https://book.fosteropenscience.eu/



Source:https://book.fosteropenscience.eu/

- **Findable**: discoverable with metadata, identifiable and locatable by means of a standard identification mechanism

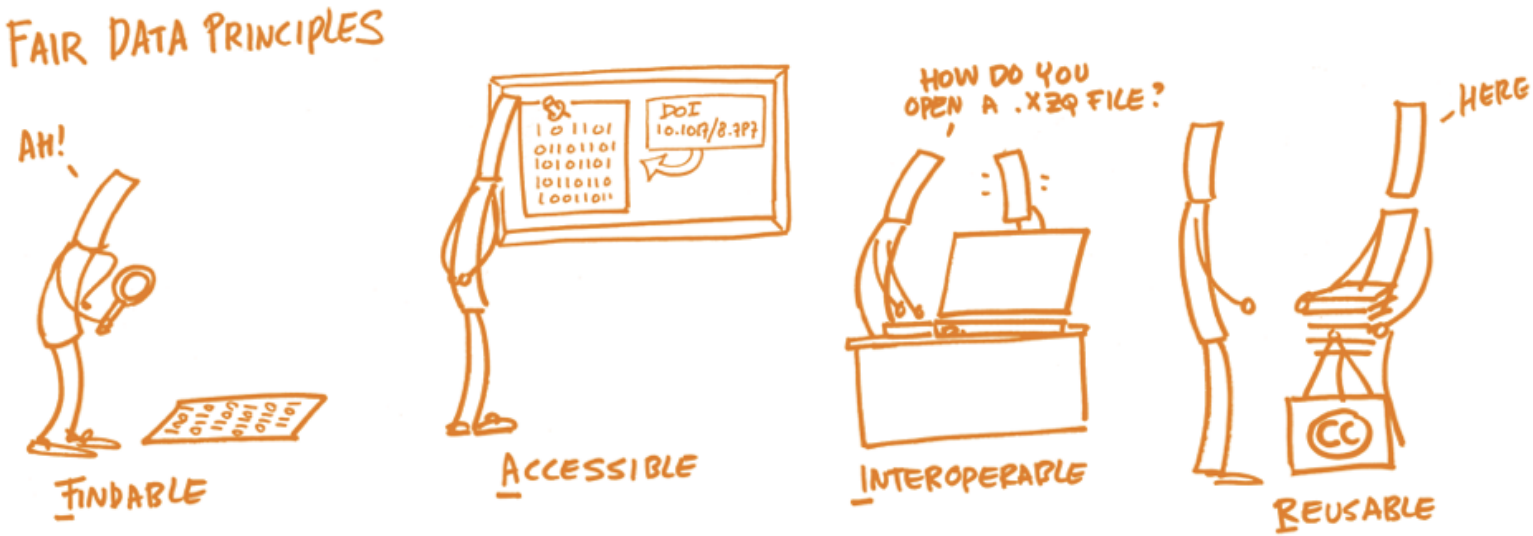
https://www.openaire.eu/how-to-make-your-data-fair



Source:https://book.fosteropenscience.eu/

- **Accessible**: always available and obtainable; even if the data is restricted, the metadata is open

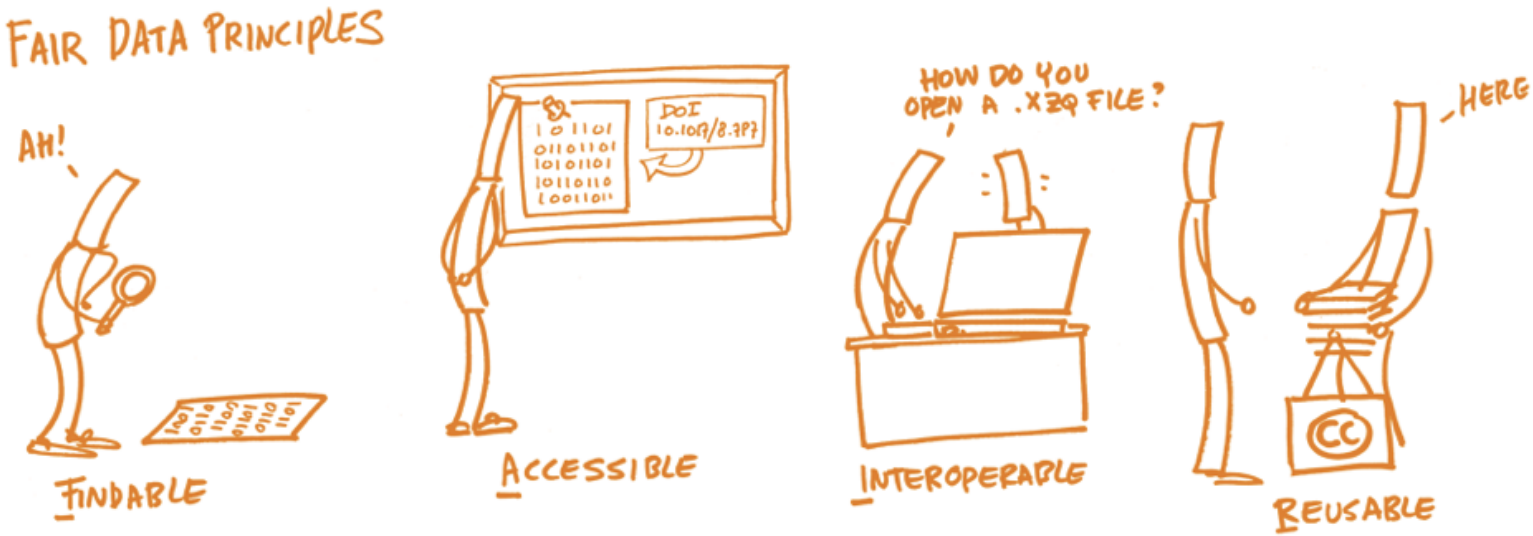
https://www.openaire.eu/how-to-make-your-data-fair



