

# Où sont les termes

Béatrice Markhoff, Arnaud Soulet



# Plan

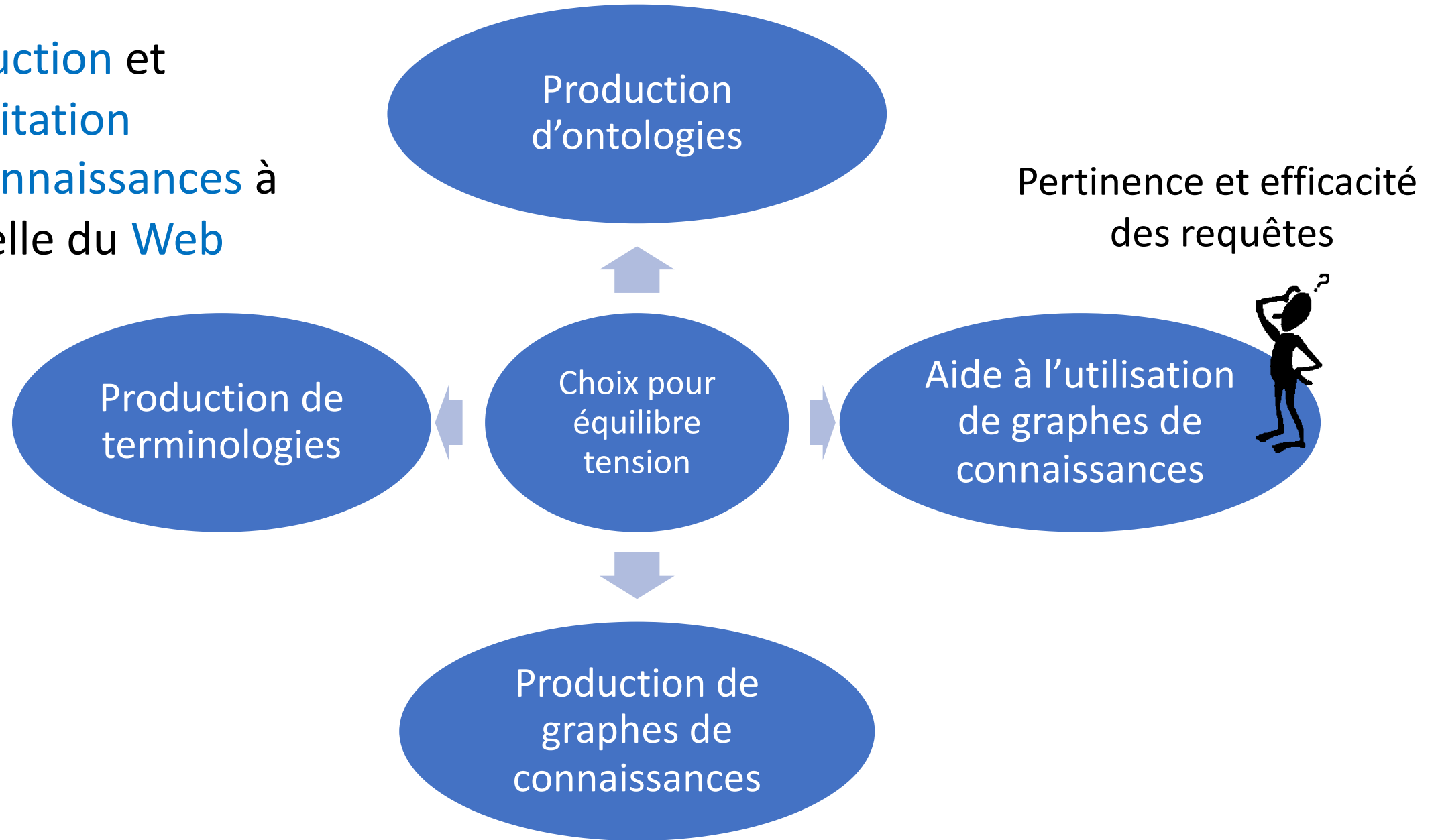
- Verrou
- Ontologies / Terminologies
- Dans le Web
- CIDOC CRM : exemple d'articulation ontologie/termes
- Conclusion

