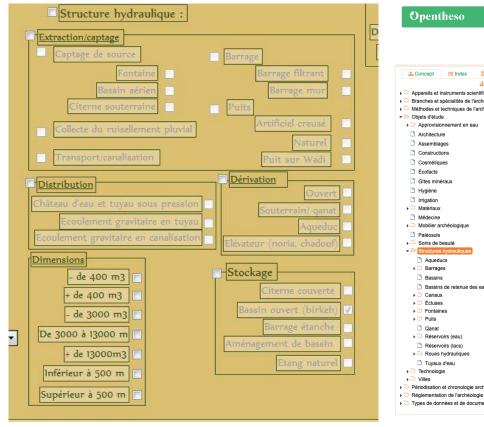
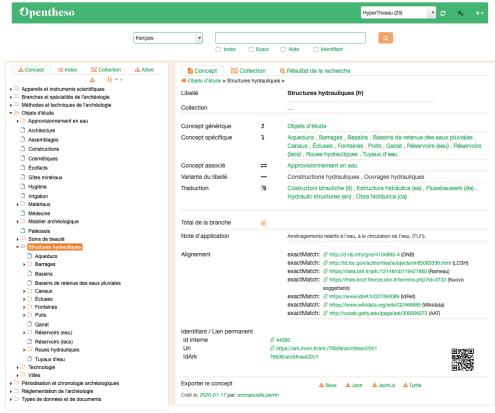
HisArc-RDF: données, vocabulaire, modélisation

Point de départ : une base de données ACCESS Point d'appui du modèle de données : un thésaurus

https://thesaurus.mom.fr/opentheso/?idc=44390&idt=25

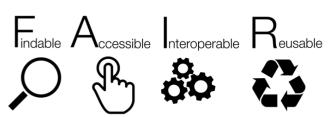




Science ouverte

Mettre en œuvre les principes « FAIR »





Infrastructure technique et épistémologie



F = métadonnées + ID unique et pérenne (dépôt sur un entrepôt public)



A = accessible à la machine et à l'Homme (protocoles, accès ouvert, pérennité)



I = formats non-propriétaires, vocabulaire contrôlé → requêtes normalisées (structure des données et vocabulaire explicites)



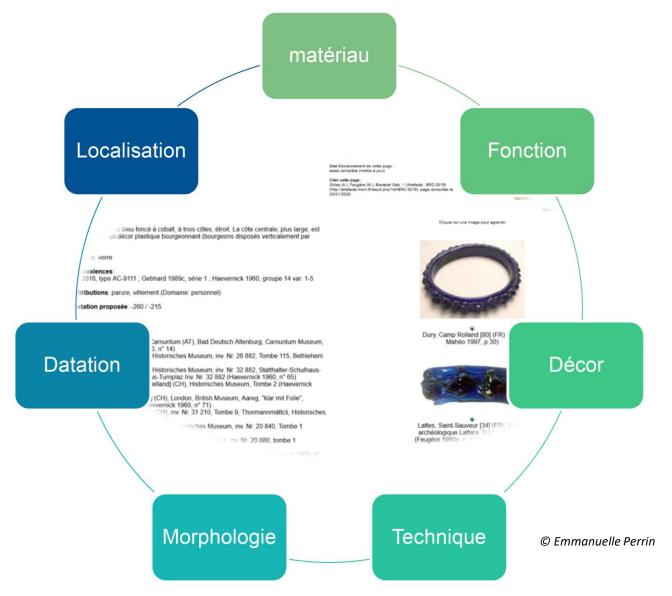
R = licences de réutilisation, description non ambiguë des données, de leurs conditions d'acquisition (épistémologie) et de leur modèle conceptuel,

- = données les plus « brutes » possibles (raw data) à chaque stade,
- = « calculabilité » → réutilisation effective » (≠ PDF, page Internet, etc.)

Ce qui donne son sens à toute la démarche, c'est le « R ».

