



HAL
open science

Consolider et enrichir ses données avec les alignements

Emmanuelle Perrin

► **To cite this version:**

| Emmanuelle Perrin. Consolider et enrichir ses données avec les alignements. 2020. hal-03712417

HAL Id: hal-03712417

<https://hal.science/hal-03712417v1>

Submitted on 3 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Opentheso

ALIGNEMENTS

Consolider et enrichir ses données avec les alignements

PAR EMMANUELLE PERRIN · PUBLIÉ 28/09/2020 · MIS À JOUR 11/05/2022

[english version in progress]

Interopérabilité, ouverture et réutilisation des données

- [Ressources pour interrelier des vocabulaires](#)
- [Les types d'équivalence du modèle SKOS](#)

Aligner automatiquement des concepts dans Opentheso

- [L'alignement automatique](#)
- [Sélection des concepts à aligner](#)
- [Sélection des référentiels externes auxquels s'interconnecter](#)
- [Ajouter un thésaurus construit sous Opentheso](#)
- [Enrichir ses données avec Wikidata](#)

Aligner manuellement des concepts dans Opentheso

[Enrichir ses données avec Wikidata](#)
[Aligner manuellement des concepts dans Opentheso](#)

[Modifier un alignement](#)

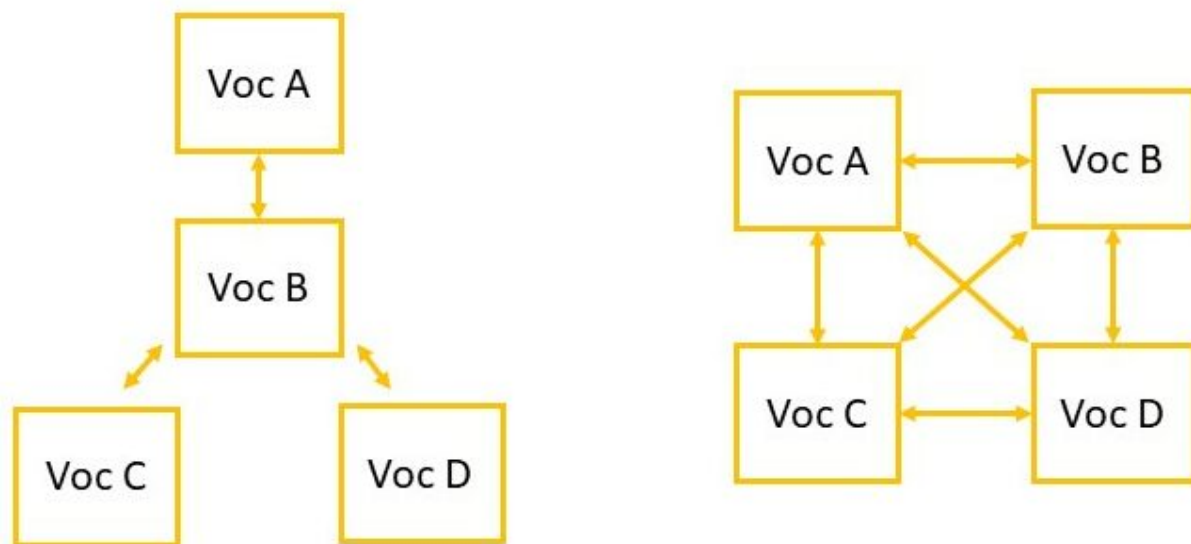
[Les coordonnées géographiques](#)

- [Ajouter ou modifier des coordonnées géographiques](#)
- [Supprimer des coordonnées géographiques](#)

[Bibliographie](#)

Interopérabilité, ouverture et réutilisation des donnée

Les propriétés d'alignement proposées dans le modèle SKOS permettent d'exprimer des liens d'équivalences entre des concepts provenant de différents systèmes d'organisation de connaissances. Ces liens forment des passerelles pour relier entre eux des vocabulaires et ainsi contribuer à l'interopérabilité, à l'ouverture, à l'enrichissement et à la réutilisation des jeux de données interconnectés.



