

# RefDivinités : Reviewing a Collaborative Project for Data and Denomination Harmonization

Alyx Taounza-Jeminet — Sébastien Dalmon  
BIS Sorbonne

alyx.taounza-jeminet@bis-sorbonne.fr — sebastien.dalmon@bis-sorbonne.fr

## A Collex-Persée Endeavour

The Sorbonne Interuniversity Library (BIS), CollEx delegate library in Ancient Studies, enters its rich collections within the SUDOC, the collective catalogue of universities, using IdRef, a metadata repository monitored by the ABES, to index them. The BIS has approached the FRANTIQ network (Federation and Resources on Antiquity), which has its own collective catalogue (CCI), indexed with the specialized thesaurus PACTOLS. This thesaurus is in the midst of a fundamental restructuring and matching with outside databases from researchers. Both partners agreed on deities and heroes from the ancient world being the scope of this first attempt at aligning the two repositories and mutually enriching them. This project was carried out with the help of a scientific committee comprised of specialists in ancient religions, allowing for a fruitful exchange between researchers and documentation professionals.

Isis welcoming Io in Canopus, fresco from the temple of Isis in Pompeii, 62-79, Naples  
Archaeological Museum | © MANN, CC-PD-Mark via Wikimedia Commons

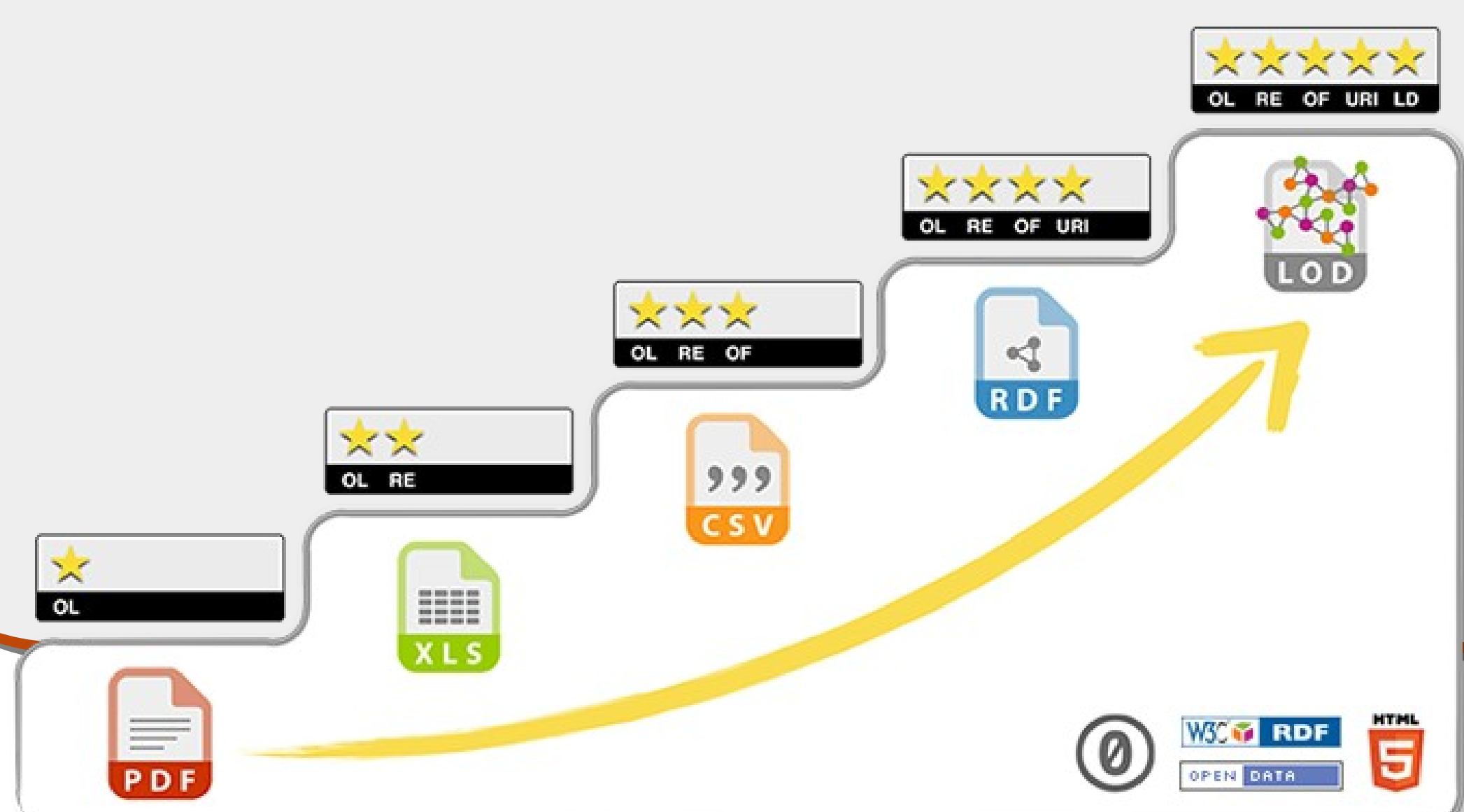
## Towards Linked Open Data for Humanities?

