

RefDivinités : bilan d'un projet collaboratif d'harmonisation des dénominations

Alyx Taounza-Jeminet — Sébastien Dalmon
BIS Sorbonne
alyx.taounza-jeminet@bis-sorbonne.fr — sebastien.dalmon@bis-sorbonne.fr

Un Projet Collex-Persée

La BIS, bibliothèque délégataire Collex pour les Sciences de l'Antiquité, signale ses riches collections dans ce domaine dans le catalogue collectif SUDOC et utilise le référentiel de l'ABES IdRef. Elle s'est rapprochée du réseau FRANTIQ (Fédération et Ressources sur l'Antiquité), qui dispose de son propre catalogue collectif indexé (CCI) en utilisant le thésaurus spécialisé PACTOLS, en cours de réorganisation et de rapprochement avec des bases de données produites par des chercheurs. Le périmètre des divinités et héros du monde antique a été choisi par les deux parties pour débiter un travail de rapprochement entre leurs deux référentiels, tout en les enrichissant. Ce projet a été mené en collaboration avec un comité scientifique composé de spécialistes des religions anciennes, permettant un dialogue fructueux entre chercheurs et professionnels de la documentation.

lo accueillie à Canope par Isis, fresque du temple d'Isis à Pompéi, 62-79, Musée archéologique de Naples | © MANN, CC-PD-Mark via Wikimedia Commons

Vers un web de données liées pour les humanités ?

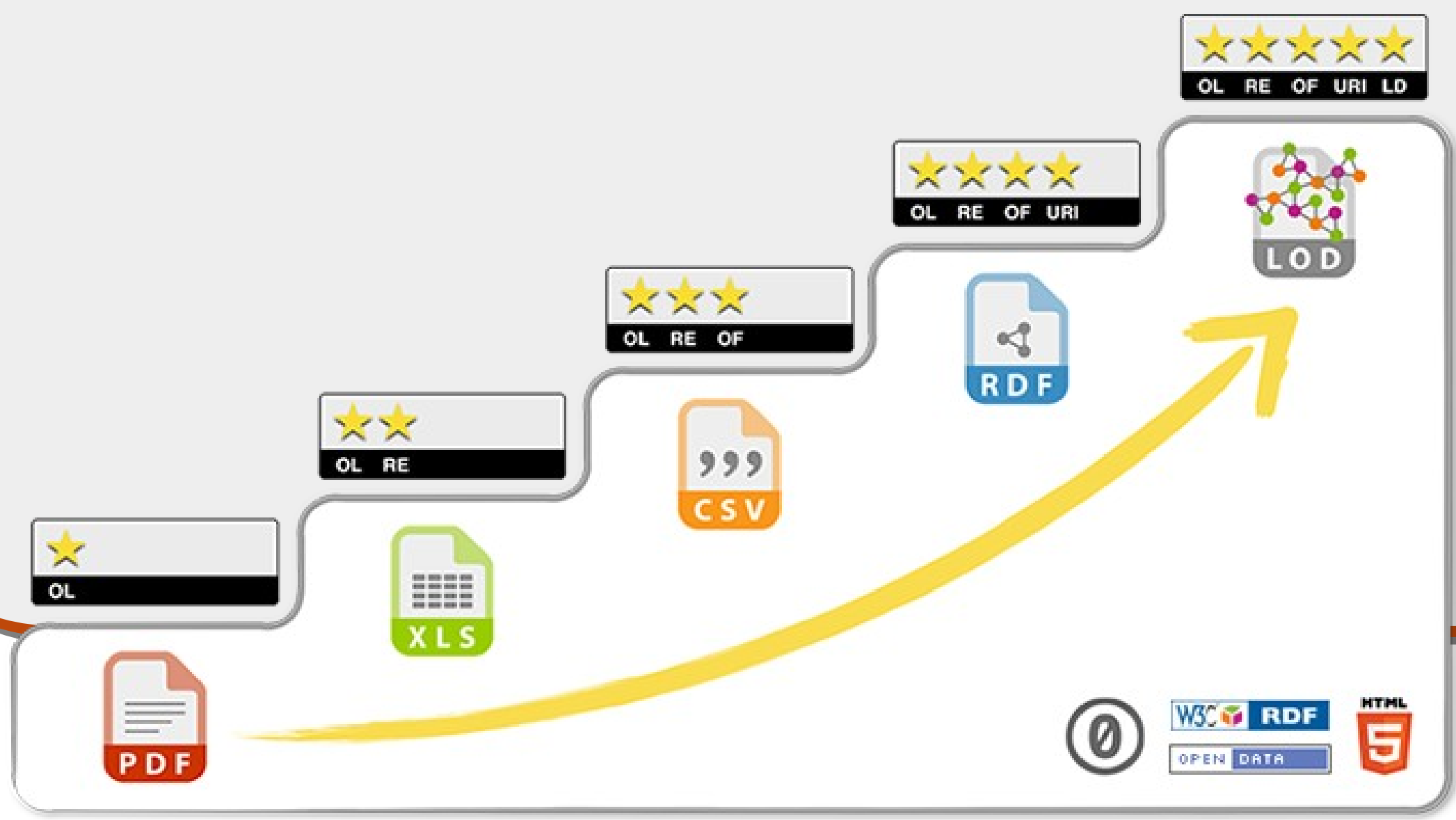
Qu'est-ce que le web sémantique ? Et le web de données liées (Linked Open Data) ?