# Verrou

↑  
Système distribué

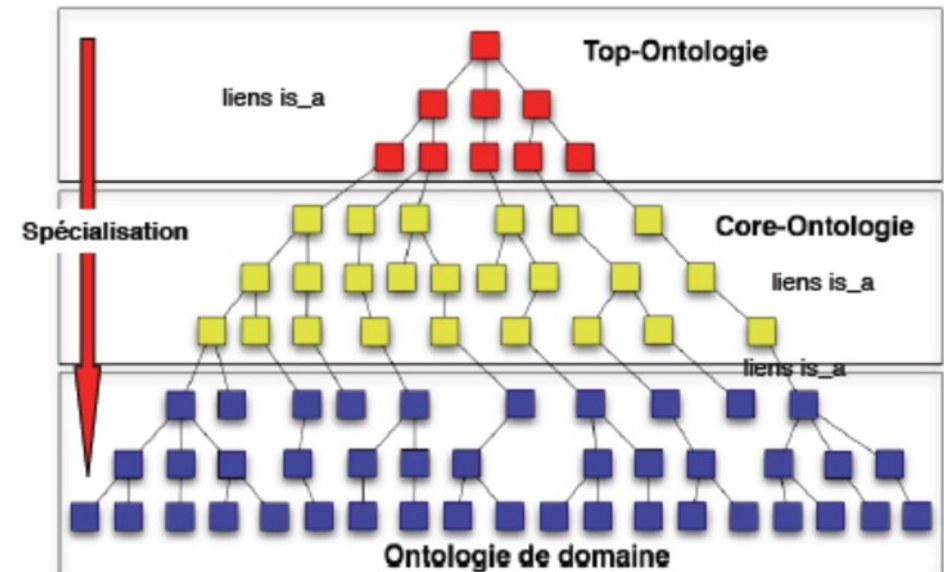
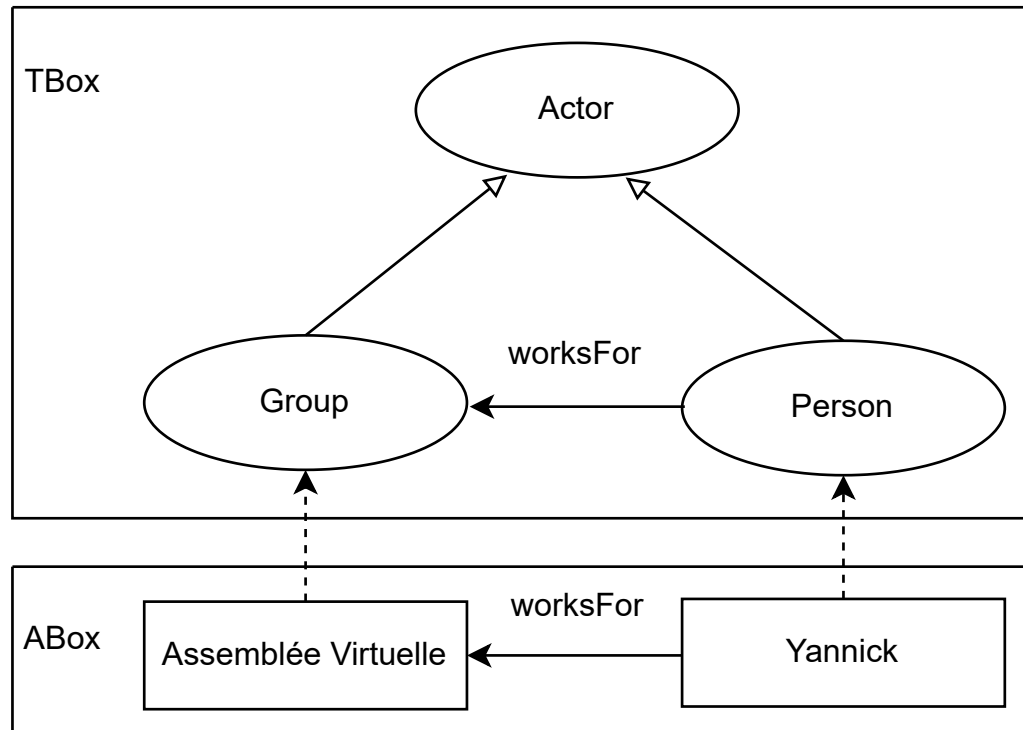
- Production et exploitation de connaissances à l'échelle du Web
- Tension entre deux nécessités
  - Interopérabilité sémantique : comment exprimer le sens partagé par les jeux de données (dans le patrimoine culturel : des objets créés par des humains)
  - Diversité et précision : comment exprimer les contenus précis propres à un domaine de connaissance plus restreint, forgés par ses spécialistes (sur des livres en bibliothèques, sur des peintures en musées, etc.)