État de l'art

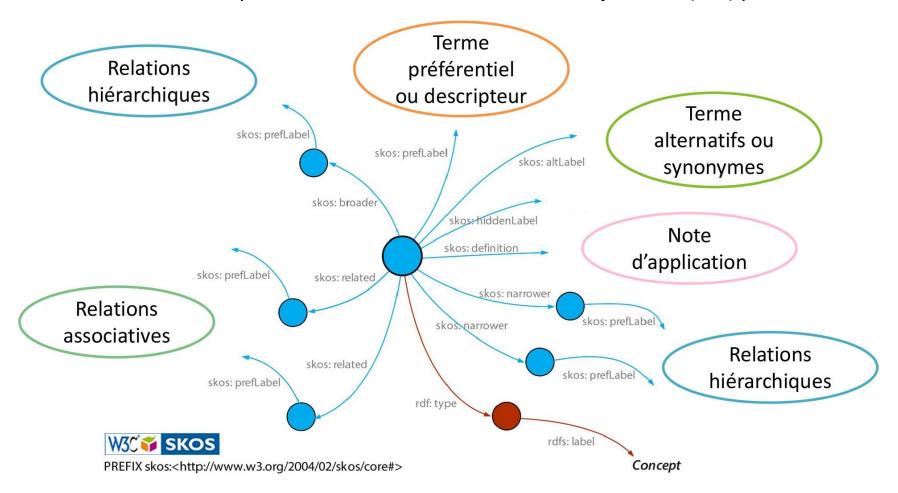
Description selon différents « points de vue » scientifiques)



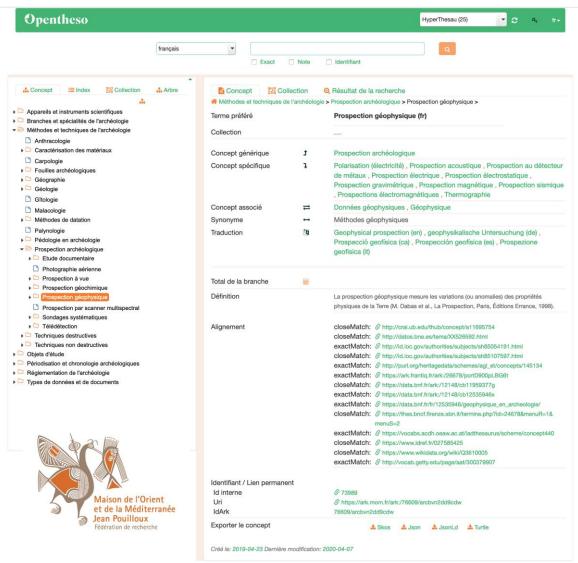
État de l'art

L'outil thésaurus (norme ISO 25964)

Un vocabulaire de **concepts** (*labels*) contrôlés par **un formalisme logique** et des **définitions** précises et référencées, une **adresse pérenne** (Ark) par label

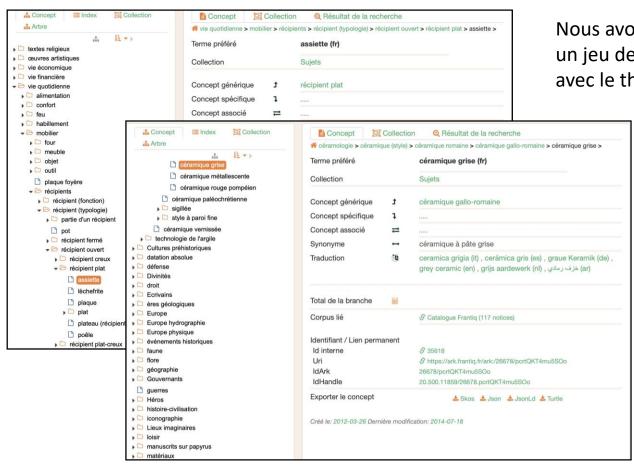


Un thésaurus « métier » pour la recherche archéologique avec Opentheso



L'archéologie, combien de thésaurus?