Source:https://book.fosteropenscience.eu/

- **Interoperable**: both syntactically parseable and semantically understandable, allowing data exchange and reuse between researchers, institutions, organisations or countries

https://www.openaire.eu/how-to-make-your-data-fair



Source:https://book.fosteropenscience.eu/

- **Reusable**: sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible and the least cumbersome integration with other data sources.

https://www.openaire.eu/how-to-make-your-data-fair

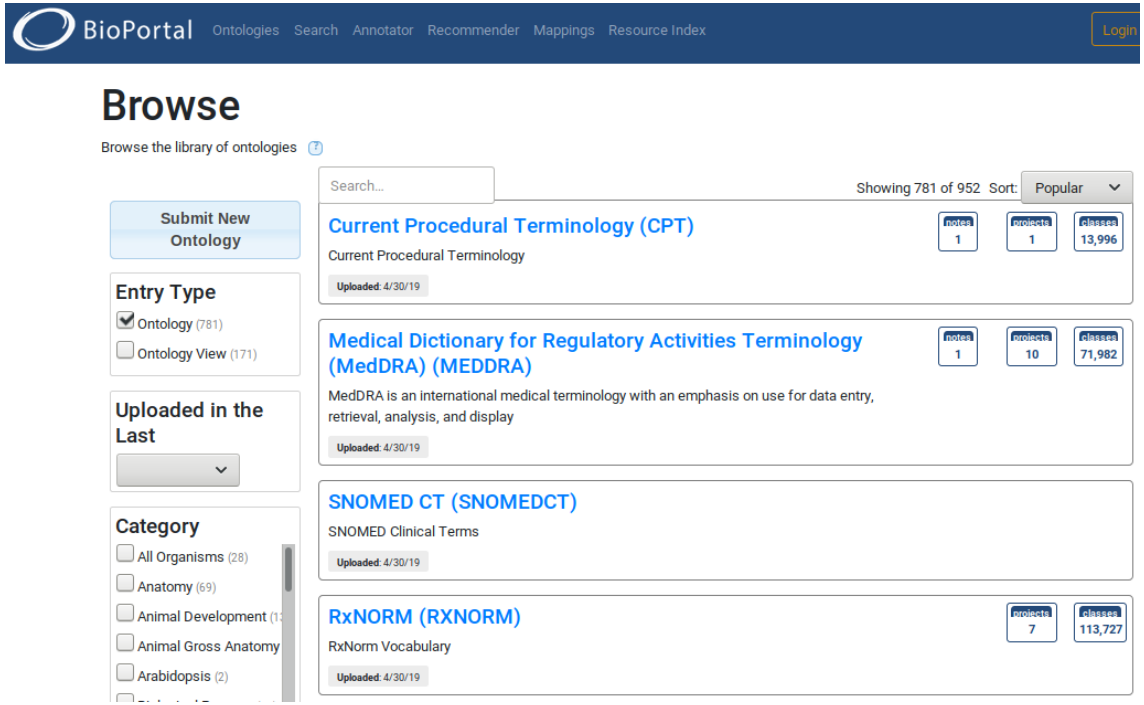
FAIR principles

$\Rightarrow$ Metadata from NCBI BioSample Stink!

- 73% of "boolean" metadata values are not actually *true false*
- 26% of "integer" metadata values cannot be parsed into integers, e.g., *JM52*, *UVPgt59.4*, *pig*
- 68% of metadata entries that are supposed to represent terms from biomedical ontologies do not actually, e.g., *presumed normal*, *wild_type*

Musen, M. Semantic Technology for Open Science: Creating an Ecosystem for FAIR Data. Ontologie, Données et Informatique médicale. May, 2019 (France).

Ontologies, vocabularies, terminologies



https://bioportal.bioontology.org/ontologies/

http://cedar.metadatacenter.net/

http://www.ontobee.org/

http://www.ebi.ac.uk/ontology-lookup/

http://obofoundry.org

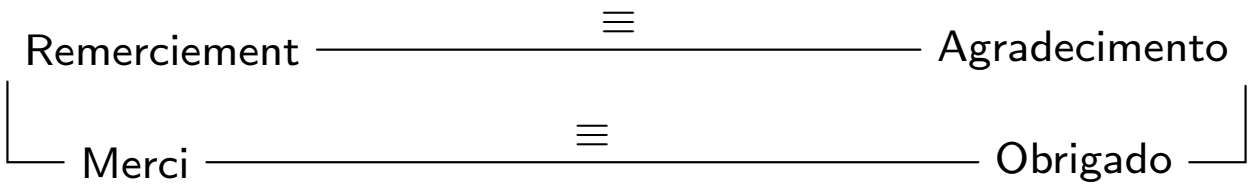
Semantics4Fair

- Construire et de réutiliser plusieurs ontologies pour prendre en compte les différents points de vue sur les données et les relations entre ces vues
  - vision des producteurs des données
  - vision des consumateurs des données
- Ces ontologies seront utilisées pour décrire les données, leur provenance et leurs usages, et serviront de base au développement de services d'interrogation et de consommation de données

Références

Daskalaki, E., Flouris, G., Fundulaki, I., and Saveta, T. (2016).
Instance matching benchmarks in the era of linked data.
Web Semantics: Science, Services and Agents on the World Wide Web.

Euzenat, J. and Shvaiko, P. (2013).
Ontology matching.
Springer-Verlag, Heidelberg (DE), 2nd edition.


Remerciement
≡
Agradecimento
Merci
≡
Obrigado