Production et  
exploitation  
de connaissances à  
l'échelle du Web



# Ontologies

- An ontology is a **formal, explicit** specification of a **shared conceptualization** (Studer et al. 1998)



# Universaux et particuliers

## Métaphysique

Les universaux sont “des **termes généraux** qui semblent désigner ce qui est commun entre diverses choses”

Ils servent à désigner les **caractéristiques communes** des entités particulières

**Les ontologies sont pensées pour définir dans la partie TBox les universaux, pour décrire les entités particulières de la ABox.**

# Engagement ontologique (“a **shared** conceptualization”)

“Se comprendre les uns les autres”

**Consensus** sur  
comment nommer  
et sur  
ce que le nom veut dire

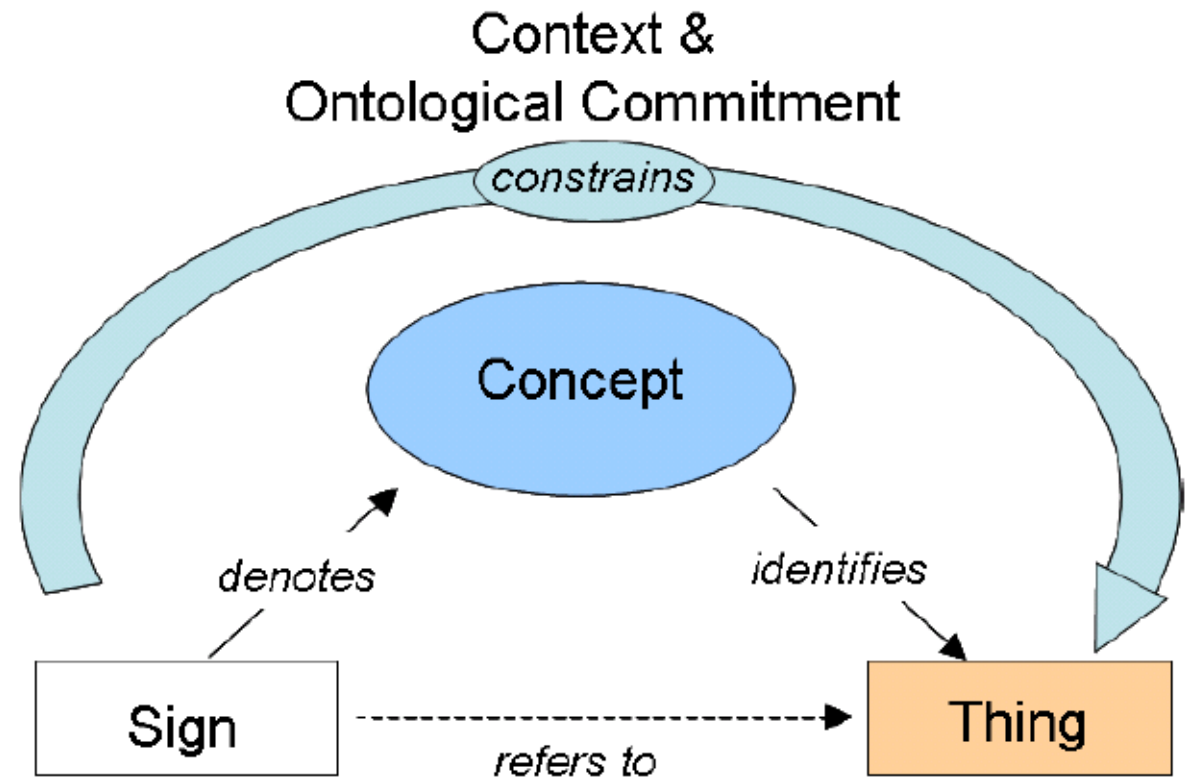


Schéma de N. Guarino et al., Handbook on Ontologies, 2009