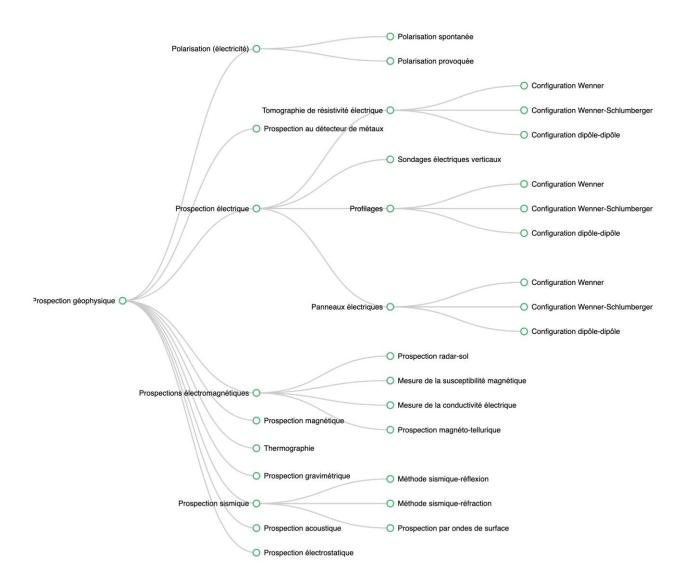
En France, nos collègues documentalistes indexent les ouvrages spécialisés dans un Catalogue collectif en ligne avec le thésaurus PACTOLS (avec Opentheso).



Nous avons essayé de décrire un jeu de tessons de céramique avec le thésaurus PACTOLS.

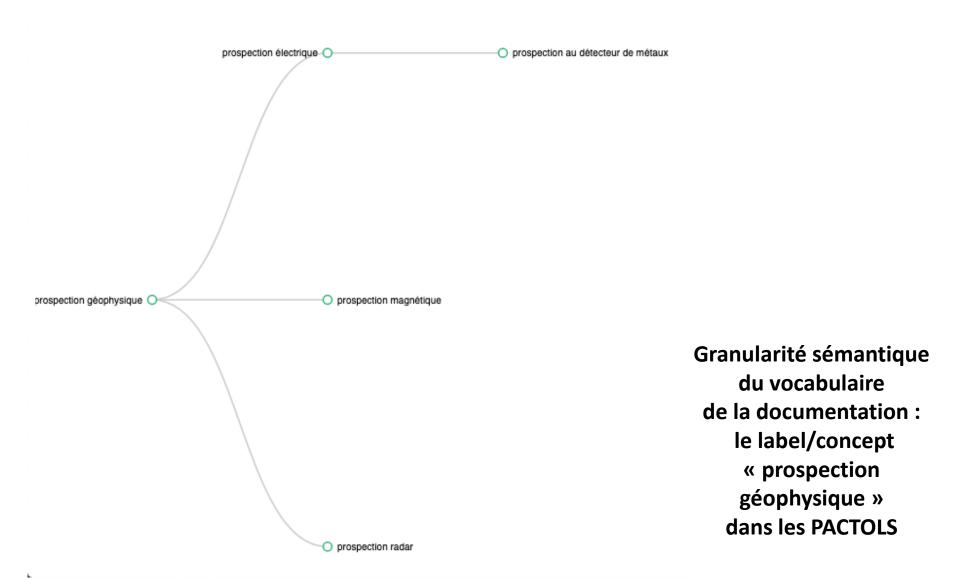
> La granularité de ce vocabulaire contrôlé n'est pas suffisante pour rendre compte de la diversité sémantique utilisée par les archéologues.