Les modèles d'interopérabilité : autour d'un pivot (à gauche) ou par paire (à droite) (d'après Antoine Isaac, "Les Référentiels : typologie et interopérabilité", Le document numérique à l'heure du web de données, ADBS, 2012, p. 94, <<https://hal.inria.fr/hal-00740282v1>>).

Il existe plusieurs modèles d'interopérabilité. Celui du pivot est par exemple représenté par le Fichier d'autorité international virtuel (VIAF), qui fait converger en un point central les notices d'autorités de nombreuses bibliothèques. La structure par paire propose un alignement entre référentiels "voisins". On trouve souvent de telles équivalences entre les notices des bibliothèques :

Voir aussi**À la BnF (1)**

Notice correspondante dans Catalogue général

Sur le Web (4)

Notice correspondante dans Bibliothèque nationale d'Espagne

Notice correspondante dans Fichier d'autorité intégré de la Bibliothèque nationale allemande

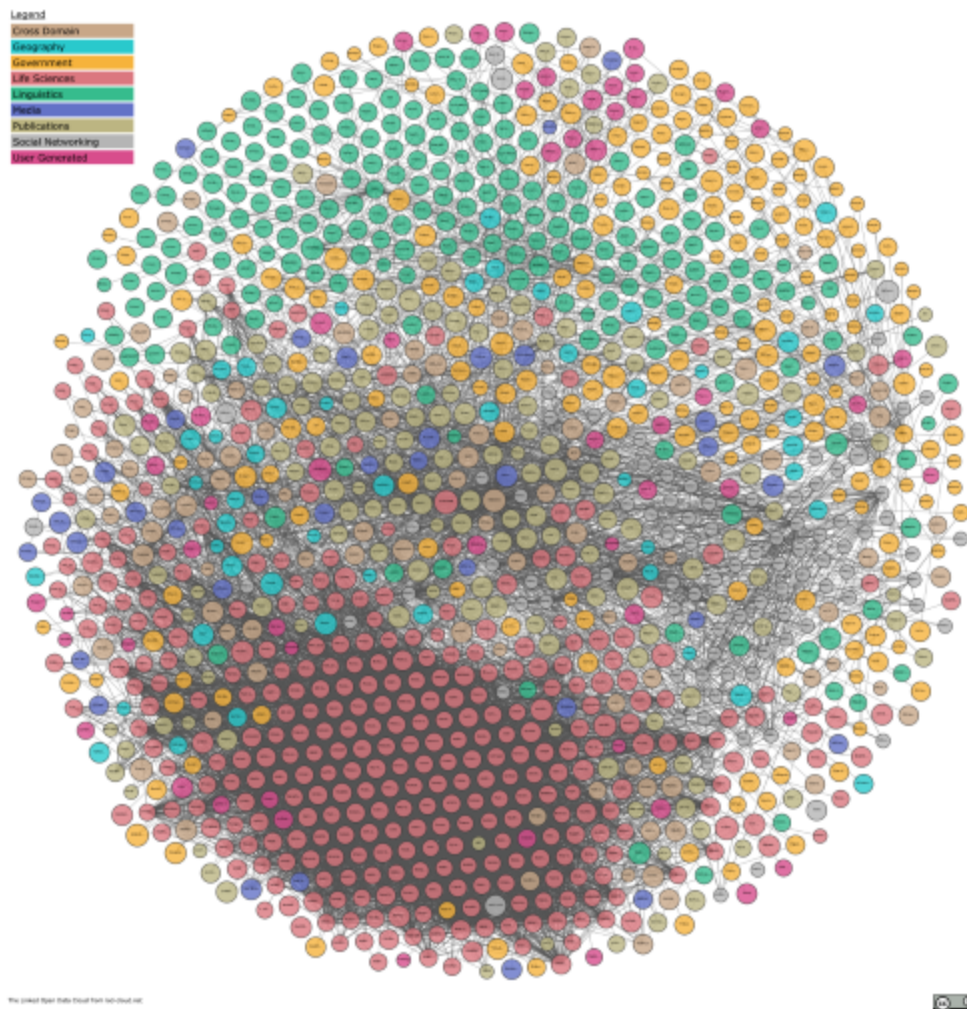
Notice correspondante dans Bibliothèque du Congrès

Notice correspondante dans Le Nuovo Soggettario

Alignements proposés dans la notice de data.bnf.fr sur les thésaurus (sources : <https://data.bnf.fr/fr/11935918/thésaurus/>).