# Terminologies

- Lorsqu'une communauté recherche un consensus sur les mots à utiliser pour décrire des éléments de son domaine, elle conçoit une terminologie.
- ISO 1087:2019 : "ensemble de désignations"
  - Désignation : représente un concept par un signe qui le représente (terme)
- ISO TC 37 LMF (ressources lexicales) et TMF (terminologies)
  - Ontologie Ontolex-lemon et extension pour terminologies
- Onto-terminologies (C. Roche et al, série de conférences TOTH)



# Plusieurs domaines en informatique

- Ingénierie des connaissances (ontologies, onto-terminologies)
- Traitement Automatique des Langues (LMF – TMF et Ontolex-Lemon)
- Bases de données : “nomenclatures” ou “référentiels” à utiliser pour renseigner les valeurs des attributs
- Bases documentaires : “systèmes d’organisation des connaissances”

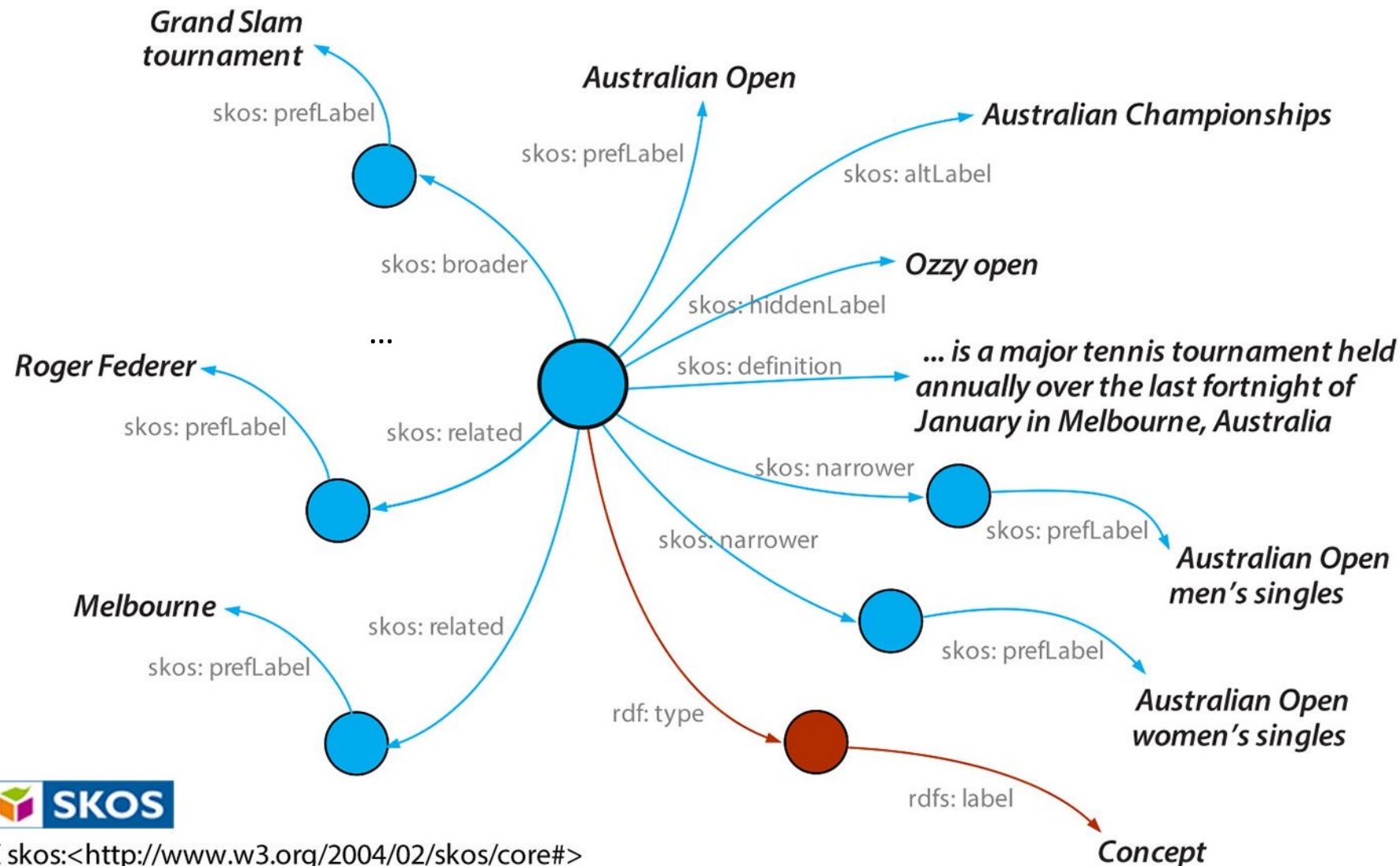
# *Dans le Web* : nombreuses terminologies sous forme d'ontologies

