



HAL
open science

Aligner pour mieux diffuser, l'expérience du thésaurus PACTOLS pour l'archéologie avec Wikidata

Christelle Molinié, Blandine Nouvel, Miled Rousset

► To cite this version:

Christelle Molinié, Blandine Nouvel, Miled Rousset. Aligner pour mieux diffuser, l'expérience du thésaurus PACTOLS pour l'archéologie avec Wikidata. Wiki, data et GLAM 2021. Webinaire, Wikimedia France; Etalab, Jun 2021, En ligne, France. halshs-03248607

HAL Id: halshs-03248607

<https://shs.hal.science/halshs-03248607>

Submitted on 3 Jun 2021

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.



Open licence - etalab

Aligner pour mieux diffuser

l'expérience du thésaurus PACTOLS pour l'archéologie
avec Wikidata

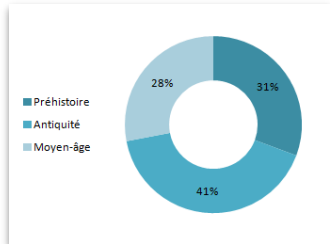
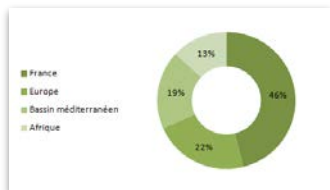


Christelle Molinié, documentaliste, Musée Saint-Raymond, Toulouse
Blandine Nouvel, Ingénieur CNRS, Centre Camille Jullian (Aix Marseille Univ, CNRS, Aix-en-Provence)
Miled Rousset, Ingénieur CNRS, Maison de l'Orient et de la Méditerranée (Université Lyon 2, CNRS, Lyon)

Wikimedia France et Etalab : webinaire « Wiki, data et GLAM 2021 », 3 juin 2021

PACTOLS, thésaurus pour l'archéologie

35 ans d'indexation des 40 bibliothèques de la Fédération et ressources sur l'Antiquité (Frantiqu, GDS du CNRS) dans le Catalogue collectif indexé.



Lexique évolutif

L'archéologie de la Préhistoire jusqu'au XXe s.
Les sciences de l'Antiquité
Europe et monde méditerranéen

60000 concepts multilingues

6 collections + 1 liste éponymes
Peuples
Anthroponymes
Chronologie
Oeuvres
Lieux
Sujets

Technique

<https://pactols.frantiqu.fr>
<https://github.com/frantiqu/PACTOLS>

Licence libre ODbL v1.0

ISO 25964 sur l'interopérabilité des thésaurus
SKOS
Import/exports - webservices REST

Identifiants pérennes à chaque concept :
Ark



projet Arkéo MOM

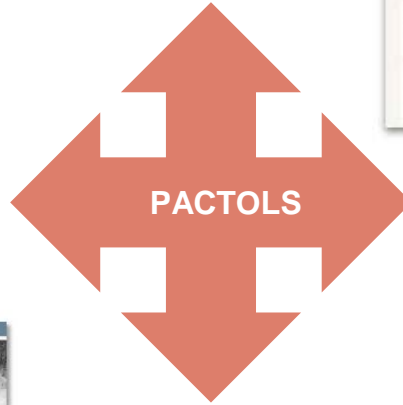
Pivot terminologique pour l'archéologie



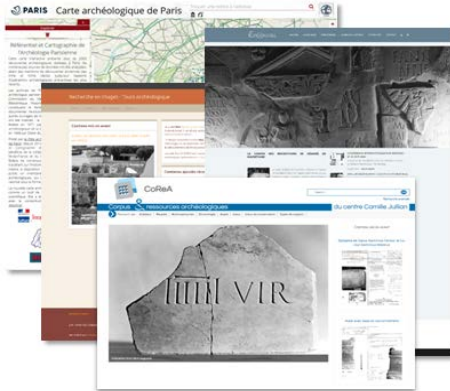
Enregistrement
de fouilles
archéologiques



Bibliothèques



Publications
scientifiques



Corpus de recherche
et patrimoine



Aligner, pourquoi et comment ?

Pour lier les données

en prenant appui sur les métadonnées thématiques, utilisées comme pivot entre plusieurs réservoirs.