Dans cet exemple, le concept "Thésaurus" de data.bnf.fr a pour équivalent celui de "Tesauros" à la bibliothèque nationale d'Espagne, celui de "Thesaurus" à la Bibliothèque nationale allemande, celui de "Thesauri" à la Bibliothèque américaine du Congrès et celui de "Thesauri" à la Bibliothèque nationale centrale de Florence.

Se connecter au réseau que forment les données ouvertes liées permet donc d'intégrer un espace global d'informations.



Le réseau des données ouvertes liées ou Linked Open Data Cloud : un espace global d'information (source : lod-cloud.net).

■ Ressources pour interrelier des vocabulaires

Référentiels	Producteurs	Adresse	Couverture
Agrovoc	FAO	http://aims.fao.org/fr/agrovoc	alimentation, agriculture, pêche, forêt, environnement
ArchiRèS	Réseau documentaire des écoles d'architecture	http://bartoc-skosmos.unibas.ch/ThesArchires/en/	Architecture
Art & Architecture Thesaurus	Getty	http://www.getty.edu/research/tools/vocabularies/aat/index.html	Art et architecture
data.bnf.fr	Bibliothèque nationale de France	https://data.bnf.fr	Notices d'autorité de la BnF et vedettes matière

Référentiels	Producteurs	Adresse	Couverture
			du vocabulaire Rameau
Datos.bne.es	Biblioteca Nacional de Espana	http://datos.bne.es/inicio.html	Catalogue bibliographique et autorités de la Bibliothèque nationale d'Espagne
Datu	EHESS	https://datu.ehess.fr/fr/	Vocabulaires institutionnels et scientifiques de l'EHESS
Deutsche National Bibliothek	Deutsche National Bibliothek	https://www.dnb.de/DE/Home/home_node.html	Catalogue bibliographique et autorités de la Bibliothèque nationale d'Allemagne
GBIF (Global Biodiversity Information Facility)	GBIF	https://www.gbif.org/	Système mondial d'information sur la biodiversité
GEMET	Agence européenne pour l'environnement	https://www.eionet.europa.eu/gemet/fr/theme/2/concepts/	Environnement
GeoNames		https://www.geonames.org/	Base de données géographiques collaborative
Getty Thesaurus of Geographic Names	Getty	http://www.getty.edu/research/tools/vocabularies/tgn/index.html	Noms géographiques
Idref	Sudoc	https://www.idref.fr	Identifiants et référentiels de l'enseignement supérieur
ISNI (International Standard Name Identifier)	ISNI-IA	https://isni.org/	Code international normalisé des noms de personnes et d'organisme