- Les “vocabulaires” standards du W3C
  - SKOS
  - PROV-O
  - ...
- Les “vocabulaires” populaires du Linked Open Data
  - Dublin Core
  - FOAF
  - Schema.org
  - ...

# Thésaurus

- ISO 25964-1 : vocabulaire structuré où les **concepts** sont représentés par des **termes**, avec des relations entre concepts (hiérarchie) et des groupements de termes synonymes
- Objectif : **indexer** et **retrouver** des documents en fonction de leur contenu
- élargir ou restreindre une requête : le concept A est **plus général** que le concept B si dans toute recherche de A les articles traitant de B devraient être retournés
  - Relations de **subsumption**, de **partition**, et parfois aussi d'**instanciation**

# Dans le Web : SKOS (et SKOS XL)



# *Dans le Web* : données et métadonnées

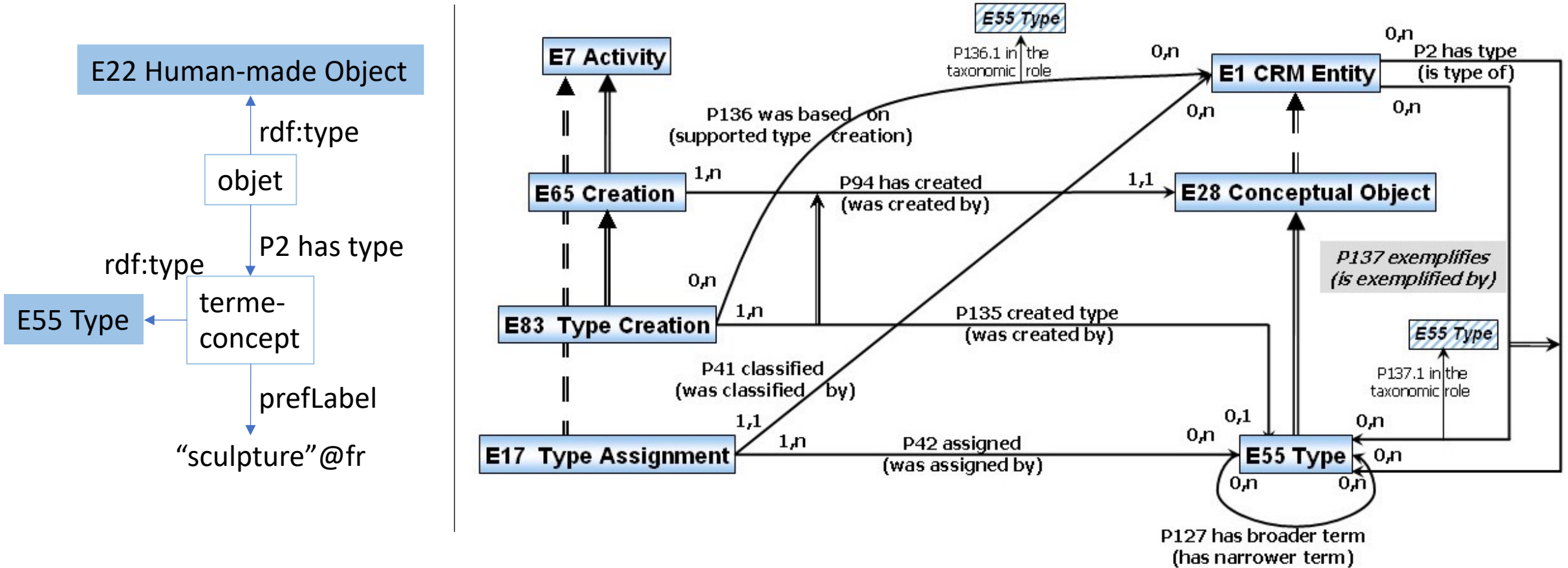
- Métadonnées
  - Éléments de TBox (classes et propriétés, définies en RDFS ou OWL)
- Données (éléments de ABox)
  - Instances de [skos:Concept](#)
  - Instances de [skosxl:Label](#)
  - Instances de [ontolex:Word](#)
  - Instance de [schema:DefinedTerm](#)
  - Instances de [crm:E55\\_Type](#)