2019 & 2020

IdRef

programmes CollEx [RefDivinités](#) avec Abes et BIS et [ArchéoRefAI](#) avec IFAO

fin 2017

Wikidata

[Projet PACTOLS/Wikidata](#), avec le support de Wikimedia France (Sylvain Boissel)

2017

Geonames

en prévision de l'abandon des Lieux, avec l'Inist

2016

Art and Architecture Thesaurus - Getty

Dans le cadre d'infrastructure européenne [ARIADNE](#), avec l'Inrap

Focus sur les alignements Wikidata

Plusieurs processus d'alignement automatisés ont été mis en oeuvre

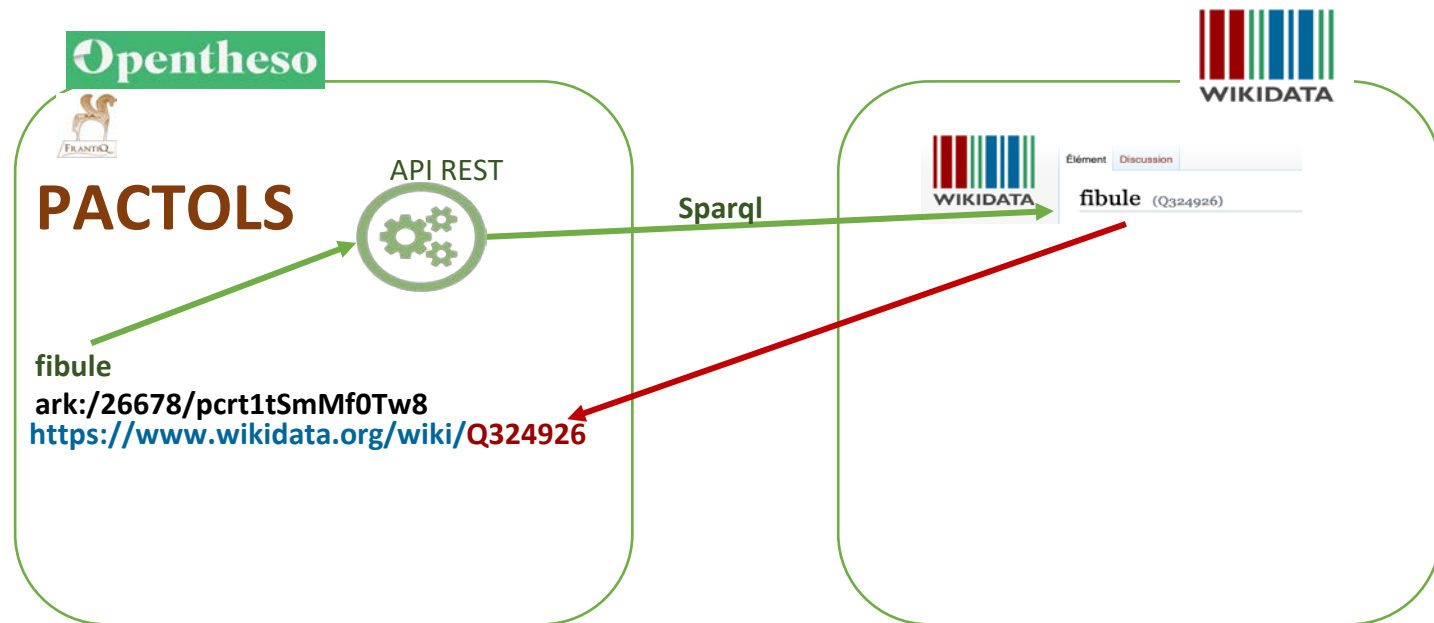
- s'appuient sur les identifiants pérennes des items des deux réservoirs
- combinent plusieurs solutions techniques offertes par Opentheso et Wikidata



Complétés de contrôles et d'une validation humaine ✓


Alignement depuis Opentheso

Phase 1



- Alignement via l'API d'Opentheso qui passe par le serveur Sparql de Wikidata
- Récupération de l'URI et de l'identifiant Wikidata
- L'URI est ajoutée uniquement côté PACTOLS


Opentheso



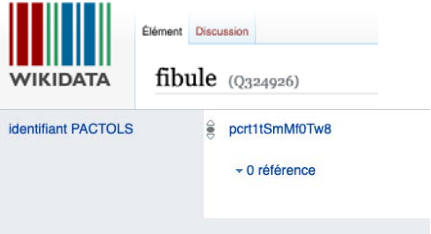
PACTOLS

fibule
ark:/26678/pcrt1tSmMf0Tw8
<https://www.wikidata.org/wiki/Q324926>

Export en CSV



WIKIDATA



Élément Discussion

fibule (Q324926)

Identifiant PACTOLS pcrt1tSmMf0Tw8

0 référence

QuickStatements

Id Ark	Id Wikidata	Label	note
pcrt1tSmMf0Tw8	Q324926	fibule	agrafe ou broche servant à attacher les vêtements
....