Un thésaurus orienté « terrain »



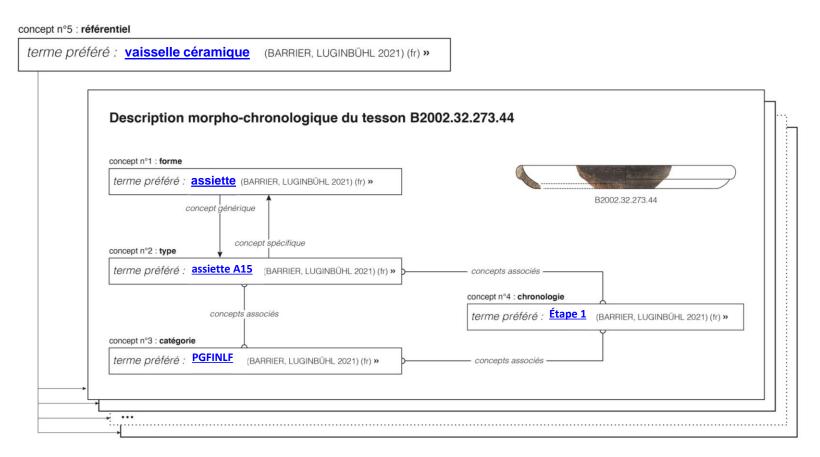
Richesse sémantique du vocabulaire de la recherche : le label/concept « prospection géophysique » dans HyperThésau

Un thésaurus orienté « doumentation »



Bibracte Ville Ouverte

Décrire une typologie de céramique gallo-romaine



Référentiel vaisselle céramique (BARRIER, LUGINBÜHL 2021) (fr)

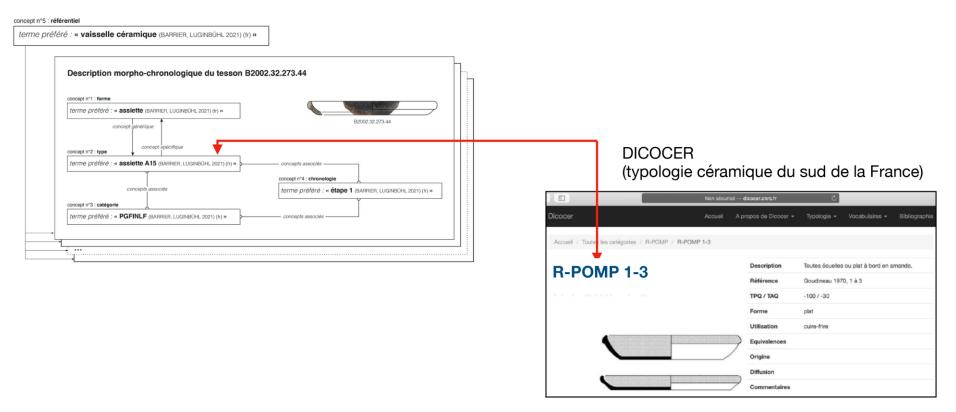
Il définit et date 53 catégories, 25 formes et 423 types de céramique répertoriés à ce jour pour la période d'occupation de l'*oppidum* de Bibracte.

Bibracte Ville Ouverte

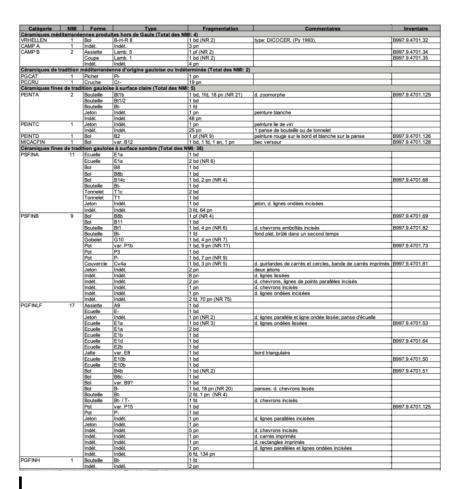
Application à *deux* typologies de céramique gallo-romaine

Des formes semblables en contexte similaire des références différentes dans des typologies ± concurrentes (voir lacovella et al. 2006)

« vaisselle céramique (BARRIER, LUGINBÜHL 2021) (fr) »