What is the Semantic Web? What about Linked Open Data?

The Semantic Web has been pushed forward since 1999 by Tim Berners-Lee — the inventor of the World Wide Web — as the next step in data sharing and system programming. It revolves around a fundamental change in scope from the sharing of documents to that of data. It forms an extension of the web, putting in place a number of standards to structure data via models and languages like RDF, OWL, XML, to make them machine-readable.

Linked data is essential to this project: using the existing web protocols such as URIs and HTTP, it enables the interconnectivity and interoperability between datasets structured according to the Semantic Web standards. Entities are sustainably identified and matched with their equivalent in other repositories, each containing their own structured metadata.

Our RefDivinités project aims at improving the description and integration of our limited corpus of entities into this web of data. To this end we intervened in PACTOLS, IdRef and Wikidata repositories.



From 5 Star Data - CC0

## Wikidata

Within the confines of this ambitious vision of a web of Linked Open Data, structured and machine-readable, Wikidata has taken a centerplace. Created in 2013 as one of Wikimedia's projects, it is now a key component of Wikipedia.

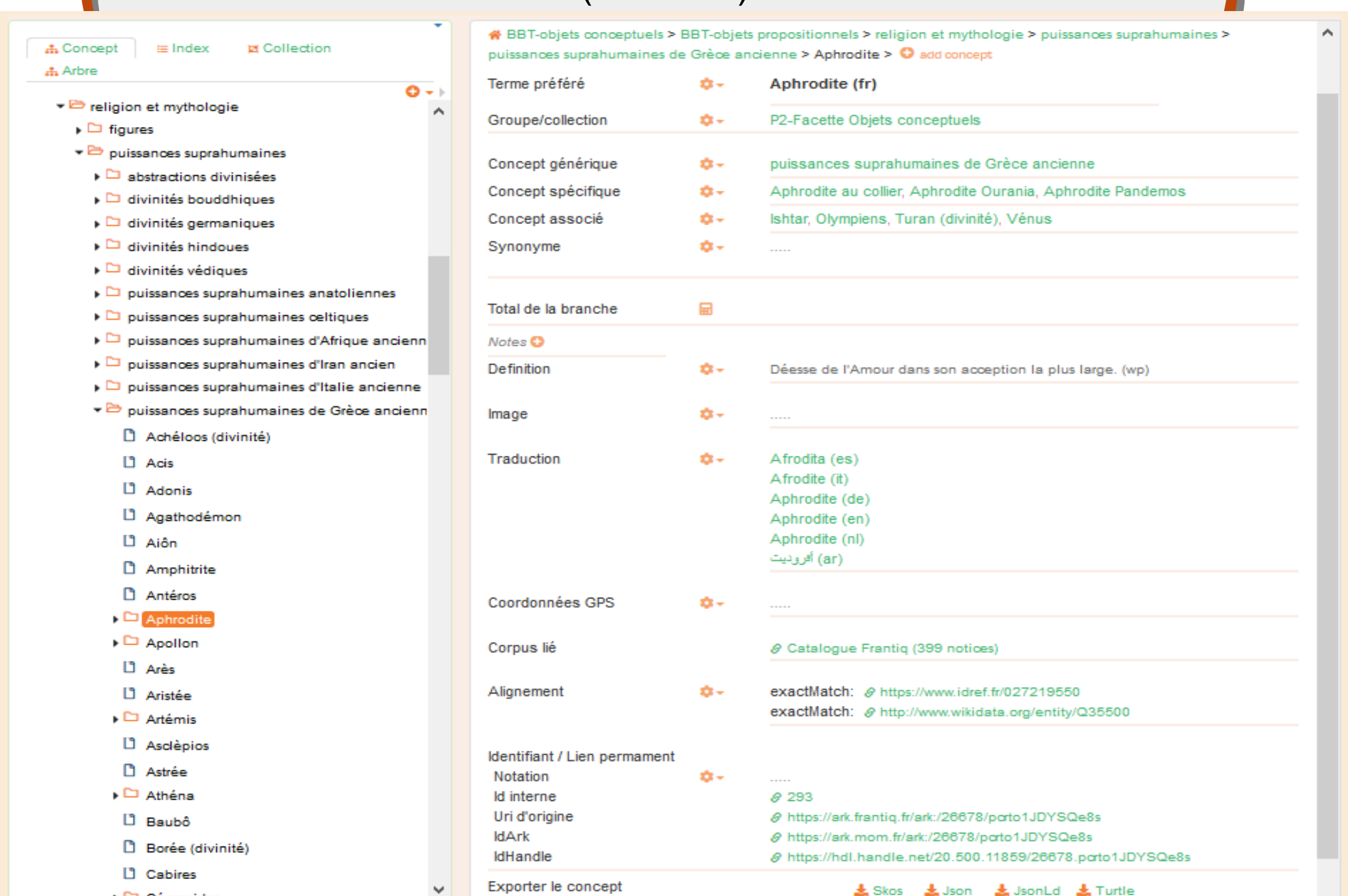
- More than 78 millions entities
- Open and free
- A structure based on the RDF standard fit for interoperability
- A hub where various datasets matches converge