- Export des alignements d'Opentheso au format tabulé CSV
- Intégration du fichier dans Wikidata en passant par QuickStatements
- Les alignements sont alors ajoutés automatiquement puisque nous disposons déjà des identifiants uniques Wikidata

Enrichissement via Mix N'Match



1 .enrichissement dans Wikidata via Mix'n Match

https://mix-n-

Mix'n'match français Bienvenue, Miled.rousset Rechercher Rechercher

PACTOLS

Thésaurus de l'archéologie et des sciences de l'Antiquité PACTOLS du réseau FRANTIQ (Fédération et Ressources sur l'Antiquité)

Importé par Miled.rousset | Rafraîchir

Entrées

Associations validées	7883	<div style="width: 75.4%;"></div> 75.4%
Préassociées	263	<div style="width: 2.6%;"></div> 2.6%
Pas de Wikidata	0	<div style="width: 0%;"></div> 0%
Non applicable pour Wikidata	1124	<div style="width: 10.7%;"></div> 10.7%
Non associées	1175	<div style="width: 11.2%;"></div> 11.2%
Total	10445	

Export via une requête Sparql de Wikidata

Opentheso

2 .récupération des alignements de Wikidata en local

Import des alignements via Opentheso

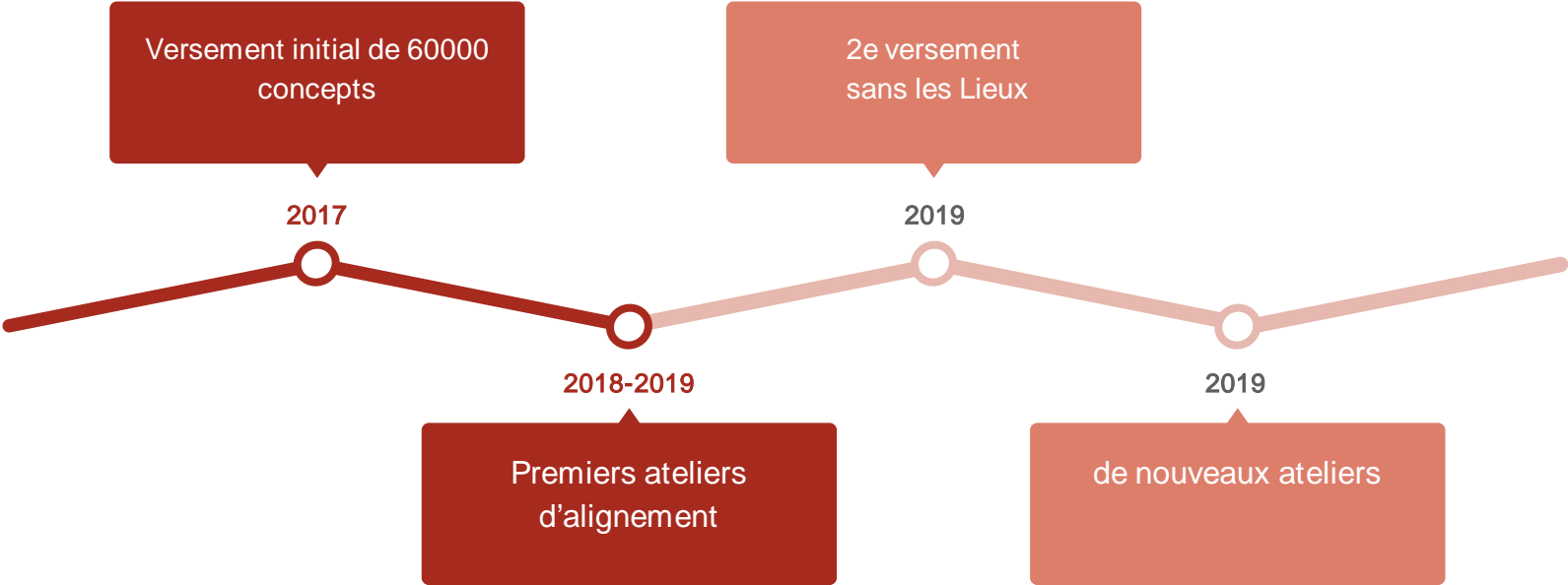
PACTOLS

fibule

ark:/26678/pcrt1tSmMf0Tw8

<https://www.wikidata.org/wiki/Q324926>

Empirisme itératif et pragmatisme



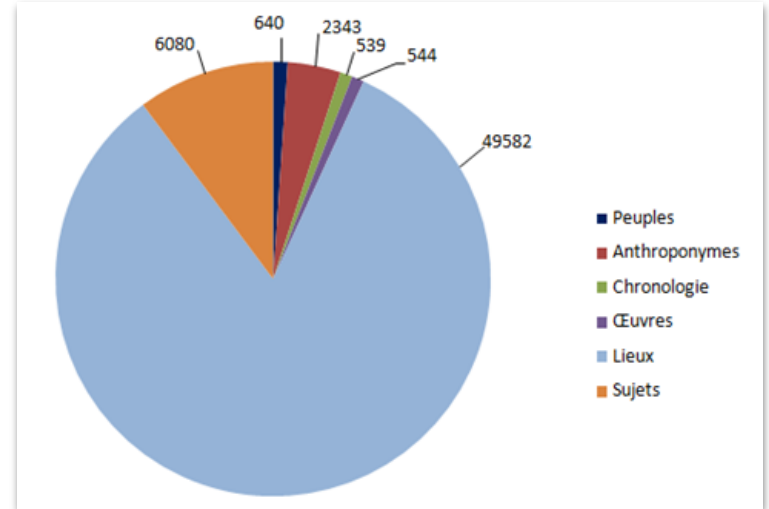
1er versement

60000 concepts

... mais les Lieux représentent 80%



- Difficulté de s'y retrouver !
- Le travail était déjà fait ou en cours pour les Lieux
- Perte de motivation face au volume d'alignements à faire...



2e versement

Pactols sans les Lieux, soit 11000 concepts

avec maintien des alignements déjà réalisés :

- 56% des concepts hors Lieux qui sont alignés Wikidata
- 76% de tous les Lieux (dont 100% pour la France certainement !)

- alignements restant plus accessibles,