# (Noms de) classes, (noms de) propriétés, (noms de) concept-termes : universaux

Profil, ou résumé du contenu d'un graphe de connaissance :

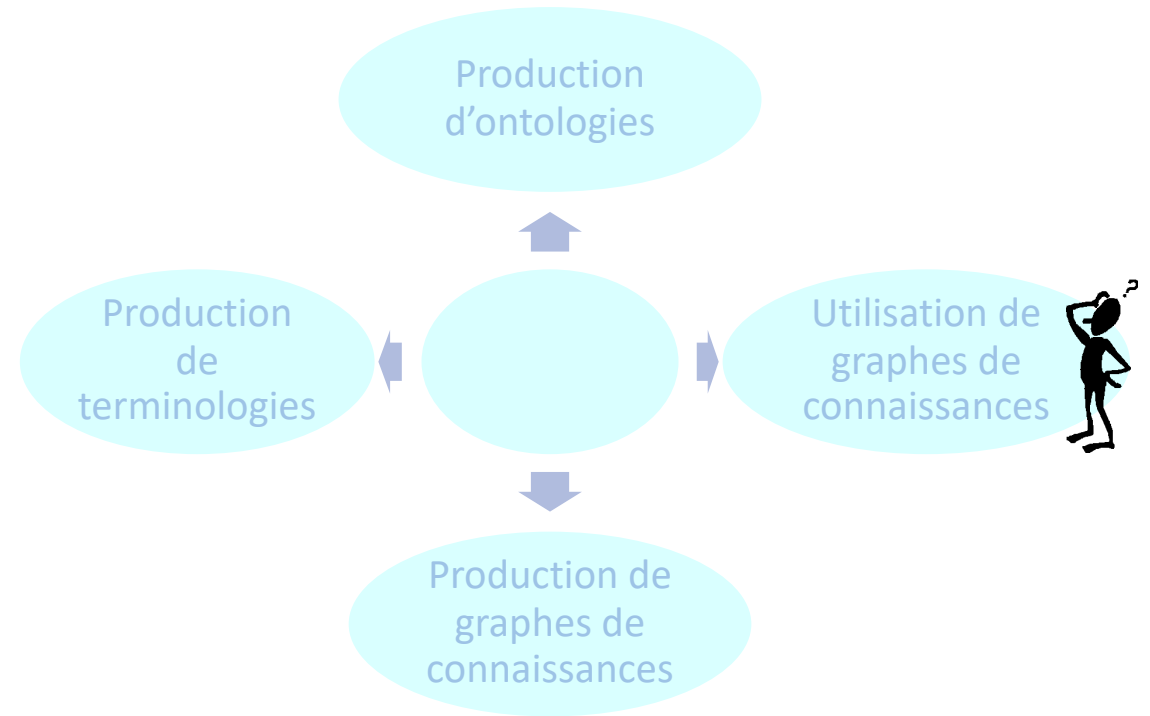
- Les universaux servent à caractériser les particuliers : en ce sens ils devraient se trouver en objet de propriétés
- Les universaux servant à caractériser les particuliers, ce sont des noeuds qui doivent avoir un grand nombre de propriétés entrantes
- Les universaux n'ont pas eux-mêmes à être décrits par des particuliers, ils sont décrits par des littéraux ou des universaux
- Tentatives de traduire cela en requête SPARQL => article à Ingénierie des Connaissances, PFIA 2022.
- Conclusion : il faudrait plus de bonnes pratiques pour guider l'utilisation de terminologies dans les graphes de connaissances