## PACTOLS

The PACTOLS thesaurus is the only controlled, free, multilingual and interoperable French vocabulary for archeology. It enables Frantiq to index its collective catalogue, but its uses are now expanding so as to offer a normalized and structured vocabulary for researchers and publishers.

The acronym stands for the sub-thesauri which PACTOLS has consisted of for decades: Peoples / Anthroponyms / Chronology / Toponyms / Works / Places / Subjects

With the help of the MASA Consortium, PACTOLS is undergoing a major transformation on the model of the Backbone Thesaurus (DARIAH).



Screen capture of Aphrodite's entry in PACTOLS

## Sorbonne Interuniversity Library (BIS)

Mutual enrichment leading to the enrichment of IdRef and PACTOLS

Bring their expertise to correct errors and suggest new entities

Presents the digital issues at stake and steers the reflexion

## Scientific committee

Composed of historians specialized in Ancient Religions  
Nicole Belayche, Corinne Bonnet, Raphaël Golosetti, Jean-Claude Lacam, Alice Mouton, Vinciane Pirenne, Gaëlle Tallet...

## IdRef

IdRef helps index several databases: SUDOC, theses.fr, Calames. It is the metadata registry run by the Abes, collecting bibliographic records and authority records for the Ministry of Higher Education databases. The BIS belongs to the network of university libraries monitoring and enriching this repository, updating the bibliographic records of its collections and the authorities, (authors or subjects) used to index them.



Screen capture of Perséphone's entry in IdRef

## Refdivinités in numbers

April 2019: 392 charted entities in IdRef and 431 in PACTOLS  
June 2020: 664 entities revised and aligned in IdRef and PACTOLS  
647 matches with Wikidata  
563 matches with VIAF  
434 matches with the BnF

FURTHERMORE  
Background work focusing on the evolution of the vocabulary  
Dialogue and improvement suggestions addressed to various partners (ABES, DARIAH, etc.)

Création: 0004-16-03-81 Modifié: 751052105-05-03-14 13:58:00 Statut: 1999-30-09-09  
Nombre des titres reliés: 77  
003 http://www.refdivinités.fr/200360  
008 latp5  
035 ##\$FRNF119344205:FRNF119344205  
106 ##\$a05118:0  
200 ##\$1ySaCratéeS:divinité  
400 ##\$1ySaGrande Mère des dieux  
400 ##\$1ySaMère des dieux  
676 ##\$a200  
810 ##\$aLava RVM: 2000-01  
822 ##\$aCytèle (Goddess)@LCSH: 1999-06

Création: 0004-16-03-81 Modifié: 751052105-10-02-20 14:26:22 Statut: 1999-30-09-09  
Nombre des titres reliés: 77  
003 http://www.refdivinités.fr/200360  
008 latp5  
035 ##\$FRNF119344205:FRNF119344205  
106 ##\$a05118:0  
200 ##\$1ySaCratéeS:divinité  
400 ##\$1ySaGrande Mère des dieux  
400 ##\$1ySaMère des dieux  
676 ##\$a200  
810 ##\$aLava RVM: 2000-01  
822 ##\$aCytèle (Goddess)@LCSH: 1999-06

## OpenTheso



OpenTheso is a software built by Miled Rousset de la Maison de l'Orient et de la Méditerranée (MOM) for Frantiq. Designed for the management of the PACTOLS thesaurus, it can be used by anyone for their own structured vocabulary.

- Open and free
- Multilingual
- Interoperable through SKOS and JSON languages
- Allows for the integration of external APIs and widgets to extract and match data

## OpenRefine



OpenRefine is an extraction and data management application. Developed since 2010 as an open source tool, it became a collaborative project in 2012 when Google withdrew its active support. It keeps evolving, offering new uses and adapting to the needs of its growing community.

- Open and free
- Allows for the processing of various types of files (csv, txt, json, xml)
- Allows for the extraction of data from the web through APIs
- Allows for the direct enrichment of Wikidata since version 3.0