Référentiels	Producteurs	Adresse	Couverture
Library of Congress Subject Headings	Library of Congress	http://id.loc.gov/authorities/subjects.html	Vedettes matières de la Bibliothèque du Congrès
Loterre	INIST	https://www.loterre.fr/	Plateforme d'exposition et de partage de terminologie scientifique
Nuovo soggettario Thesaurus	Biblioteca Nazionale Centrale di Firenze	https://thes.bncf.firenze.sbn.it/ricerca.php	Notices d'autorité de la Bibliothèque nationale centrale de Florence
Pactols	GDS Frantiq	https://pactols.frantiq.fr/opentheso/	Archéologie
PeriodO	Adam Rabinowitz (University of Texas at Austin) and Ryan Shaw (University of North Carolina at Chapel Hill)	https://test.perio.do/#p/Canonical/	Périodes historiques
Thesaurus de la Universitat de Barcelona	Universitat de Barcelona	http://bartoc-skosmos.unibas.ch/thub/en/?clang=ca	Thésaurus pluridisciplinaire utilisé par l'université de Barcelone pour indexer ses ressources
Thésaurus de l'Unesco	Unesco	http://vocabularies.unesco.org/browser/thesaurus/fr/	éducation, culture, sciences naturelles, sciences sociales et humaines, communication, information
VIAF (Virtual International Authority File)	Bibliothèques nationales	https://viaf.org/	Fichier d'autorité international
Vocabulaires du ministère de la Culture et de la	Ministère de la Culture et de la Communication	http://data.culture.fr/thesaurus/	Archives, arts, architecture, objets mobiliers, photographie,

Référentiels	Producteurs	Adresse	Couverture
Communication			patrimoine
Wikidata		https://www.wikidata.org	Base de données de Wikipédia

SKOS Datasets : <https://www.w3.org/2001/sw/wiki/SKOS/Datasets>

Linked Open Vocabularies : <https://lov.linkeddata.es/dataset/lov>

sameAs : <http://sameas.org/>

■ Les types d'équivalence du modèle SKOS

Le modèle SKOS propose cinq types d'équivalence :

- **skos:exactMatch** : équivalence exacte entre deux concepts.
- **skos:closeMatch** : équivalence approximative ou partielle entre deux concepts.
- **skos:broadMatch** : lien d'alignement générique entre deux concepts.
- **<A> skos:broadMatch ** : B est plus générique que A.
- **skos:narrowMatch** : lien d'alignement spécifique entre deux concepts.
- **<A> skos:narrowMatch ** : B est plus spécifique que A.
- **skos:relatedMatch** : lien d'alignement associatif entre deux concepts.

Aligner automatiquement des concepts avec Opentheso

■ L'alignement automatique

1) Pour ajouter un alignement automatique, cliquer sur

2) Sélectionner la source d'alignement, puis cliquer sur

3) Sélectionner l'alignement avec

4) Contrôler les informations (possibilité de décocher les informations non désirées pour l'alignement) puis, valider.

Ajouter un alignement automatique

■ Sélection des référentiels externes auxquels s'interconnecter

1) Pour gérer les sources de l'alignement, cliquer sur « Gérer les sources ».

2) Dans la fenêtre, sélectionner un ou plusieurs thésaurus, puis valider.

Wikidata_sparql et Wikidata_rest sont différents uniquement par leur méthode de requête (à choisir en fonction de votre serveur). La source « bnf_instrumentMusique ne fonctionne pas.

Gérer les sources d'alignement automatique

■ Ajouter un thésaurus construit sous Opentheso

1) Pour ajouter une source d'alignement, cliquer sur « Ajouter une source ».

2) Il n'est possible d'ajouter que des instances d'Opentheso. L'URL du serveur doit être de type : <https://monserveur.fr/opentheso>

3) Dans la fenêtre, indiquer le nom, l'URL et l'identifiant du thésaurus, puis valider en cliquant sur « Ajouter ».

Ajouter une source d'alignement automatique

■ Enrichir ses données avec Wikidata

En cliquant sur le lien, il est possible de consulter la page.

Sélectionner l'alignement correspondant au concept

Alignement automatique avec Wikidata – sélectionner l'alignement

1) Sélectionner le type d'alignement.

ID lang	Nom	Nom complet
af	azak	بنق
de	Nahrungsmittel	Haselnuss
el		φουτούκι
en	foodstuffs	hazelnut
he		תבואה
it	cibo	nocciola
nl	voedingsmiddelen	hazelnoot
ru		лесной орех
es	alimentos	avellana

2) Sélectionner les données (traductions, définitions et images)

3) Valider l'alignement.

Alignement automatique avec Wikidata – sélection du type d'alignement et des données

Les enrichissements de Wikidata ont été intégrés dans la notice du concept.