Mis en avant depuis 1999 par Tim Berners-Lee, l'inventeur du World Wide Web, comme la prochaine évolution de nos systèmes d'information et de partage de données, le Web sémantique propose un partage non plus axé sur les documents mais sur les données.

Le Web sémantique est une extension du web : il s'agit d'un certain nombre de standards mis en place pour structurer les données (grâce à des modèles et langages comme RDF, OWL, XML) et permettre la lecture de celles-ci par les machines.

Le Web de données liées est la pierre angulaire de ce projet : en utilisant les protocoles existants du web (URIs et HTTP), il permet au web sémantique de faire réseau, par l'identification pérenne et l'alignement des entités invitant à l'interopérabilité au sein de et entre différents sets de données.

Au cœur de notre projet se trouve donc l'idée d'une meilleure identification, description et intégration du corpus d'entités au sein de ce web de données liées. Ceci est rendu possible grâce à leur traitement dans les datasets que forment PACTOLS, IdRef et Wikidata.



Wikidata

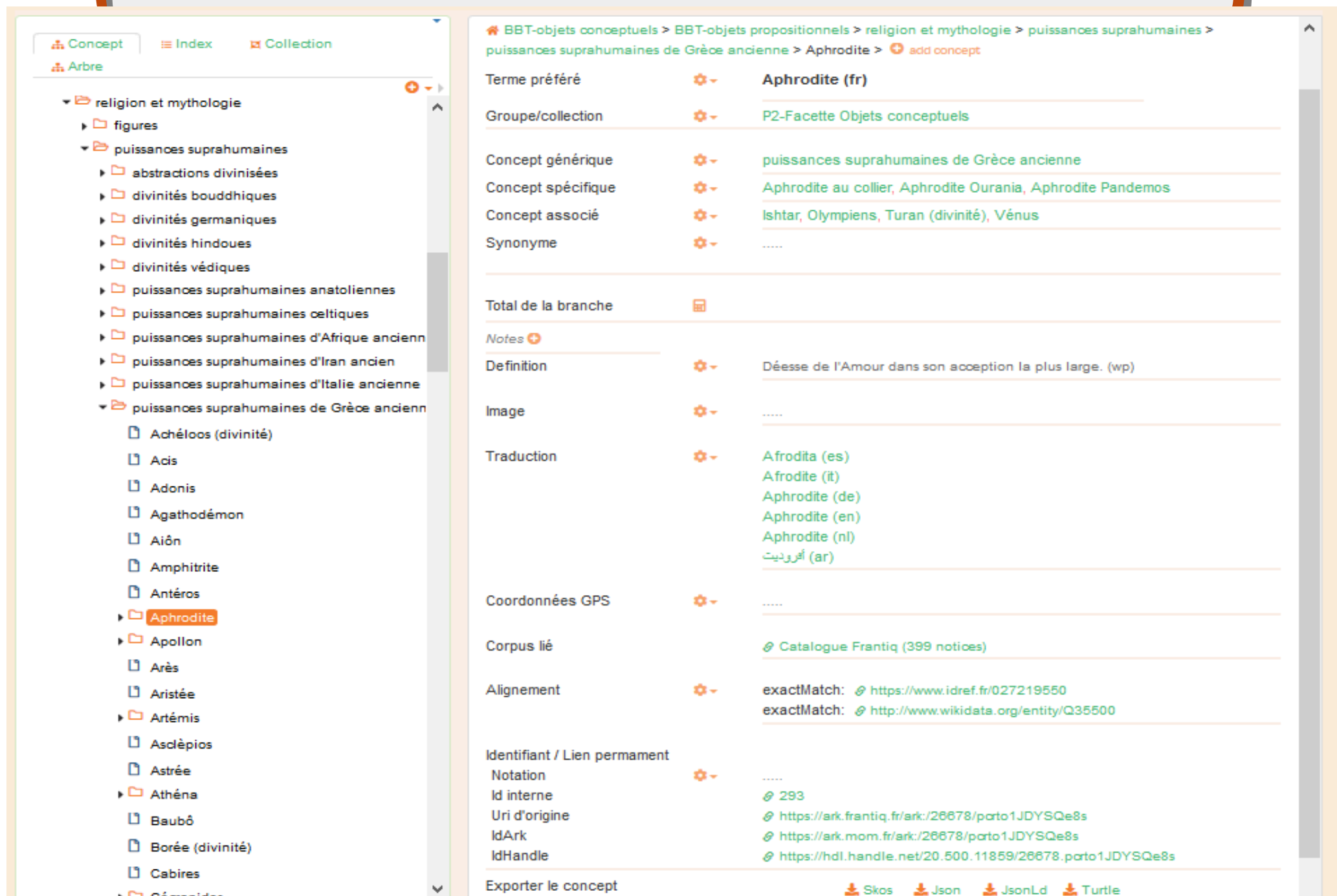
Au sein de cet ambitieux projet d'un web de données ouvert, structuré et lisible par la machine, Wikidata a pris une place centrale. L'un des projets de Wikimedia, créé en 2013, il sert désormais de colonne vertébrale à Wikipédia.

- Plus de 78 millions d'entités
- Libre et gratuit
- Une structure plus complexe que le RDF mais adaptée pour l'interopérabilité
- Un hub pour l'alignement entre différents datasets

PACTOLS

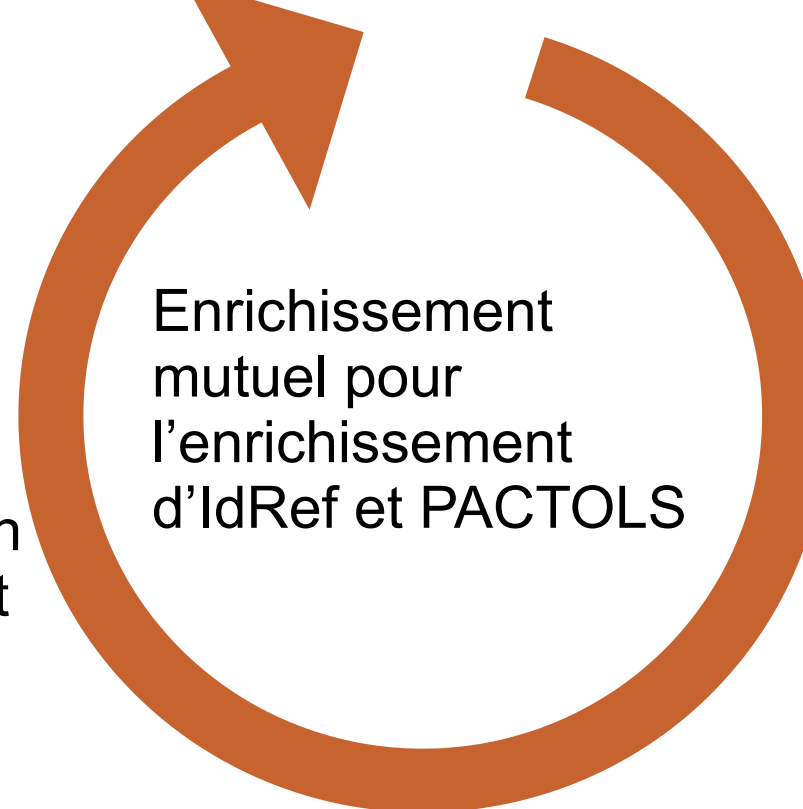
PACTOLS est l'unique vocabulaire français spécialisé dans l'archéologie qui soit contrôlé, gratuit, multilingue et interopérable. Il permet au réseau Frantiq d'indexer son Catalogue Collectif, mais ses utilisations se multiplient pour offrir un vocabulaire normalisé aux mondes de l'édition ainsi que de la recherche.