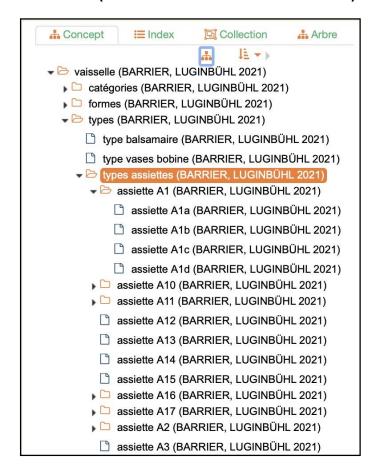
Une chaîne opératoire « humain-machine-humain »



Jeu de données dynamiques (feuille de calcul téléchargeable depuis un entrepôt Nakala), exprimées avec les valeurs d'un vocabulaire contrôlé, partagé et pérenne

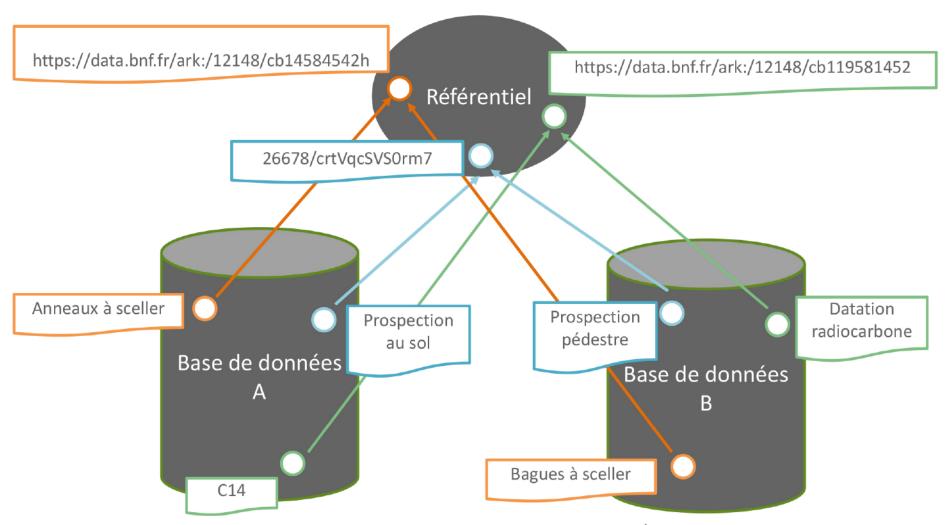


(voir Wilkinson et al. 2016)



État de l'art

Vers un vocabulaire « pivot » pour la recherche archéologique

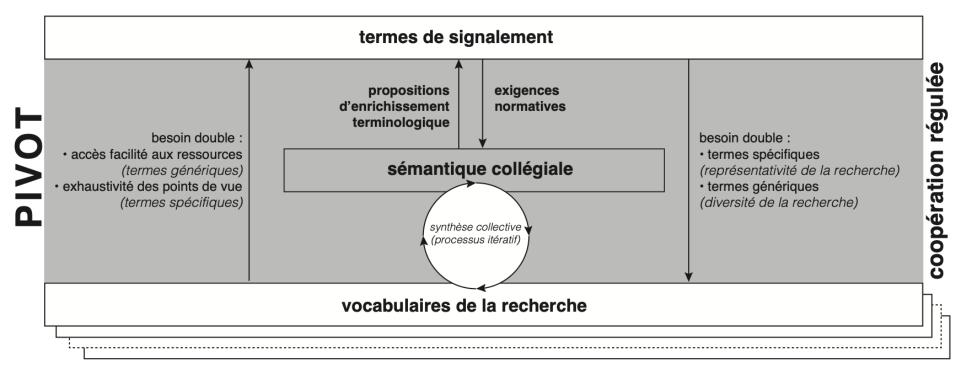


D'après E. Bermès, Le Web sémantique en bibliothèque, Éditions du cercle de la librairie, 2013.