# La proposition de CIDOC CRM : E55 Type



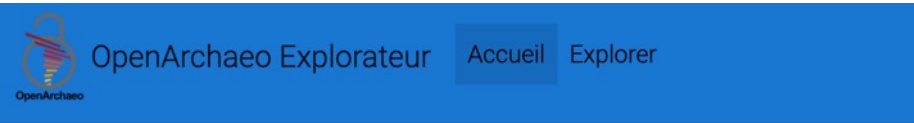
# Conclusion - perspectives

- Cette question “où sont les termes” est riche de perspectives
- Exemples de problématiques liées :
  - Conceptions d'ontologies
  - **Workflows** de production de graphes de connaissances
  - Spécification générique d'une interface d'interrogation **dans les termes de spécialité**
  - Génération automatique de résumés **montrant les termes**
  - Support à la détection et l'exploitation “d'entités d'intérêt” en TAL





# Publication et interrogation : OpenArcheao



## Plateforme web sémantique

MASA (Mémoire des Archéologues et des Sites Archéologiques), cor Web sémantique, et le [Laboratoire d'Informatique Fondamentale et /](#) Exploitant les triplestores archéologiques appariés avec le [CIDOC](#) de

### Interface Utilisateur : l'Explorateur

[ACCÉDER À L'EXPLORATEUR](#)

L'explorateur, résolument convivial, est basé sur un système d'icônes et de concepts bien maîtrisés par les archéologues (Unité stratigraphique, Mobilier, Fait, Site...). Pour produire une requête SPARQL à partir du langage naturel, OpenArcheao a fait l'objet du développement du composant [SPARNATURAL](#), modulable et réutilisable. Ce système permet ainsi de générer une requête de manière intuitive, sans connaissance du langage SPARQL, mais en ayant une connaissance des principaux concepts de l'archéologie.

The interface shows a query builder with two rows of concepts connected by 'Et' (And) and 'trouvé à' (found at) relationships. The first row contains 'Sepulture' (Burial) and 'Site' (Site). The second row contains 'Site' (Site) and 'Acteur' (Actor) 'Elisabeth Lorans'. A 'LANCER LA RECHERCHE' (Launch Search) button is at the bottom right, along with a 'Tableau' (Table) dropdown menu and the text 'affichage des résultats :'. Each concept in the query builder has a small 'x' icon to remove it.

✓ Requête terminée ! - [Afficher/masquer la requête SPARQL](#)

[Response](#) **[Table](#)** [Pivot Table](#) [Geo](#) [Google Chart](#) [Timeline](#) [Download](#) [Code](#)

Showing 1 to 50 of 722 entries

Search:  Show **50** entries

	this	thisLabel
1	<a href="http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000001">http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000001</a>	Sépulture AJ000001
2	<a href="http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000002">http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000002</a>	Sépulture AJ000002