L'acronyme désigne les différents thésaurus qui composent aujourd'hui PACTOLS :
Peuples / Anthroponymes / Chronologie / Toponymes / Œuvres / Lieux / Sujets
Depuis 2017, avec le soutien du consortium MASA, PACTOLS est en pleine restructuration, sur les bases d'un autre modèle, le Backbone Thesaurus.



Capture d'écran de la notice Aphrodite dans PACTOLS

Bibliothèque de la Sorbonne (BIS)



Présente les enjeux numériques et oriente la réflexion

Apporte son expertise et propose de nouvelles entités

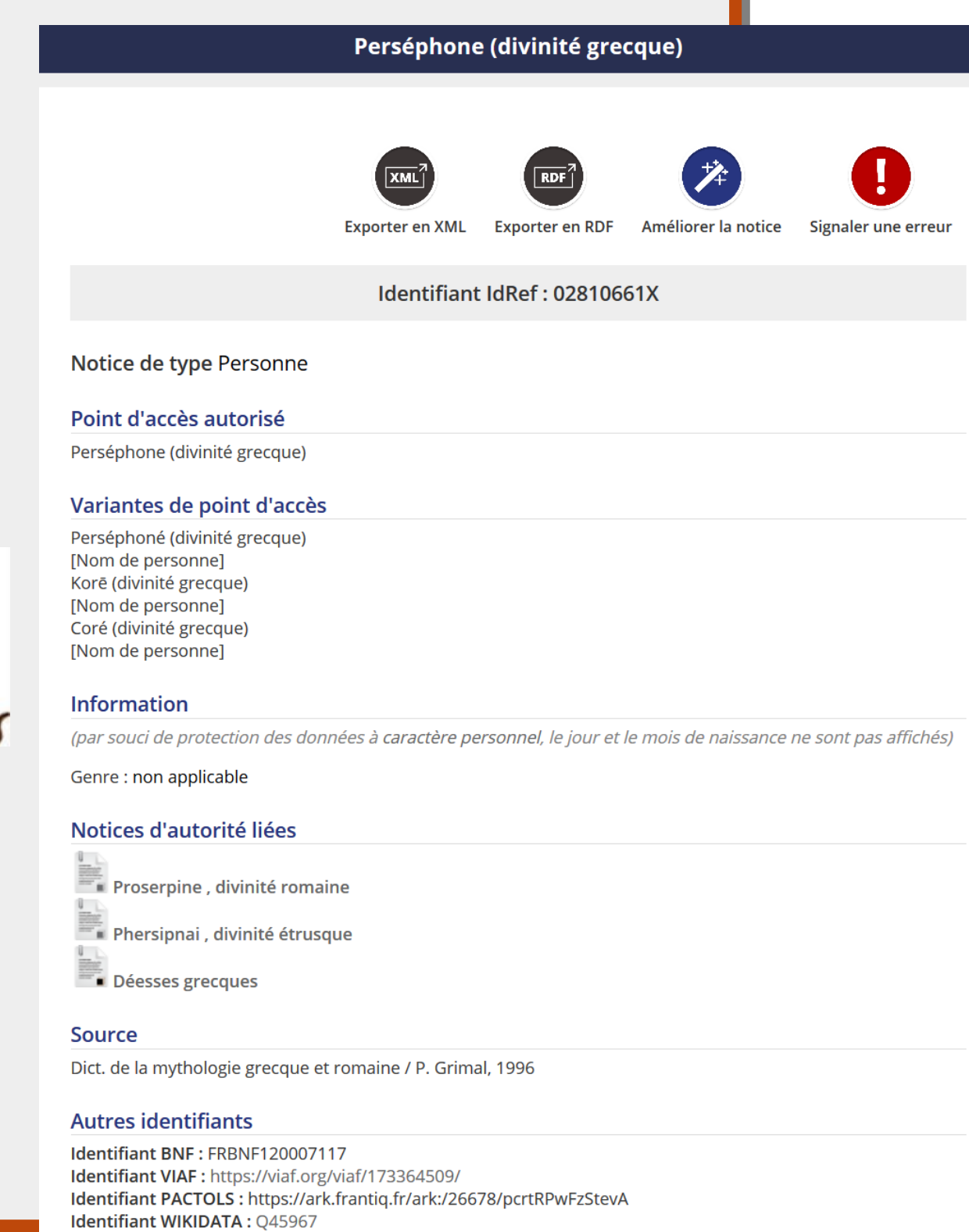
Conseil Scientifique

Composé d'historiens et historiennes spécialistes des religions antiques :
Nicole Belayche, Corinne Bonnet, Raphaël Golosetti, Jean-Claude Lacam, Alice Mouton, Vinciane Pirenne-Delforge, Gaëlle Tallet...

IdRef

IdRef structure plusieurs bases : SUDOC, theses.fr, Calames. Il s'agit du référentiel de notices bibliographiques et d'autorités du Ministère de l'Enseignement supérieur, piloté par l'Abes.

En tant que bibliothèque universitaire, la BIS appartient au réseau qui veille à l'enrichissement constant de cette base, aussi bien dans les notices d'ouvrages qu'elle indexe que dans les autorités (auteurs ou sujets), qui servent à les décrire.



Capture d'écran de la notice Perséphone dans IdRef

En chiffres

En avril 2019 : 392 notices signalées dans IdRef et 431 dans PACTOLS
En juin 2020 : 664 notices reprises et alignées dans IdRef et PACTOLS
647 alignements avec Wikidata
563 alignements avec VIAF
434 alignements avec la BnF

MAIS AUSSI