Traductions

Définition

Image

Alignement

Enrichissement d'un concept

Aligner manuellement des concepts dans Opentheso

1) Pour ajouter un alignement manuel, cliquer sur les paramètres.

2) Cliquer sur « Ajouter un alignement manuel ».

Ajouter un alignement manuel

Le concept de condiment à la Bibliothèque américaine du Congrès est identifié par un URI (Uniform Resource Identifier, identifiant uniforme de ressource).

3) Dans la fenêtre, sélectionner un type d'équivalence, et indiquer l'URI du concept à aligner et le système d'information dont ce concept est issu, puis valider.

Les alignements sont bâtis sur les identifiants (URI) des concepts.

■ Modifier un alignement

1) Pour modifier un alignement, cliquer sur les paramètres.

2) Cliquer sur « Modifier l'alignement ».

3) Dans la fenêtre, modifier l'URI ou l'équivalence puis valider.

Modifier un alignement manuel

■ Supprimer un alignement

1) Pour supprimer un alignement, cliquer sur les paramètres.

2) Cliquer sur « Supprimer un alignement ».

3) Dans la fenêtre, supprimer l'alignement en cliquant sur la corbeille.

Supprimer un alignement manuel

Les coordonnées géographiques

Les coordonnées géographiques sont exprimées en degrés décimaux.

■ Ajouter ou modifier des coordonnées géographiques

1) Pour ajouter ou modifier des coordonnées GPS, cliquer sur les paramètres.

2) Cliquer sur « Ajouter/modifier les coordonnées GPS ».

3) Dans la fenêtre, indiquer ou modifier les coordonnées puis valider.

Les coordonnées géographiques sont exprimées en degrés décimaux.

4) Le lieu est également géolocalisé sur une carte.

Coordonnées GPS

Nom du concept : Argentine

Latitude : 0.0

Longitude : 0.0

Fermer Valider

Coordonnées GPS

Latitude: -34.9964963 Longitude: -64.9672817

Ajouter des coordonnées géographiques

■ Supprimer des coordonnées géographiques

1) Pour supprimer des coordonnées GPS, cliquer sur les paramètres.

2) Cliquer sur « supprimer les coordonnées ».

3) Dans la fenêtre, supprimer les coordonnées en cliquant sur la corbeille.

Supprimer les coordonnées GPS

Nom du concept : Argentine

lat: -34.9964963 long: -64.9672817

Supprimer les coordonnées GPS

Supprimer des coordonnées géographiques

Citer cet article : « Consolider et enrichir ses données avec les alignements », par Emmanuelle Perrin. Publié sur *Opentheso* le 28/09/2020. Lien : <https://opentheso.hypotheses.org/52>.

Bibliographie

Angjeli, Anila et Isaac, Antoine, « Web sémantique et interopérabilité des vocabulaires : une expérimentation dans le domaine des enluminures », *World Library And Information Congress: 74th IFLA General Conference And Council, 2008* [[Lire en ligne](#)].

ATHENA et Linked Heritage, *Votre terminologie comme élément du web sémantique : Recommandations de conception et de management*, s.d. [[Lire en ligne](#)].

"Autorités, identifiants, entités. L'expansion des référentiels, *Arabesques*, 85, avril-juin 2017 [[Lire en ligne](#)].

Bermès, Emmanuelle, *Le Web sémantique en bibliothèque*, Bibliothèques, Paris, 2013 [[Lire en ligne](#)].

Isaac, Antoine, "Les Référentiels : typologie et interopérabilité", in INRIA, *Le Document numérique à l'heure du web de données*, ADBS, 2012, p85-104 [[Lire en ligne](#)].

Zeng, Marcia Lei, « Interoperability », *Isko Encyclopedia of Knowledge Organisation*, 46, 2, 2019, p. 122-146 [[Lire en ligne](#)].



Rechercher dans OpenEdition Search
Vous allez être redirigé vers OpenEdition Search

 Dans tout OpenEdition Dans Opentheso