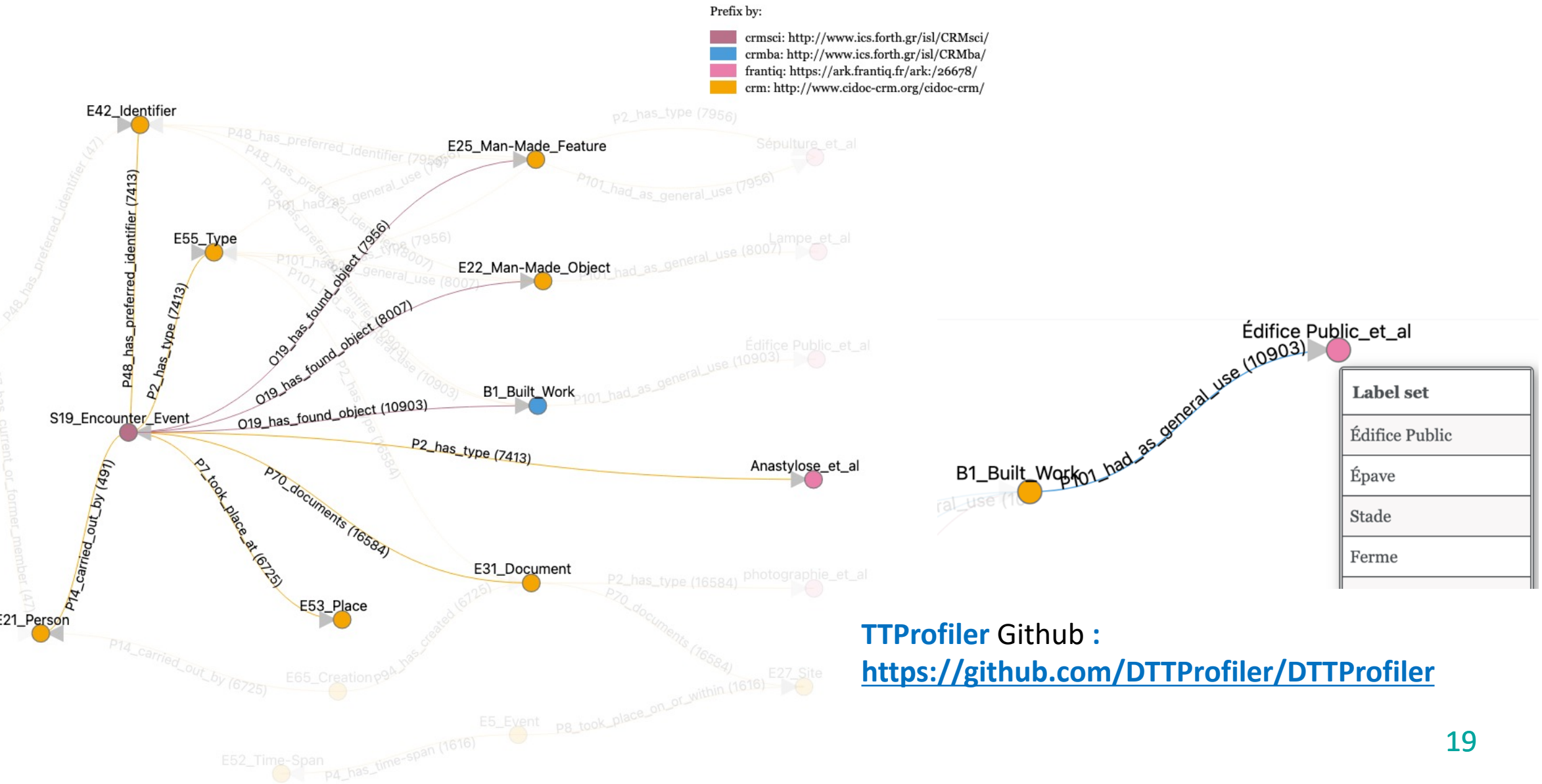
# compréhension du contenu

- Difficile pour un développeur de savoir ce qui peut être interrogé dans le graphe

```
SELECT DISTINCT ?person ?personLabel
FROM NAMED <http://openarchaeo.huma-num.fr/federation/sources/arsol>
WHERE {
    ?person a crm:E21_Person .
    ?person skos:prefLabel ?personLabel .
    ?person crm:P14i_performed/crm:P8_took_place_on_or_within ?site .
    ?site crm:P8i_witnessed/crmsci:019_has_found_object ?sepulture .
    ?sepulture a crm:E25_Man-Made_Feature .
    ?sepulture crm:P2_has_type <https://ark.frantiq.fr/ark:/26678
/pcrt795b632nWw> .
}
```



Profils <https://kgsumviz.univ-tours.fr>



# Merci



# Terminologies dans le Web

- UMLS (plusieurs dizaines d'énormes terminologies)
- MESH (index de Medline et PubMed)
- SNOMED CT (centaines de milliers de concepts, env. Clinique)
- AGROVOC (38 000 concepts, 800 000 termes dans 40 langages)
- ESCO
- AAT du Getty
- PACTOLS
- ConceptNet : <https://conceptnet.io/>