Un travail de fond concernant l'évolution de la terminologie.
Des suggestions tournées vers les différents acteurs (ABES, DARIAH, etc.)



OpenTheso



OpenTheso est un logiciel conçu par Miled Rousset de la Maison de l'Orient et de la Méditerranée pour Frantiq. Il permet la gestion du thésaurus PACTOLS mais peut être utilisé par n'importe qui pour gérer son propre vocabulaire structuré.

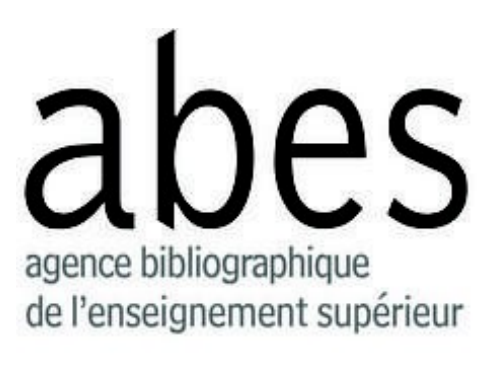
- Libre et gratuit
- Multilingue
- Interopérable avec les langages SKOS et JSON
- Permet l'intégration d'API externes et l'utilisation de widgets pour l'alignement de données

OpenRefine



OpenRefine est un outil d'extraction et de traitement de données. Développée depuis 2010 en open source, l'application devient un projet collaboratif en 2012 quand Google retire sa participation active. L'application ne cesse de s'améliorer, proposant de nouveaux outils et s'adaptant aux besoins de sa communauté d'utilisateurs et utilisatrices.

- Libre et gratuit
- Permet le traitement de différents types de fichiers (csv, txt, json, xml)
- Permet l'extraction de données depuis le web grâce à des APIs
- Permet l'enrichissement de Wikidata



Sébastien Dalmon, BIS, sebastien.dalmon@bis-sorbonne.fr
François Haas, BIS, francois.haas@bis-sorbonne.fr
Blandine Nouvel, Centre Camille Jullian UMR 7299, Frantiq, blandine.nouvel@frantiq.fr
Evelyne Sinigaglia, MSH Mondes Nanterre USR 3225, Frantiq, evelyne.sinigaglia@cncrs.fr
Alyx Taounza-Jeminet, BIS, alyx.taounza-jeminet@bis-sorbonne.fr