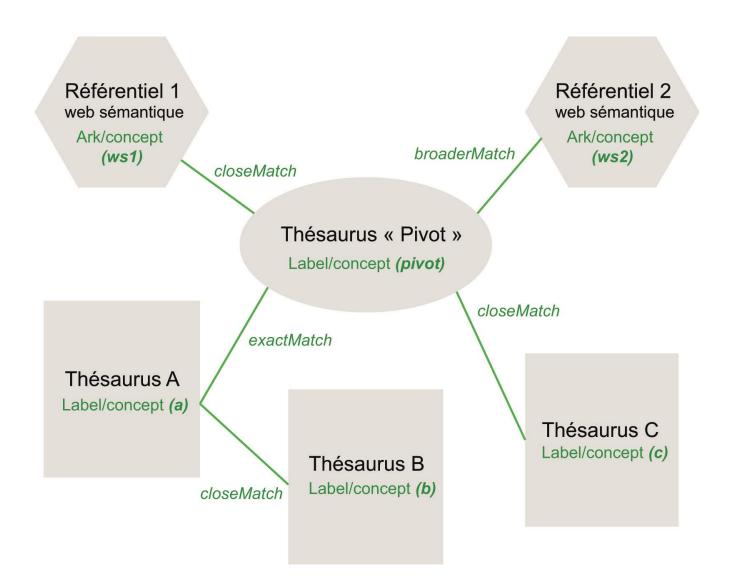
Faire converger les différents vocabulaires de la recherche

La proposition d'HyperThésau pour réconcilier des besoins sémantiques spécifiques : créer un thésaurus intermédiaire « pivot »

Une sémantique collégiale **produite par des archéologues** dans une démarche **« facilitée » par les « ingénieurs de la terminologie »** (Bachimont 2000)



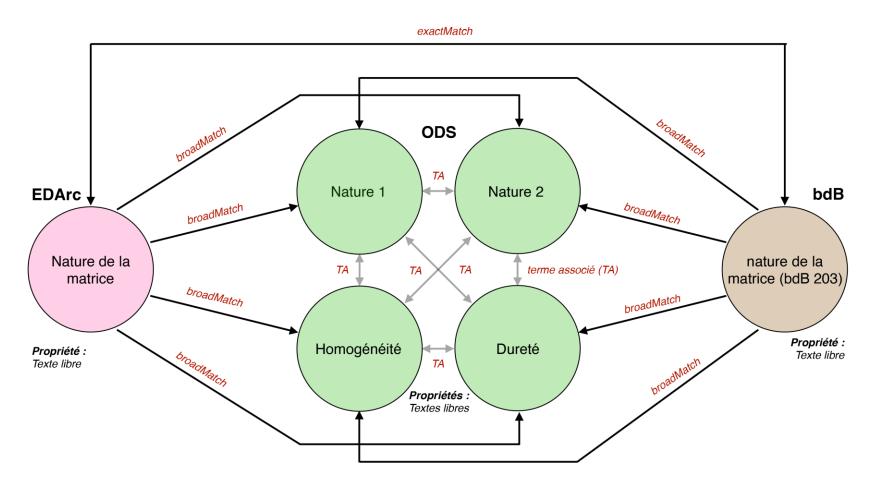
Convergence et divergences : « l'union est un combat »



En réalité, peu d'équivalences exactes entre les vocabulaires des différentes équipes de recherche et les vocabulaires génériques du web sémantique

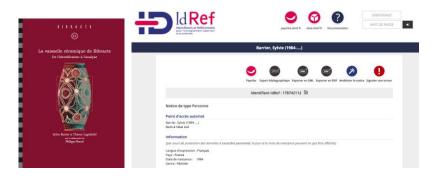
« Aligner » les systèmes d'information archéologiques

Les écarts de vocabulaires expriment des différences de relations logiques entre les tables des SIA (« modèles » de données)



Synthèse

Deux catégories de vocabulaire : documentaire et scientifique



Les données documentaires sont intangibles.

Elles font référence.

Elles sont parfaitement *stables* dans le temps.





Les données scientifiques sont hétérogènes et évolutives. Elles ne valent que dans le cadre précisé par leur publication (Desfarges, Helly 1991) et sont *instables* dans le temps.