- remobilisation des équipes !

Les ateliers collaboratifs

Organisation de plusieurs ateliers d'alignement :

[Magna Grecia Open data](#)

Aix-en-Provence mars 2018

[Atelier FRANTIQ](#)

MAE Nanterre, septembre 2018

[Journées MASA](#)

MAE, septembre 2019

[Atelier en ligne](#)

janvier 2021



Atelier 1er mars 2018, Aix-en-Provence Photo C Molinié CC BY-SA

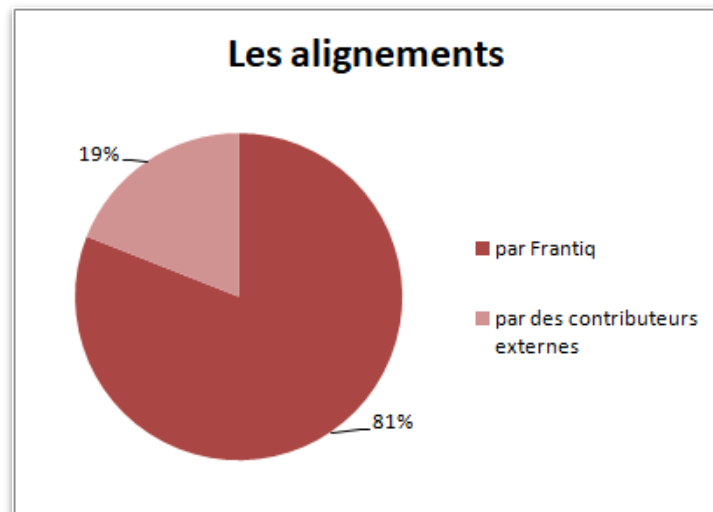
Les contributions sur Mix n'Match

40 contributeurs (avec + de 5 contributions)

7900 alignements = 75,4% réalisés

81%

- 20 membres de FRANTIQU
6400 alignements
- 20 Wikimédiens
1500 alignements



Un premier bilan

- 1ère phase d'alignements de masse quasi terminée
- 2ème phase : ajuster, affiner et traiter les termes non alignés :
 - approfondir la recherche en incluant les traductions à l'alignement automatique
 - évaluer la pertinence à créer de nouveaux éléments sans équivalent sur Wikidata par ex. : tegula et [imbrex](#) ; [Treilles](#)
 - traiter les éléments Wikidata alignés avec plus d'un concept PACTOLS par ex. : [divinité](#) aligné avec les concepts dieux et divinités des PACTOLS
- Inclure l'alignement Wikidata à la gestion quotidienne du thésaurus

Des enrichissements mutuels

Sur PACTOLS :

- Récupérer les identifiants d'autres référentiels (IdRef, AAT, Geonames, BnF ...)
- Bénéficier des ressources de l'écosystème Wikimédien (images, définitions et traductions)
- Partager largement les ressources FRANTIQ (catalogue biblio et thésaurus)
- et les inscrire dans le web de données

Sur Wikidata :

- Verser les identifiants des référentiels préalablement alignés avec les PACTOLS (IdRef, Geonames, Insee)
- Apporter l'expertise des membres de FRANTIQ pour créer de nouveaux éléments, enrichir les traductions, corriger et désambiguïser

Merci !

